University of Reading

The effect of L2 exposure on processing and perceiving aspect marking and reflexive binding in Mandarin-English bilinguals

PhD

Department of English Language and Applied Linguistics

Supervisors: Dr Clare Wright, Prof Jason Rothman, Dr Ian Cunnings, Dr Jacqueline Laws

Shi Zhang

September 2018

Declaration of original authorship

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Shi Zhang

Acknowledgements

I have been indebted to many people throughout my PhD. First, I would like to express my special appreciation and thanks to my supervisors, Dr Clare Wright, Professor Jason Rothman, Dr Ian Cunnings, and Dr Jacqueline Laws. I am very grateful to them for their academic support, encouragement and insightful feedback. Without their help, the completion of this thesis would not have been possible.

My sincere thanks also goes to my friends, Pengchong Zhang and Yun Wang, for their tremendous help throughout the journey to my PhD. I am also thankful to my participants in the UK and China, for being willing to participate in this study.

Finally, I would like to thank my wife, for her love and support, and playing Splatoon 2 with me.

Table of contents

List of tables

List of figures

Abstract

Extended bilingualism can induce first language (L1) attrition, and lead late sequential bilinguals to deviate from monolingual peers in terms of L1 representation and processing (Schmid & Köpke, 2017). Various theoretical frameworks (e.g. the Interface Hypothesis, Sorace, 2011) have been proposed to explain and predict which L1 domains are vulnerable to attrition, but they still need more empirical evaluation. Moreover, these frameworks have rarely discussed an important issue, i.e. whether and how individual differences in cognitive abilities might explain individual variation in L1 attrition.

Framed within the Interface Hypothesis (IH), this study investigated 14 L1 Mandarin-L2 English bilinguals, and examined whether L1 attrition was more likely to happen with the long-distance binding reflexive *ziji*, a structure at the "external" syntax-pragmatics interface, rather than perfective and durative aspect marking in simple declarative sentences, a structure at the "internal" syntax-lexicon interface. This study also examined whether individual differences in a cognitive ability, namely working memory (WM) capacity, might explain potential variations in L1 attrition. Both off-line and on-line measures were employed in a novel combination including eye-tracking, to assess the perception, production/ interpretation and processing of these structures.

In line with the IH, the bilinguals only demonstrated L1 attrition in interpreting and processing *ziji*, which assumingly involve syntax and pragmatics. However, the IH cannot fully explain the specific patterns of L1 attrition observed in this study. Furthermore, the results partially supported our hypothesis that individual differences in cognitive abilities could explain variation in L1 attrition. The interaction between

WM capacity and other factors explained variation in processing *ziji*, but not in interpreting *ziji* off-line. Nevertheless, this finding highlighted the possible role of cognitive factors in explaining variation in L1 attrition. We conclude with suggestions in how future research can be informed by taking a more a holistic, cross-disciplinary approach to L1 attrition.

1. Introduction

During the last three decades, there has been a rapid growth in research on first language (L1) attrition in bilingual speakers. In empirical studies, L1 attrition usually refers to bilinguals' or multilinguals' deviation from monolingual norms in perceiving, processing and/or producing the L1 due to the influence from a second language (L2) and/or the disuse of the L1 (i.e. not due to pathological reasons such as aphasia or dementia, see e.g. Schmid, 2011; Schmid & Köpke, 2017a). This research topic continues to receive increasing interest, because it allows us to re-examine important issues in linguistics, such as the relationship between the multiple languages in bi-/multilinguals' mind, language processing in bi-/monolinguals and the nature of crosslinguistic influence, from a different perspective.

Prior to the early 2000s, it was largely taken to be the case that attrition had limitations; narrow syntax and/or entrenched L1 processing routines were somehow impervious (or highly resistant) to attrition effects (Tsimpli, Sorace, Heycock, & Filiaci, 2004). Recent research, however, questions this on empirical grounds. A series of studies on how an L2 might affect L1 perception/processing in adult L2 learners revealed that crosslinguistic influence could happen in the L2-to-L1 direction (e.g. processing strategies for high vs low attachment of relative clause interpretation), and that the native, early-acquired L1 was not impervious to change for processing in bilinguals (Dussias, 2004; Dussias & Sagarra, 2007). The same has been shown more recently to happen also at the level of grammatical representations where, in severe contexts of L1 isolation, attrition can happen across virtually all domains of an L1 grammar, even in narrow syntax (Iverson, 2012; Iverson & Miller, 2017). Furthermore, studies on the relationship between onset of age of L2 acquisition/immersion and L1 attrition suggest that, an earlier onset age of L2 acquisition/immersion does not only increase the possibility of acquiring the L2 in a target-like manner, but also increase the possibility of experiencing more severe L1 attrition (Montrul, 2008). Findings such as the above challenge the earlier view that an L1 is invulnerable to change once it is acquired, highlight the interactions between multiple languages in bilingual minds, and help scholars to devise more precise models of bilingualism.

While researchers widely acknowledge the importance of studying L1 attrition, the findings from this line of research are far from conclusive. For example, in the domain of morphosyntax, although existing studies suggest that it is rare to observe L1 attrition among late sequential bilinguals, who acquire an L2 after puberty (as defined by Montrul, 2008, we will turn back to this definition in Section 2.1),[1] a large number of studies still reported that late sequential bilinguals could deviate from their monolingual peers in processing or interpreting certain types of L1 morphosyntax, but

---

[1] The scenario is different when early bilinguals are concerned. Early bilinguals may demonstrate different levels of non-monolingual-like perception/processing/production for a wide range of L1 morphosyntactic structures, but such deviation could derive from complex reasons such as lacking specific L1 input during childhood, and should not be simply interpreted as L1 attrition. In fact, research on this type of bilinguals falls into its own subfield known as heritage language acquisition, and relevant discussions could be found in Kupisch & Rothman (2016), Montrul (2008, 2016), Rothman (2009), Polinsky (2011, 2018), amongst others.

not in all domains of grammar equally (for overviews, see Gürel, 2008; Schmid, 2011; and Schmid & Köpke, 2017a). At present, there is an unresolved controversy about how to explain and predict this selectivity of L1 attrition in morphosyntax, i.e. which types of L1 morphosyntax are vulnerable to L1 attrition, while other types are not.

Among the attempts to providing a theoretical framework for L1 attrition, the Interface Hypothesis (Sorace, 2011, henceforth the IH) seems to be the most frequently and systematically examined one so far. Based on a number of studies on bilinguals' interpretation, processing and production of forward and backward anaphors (Argyri & Sorace, 2007; Serratrice, Sorace, & Paoli, 2004; Tsimpli et al., 2004, amongst others), Sorace (2011) suggests that "language structures involving an interface between syntax and other cognitive domains are less likely to be acquired completely than structures that do not involve this interface", and that these structures are more likely to be subject to L1 attrition "in the very early stages" (Sorace, 2011, p. 1). Furthermore, Sorace (2011) argues that language structures which involve an "external" interface, e.g. between syntax and pragmatics/discourse, are more vulnerable to L1 attrition than those which involve an "internal" interface between syntax and lexicon/semantics (Sorace, 2011, pp. 7–9). This hypothesis is highly testable, and it has inspired a number of systematic research on L1 morphosyntax. Some of these studies reported findings in support of the IH (Chamorro et al., 2016a, 2016b), but others observed that attrition could also happen for language structures at "internal" interfaces (Domínguez, 2013; Iverson, 2012; Ko, 2014).

With respect to these and other conflicting findings the IH should not expect, some scholars argue that the IH needs modification to enhance its explanatory and predictive power (e.g. Domínguez, 2013; Montrul, 2011; Rothman, 2009; White, 2011). Moreover, others have argued that alternative language acquisition models, such as the Feature Reassembly Hypothesis (Lardiere, 2005, 2009, see discussion in Section 2.2.1), could and should be applied in L1 attrition research to better explain the collective findings (Schmid & Köpke, 2017a). In order to examine these claims, empirical evidence from more language structures in different languages would be necessary, and this dissertation intends to fill this gap.

Meanwhile, although the IH and competing proposals differ in theoretical foundation and address the selectivity of attrition in L1 morphosyntax differently, they all seem to assume that the potential source of attrition in L1 morphosyntax representation and/or processing include (1) crosslinguistic influence from the L2, (2) variations in cognitive abilities, such as potential bilingual/monolingual differences in the availability or allocation of cognitive resources. Many empirical studies have examined the first type of source, yet very few have directly investigated the second type of source. It should be noted that, examining the second type of source is very important, as it could contribute to constructing theoretical models which also explain individual variations in L1 attrition. Indeed, if the potential difference in the availability/allocation of cognitive resources is a source of L1 attrition, then, at least at the processing level, individual differences in cognitive resources could lead to different outcomes of L1 attrition in bilinguals with similar language background. However, even in existing studies concerning L1 processing in bilinguals, very few have included any factor related to cognitive abilities, and looked into whether such factors played a role in L1 attrition. An exploration of how cognitive factors may affect L1 attrition should help us to construct a more precise and powerful theoretical framework of L1 attrition.

The present study has two major objectives. The first objective is to further examine the selectivity of L1 attrition in morphosyntax in typical cases - i.e. not in the context of linguistic isolation, where no L1 can be accessed (see Iverson, 2012; Iverson & Miller, 2017). Therefore this study will investigate whether L1 attrition could be observed for two language structures in Mandarin Chinese (henceforth Mandarin), namely perfective and durative aspect marking and long-distance (LD) reflexive binding, in a less investigated population, which is late sequential L2 English speakers of L1 Mandarin (henceforth Mandarin-English bilinguals), who lived in an English-dominant environment but had regular access to L1 Mandarin in daily life. Achieving this objective will provide more empirical evidence, which could be used to examine and further develop current theories of L1 attrition. The second objective is to explore whether and how individual differences in one type of cognitive resources, namely working memory (WM) capacity, might play a role in L1 attrition. In doing so, we will make a step forward to a better understanding about whether and how cognitive factors may affect L1 attrition.

As will be discussed in Section 2.2 and alluded to already, this study will be framed within the IH because, comparing to its competing theories (see Section 2.2 for discussion), the IH allows us to formulate predictions in a more testable way, by constraining L1 attrition based on if the target language structures involve "external" interfaces or not. However, as stated in the objectives, our main goal is to empirically explore L1 attrition in the selected language structures to gain insights about the bilingual mind above and beyond specifically testable implications of the IH alone. In this this context, aspect marking and LD binding are two ideal language structures for testing the IH, because aspect marking (as tested in this study) strictly involves an "internal", syntax-semantics interface, whereas LD binding in Mandarin is more privy to an "external" interface, which involves syntax, semantics and pragmatics. Indeed, aspect marking can involve pragmatics/discourse, because it relates to the viewpoint of the speaker and hearer; however, in the way we test it herein, it relates strictly to the interaction between the grammatical aspect marker and the lexical aspect encoded in verbs/predicates in simple declarative sentences (see Section 3.6-3.8). Therefore, by looking into these two language structures, we will be able to examine whether the IH is suitable for predictingand explainingL1 attrition in these less investigated language structures, and explore whether the competing theories would provide better explanations for the potential L1 attrition in these structures.

Moreover, these two language structures are also good candidates for examining the relationship between individual differences in cognitive abilities and variations in L1 attrition, given their relative complexities in structure and processing terms. Compared to the aspect marking phenomenon examined in this study, which only requires speakers to integrate the information encoded in lexical verbs/predicates and that encoded in grammatical aspect markers while reading simple declarative sentences, LD binding on all accounts is formally more complex - it is a structure that displays inherent linear distance, and it requires speakers to hold multiple types of information during the processing of the LD binding reflexive (see Li & Zhou, 2010). Therefore, LD binding will tax working memory (WM) in a way that aspect marking should not, and studying these two language structures will not only allow us to understand the relationship between individual differences in cognitive abilities and variations in L1 attrition in general, but also allow us to understand how this relationship may vary depending on linguistic factors, i.e. language structures.

Following proposals advocating using both off-line and on-line measures in L1 attrition research (e.g. Iverson & Miller, 2017), this study uses tasks of both types to look into the perception, production and processing of the target language structures in Mandarin-English bilinguals. In doing so, this dissertation stands out as relatively novel, as currently there are very few studies in L1 attrition that combine off-line and on-line experimentation (see also Miller, 2017). Combining methods is especially welcome since one cannot and should not take for granted that what is shown in off-line behavioural experimentation will be confirmed in on-line measures, be them behavioural (reaction time/eye-tracking) or neurological (EEG/ERP) (Miller, 2017; Puig-Mayenco et al., 2018; Villegas, 2014).

Chapter 2 presents the background of this study. We will first define L1 attrition, and then describe the IH and some alternative theoretical frameworks. We will then point out that, while all these theoretical frameworks seemed to assume that cognitive factors played a role in L1 attrition, none have explicitly explained whether and how this would be manifested in empirical research. Following this, we will briefly introduce the concept of WM capacity, and discuss why this cognitive ability may affect L1 attrition overall and at the individual level. The following two sections of this chapter introduce the properties of the two language structures (i.e. perfective and durative aspect marking and LD binding reflexive in Mandarin) to be investigated, and review a selection of relevant literature on the acquisition and processing of these two structures. Finally, we will present the research questions and hypotheses.

Chapter 3 describes the methodology of this study. It will start with the language background of our participants, followed by the research design. A questionnaire was used to collect information about the participants' language background, and a digits-back recall task was used to measure their WM capacity. An abridged HSK-3 test was also used to test whether the participants had sufficient reading and listening skills for this study. With respect to the two concerned language structures, an acceptability judgement task (AJT) using a 5-point Likert scale, a written cloze task and a sentence-picture matching task were employed to assess the participants' perception, production and processing of perfective and durative aspect marking, while two interpretation tasks and a visual-world paradigm eye-tracking task were used to assess their interpretation and processing of the LD binding reflexive. The materials of each task will be presented in turn. The last two sections will describe the procedures and ethics of this study.

Chapter 4 describes the participants' language background and the results of the digits-back recall task, and the HSK-3 test. It also presents and analyzes the results of the tasks concerning aspect marking. Following this, Chapter 5 presents and analyzes the results of the tasks concerning the LD binding reflexive.

Chapter 6 discusses the implications of our findings. More specifically, it discusses whether the between-group differences observed in the tasks concerning aspect marking and LD binding should be interpreted as L1 attrition effects, as well as whether and how individual differences in WM capacity explained the variations in L1 attrition. This chapter also argues that the IH and alternative frameworks may not be capable of providing a sufficient explanation of L1 attrition, and that developing a holistic, multidimensional theoretical framework is necessary for future research. In addition to these issues, this chapter also discusses the appropriateness of Schmid & Köpke (2017a)'s definition of L1 attrition. In the final sections of this chapter, we first briefly discuss some issues that are not of direct interest to the present study, but which have implications for Chinese linguistics, psycholinguistics and research methods; then we point out the limitations of this study, and suggest how future studies could overcome them. The final chapter summarizes this thesis and provides a conclusion.

2. Literature review

2.1. Defining L1 attrition

As Gürel (2008) noted, L1 attrition is “a multi-dimensional phenomenon” (p. 432). In sociolinguistic research, L1 attrition is an intergenerational phenomenon, and it generally refers to the death of an ethnic minority language or the shift from one language to another in bi-/multilingual societies, where one language is dominant for social/political reasons. In research concerning bilingualism at the individual level, L1 attrition is an intragenerational phenomenon. A widely adopted definition defines L1 attrition as the non-pathological “loss of, or changes to, grammatical and other features of a [first] language as a result of declining use by speakers who have changed their linguistic environment and language habits” (Schmid, 2011). Language attrition at these two levels may be relevant to each other, as L1 attrition at the individual level could be a precursor of language death/shift at the societal level (Seliger, 1996). In this study, we are only concerned with L1 attrition at the individual level, as our main interest is in how linguistic features and cognitive abilities may affect L1 perception/production/processing in individual bilingual speakers. In the rest of this thesis, we will use the term “L1 attrition” to refer to L1 attrition at the individual level, unless otherwise stated.

L1 attrition can happen in early simultaneous bilinguals, who simultaneously acquire two languages before the age of 3, as well as in early sequential bilinguals, who begin to acquire an L2 after having acquired some aspect of an L1, arguably between the age of 4 and puberty (see Polinsky, 2011). Furthermore, attrition can also be observed in late sequential bilinguals, who begin to acquire an L2 in an immersive setting after having “fully” acquired the L1, assumingly after puberty (Montrul, 2008)[2]. Although L1 attrition could happen in both early and late bilinguals, increasing evidence suggests that the deviation from monolinguals observed in early bilinguals may not only result from individual L1 attrition, but also stemming from the byproduct of previous generational attrition (i.e. L1 attrition experienced by the parents/caregivers of early bilinguals), which affects the input base against which bilinguals in heritage environments construct their grammars (Bayram, Prada, Cabo, & Rothman, 2018; Pascual y Cabo & Rothman, 2012; Rothman & Treffers-Daller, 2014). Indeed, it is difficult to tease apart the impact from these two different processes on early bilinguals. However, such confusion could be avoided if late sequential bilinguals were concerned, as these bilinguals are in a position where secondary factors that contribute to the result of apparent L1 attrition are not as complex - after all, they were once native monolinguals for a significant period of time. As this study concerns whether language structures are subject to L1 attrition after they have been acquired to an adult-like and monolingual-like degree, we will focus on late sequential bilinguals.

While L1 attrition is broadly defined as bilingualism-caused changes to an L1, there is indeed a debate about what types of changes to L1 can be considered as attrition. For example, Seliger & Vago (1991) argue that only permanent, irreversible

[2] The validity of classifying bilinguals based on the chronological onset age of bilingualism may still require further verification, but Montrul's classification is adopted here nevertheless, because current studies suggest that the onset age of bilingualism associates with the outcomes of L1/L2 acquisition and attrition, and it enables us to label different types of bilinguals in an easy way.

changes at the representational level should count as true L1 attrition. In contrast, processing-based accounts of L1 attrition such as Paradis (2007) argue that L1 attrition is the inability to access L1 knowledge during processing, rather than the irreversible changes to an L1 at the representational level. In a recent proposal, Schmid & Köpke (2017a) do not distinguish between changes to L1 at the representational and those at the processing level, and consider both types of changes as L1 attrition. They propose that "[L1] attrition effects begin as soon as L2 development sets in", and such effects may, but do not necessarily lead to apparent changes to L1 knowledge, processing or production; therefore, "to make such a distinction between online/transient and representational/permanent effects of the L2 on the L1, with only the latter being considered instances of attrition, is both artificial and unhelpful, as they merely represent developmental stages on the same continuum" (p. 641). Based on this definition of L1 attrition, Schmid & Köpke (2017a) further suggest that "every bilingual is an L1 attriter" (p. 641).

Although it is questionable whether the claim "every bilingual is an L1 attriter" can hold true, especially for the bilinguals who use L1s in professional contexts and do not show observable attrition effects (for example, see Miličević & Kraš, 2017), in this thesis we will adopt Schmid & Köpke (2017a)'s definition of L1 attrition, and treat any bilingualism-caused processing and representational changes to L1 as L1 attrition for two reasons. On the one hand, it is doubtful whether permanent, irreversible changes to an L1 could actually happen in late sequential bilinguals. A recent study by Chamorro et al. (2016a) suggests that brief re-exposure to an L1 may decrease attrition effects in on-line processing, and future studies may find evidence that changes to an L1 at the representational level are not permanent or irreversible. On the other hand, processing-based definition of L1 attrition like Paradis (2007) missed the fact that changes to an L1 could happen at the representational level (Iverson, 2012). As the definitions like Seliger & Vago (1991) or Paradis (2007) failed to capture the recently discovered properties of L1 attrition, for now it might be best to follow Schmid & Köpke (2017a) to treat L1 attrition as a continuum. Therefore, in this study, we do not differentiate between representational and processing changes to L1, and treat all changes as L1 attrition.

The next section will review a selection of theoretical frameworks, and introduce how these theoretical frameworks explains and predicts L1 attrition in morphosyntax. In line with Schmid & Köpke (2017a), these theoretical frameworks are broadly categorized into formal and other cognitive approaches to L1 attrition, depending on whether they explain and predict L1 attrition based on the formal linguistic features of language structures. The formal approaches to L1 attrition include the IH proposed by Sorace (2011), Domínguez (2013)'s criticism and modification to the IH and the Feature Reassembly Hypothesis (Lardiere, 2005, 2009), and the cognitive approaches include the Activation Threshold Hypothesis (Paradis, 2007) and the Unified Competition Model (MacWhinney, 2012). The merits and problems of these frameworks will be discussed, and we will point out why this study is framed within Sorace's version of the IH, as well as why it is important to explore the relationship between individual differences in cognitive abilities and variations in L1 attrition.

2.2.Theoretical frameworks

2.2.1. Formal approaches to L1 attrition

The current version of the IH, as proposed in Sorace (2011), argues that, comparing to the language structures involving an "internal" interface between syntax and language-internal modules (e.g. syntax-lexicon), those structures involving an “external” interface between syntax and other cognitive domains (e.g. syntax-pragmatics) are more difficult to be acquired completely and more vulnerable to L1 attrition; therefore, near-native L2 learners and L1 attriters are more likely to show optionality and indeterminacy for the structures at the “external” interface than those at the “internal” interface. Sorace (2011) also speculates that such optionality and indeterminacy may result from bilinguals being less efficient in integrating information from multiple sources during language processing, rather than lacking relevant grammatical knowledge. The reduced efficiency may be caused by that the bilinguals’ “knowledge of or access to computational constraints within the language module is less detailed and/or less automatic”, and/or that “they have fewer general cognitive resources to deploy on the integration of different types of information” (Sorace, 2011, p. 15).

A number of studies reported supportive evidence for this claim. For example, Tsimpli et al. (2004) tested the interpretation of null vs overt subject pronouns in a group of 20 Italian-English bilinguals, who had a minimum length of stay in the UK of six years. The researchers predicted that, the bilinguals would differ from Italian monolinguals in interpreting overt subject pronouns, because it involves topic shifting and focus (therefore it is at the syntax-pragmatics interface, if we follow Sorace, 2011); however, these bilinguals would not differ from the monolinguals in interpreting null subject pronouns, of which the interpretation do not involve topic shifting or focus and is regulated by syntax.

In order to examine this prediction, the participants were asked to complete a picture verification task. In this task, they were presented with a set of sentences, and each sentence was presented together with a set of three pictures; these participants were asked to choose which picture(s) matched the meaning of each sentence. The sentences always consisted of one main and one subordinate clause, with the subject of the main clause always being an NP, and the subject of the subordinate clauses either being an overt pronoun or a null subject (see Example 1). The results showed that the bilinguals significantly differed from the monolinguals in interpreting the overt pronouns, as they were more likely to interpret the overt pronouns as a continued topic rather than a new topic; by contrast, the bilinguals performed like monolinguals in interpreting the null subjects. This finding suggests that the bilinguals' interpretation of Italian subject pronouns is more vulnerable to attrition when pragmatic factors are involved. According to Sorace (2011), this finding supported the idea that language structures at the "external" interfaces (i.e. involving syntax and pragmatics, as in this case) are more vulnerable to L1 attrition.

(1) a. Quando $lei_{k/l}/pro_i$ attraversa la strada, l'anziana $signora_i$ salute la $ragazza_k$.

while she crosses the street, the old woman greets the girl

"While she/*pro* crosses the street, the old woman greets the girl."

b. L'anziana $signora_i$ salute la $ragazza_k$ quando $lei_{k/l}/pro_i$ attraversa la strada.

the old woman greets the girl when she crosses the street

"The old woman greets the girl when she/*pro* crosses the street."

(Tsimpli et al., 2004, p. 266)

A more recent study also reported that structures at the “external” interfaces were more vulnerable to L1 attrition. Perpiñán (2011) studied the production and comprehension of subject-verb inversion in matrix questions and relative clauses in Spanish, and tested 13 Spanish-English bilinguals who lived in the US for an average of five years. The researcher expected no attrition effect for subject-verb inversion in matrix questions because it is syntactically obligatory, but she predicted an attrition effect for subject-verb inversion in relative clauses, because it was regulated by syntax, phonological rules and a pragmatic distinction (topic/focus). The results showed that, the bilinguals did not differ from the monolinguals in orally producing subject-verb inversion in matrix questions in an elicitation task, or comprehending subject-verb inversion in relative clauses in a self-paced reading task, but they were significantly less likely to produce subject-verb inversion in relative clauses than the monolinguals in oral and written elicitation tasks.

Using an eye-tracking reading task and a following acceptability judgement task, Chamorro et al. (2016a) and Chamorro et al. (2016b) tested a group of 24 Spanish-English bilinguals who lived in the UK for an average of 7 years, and examined the processing and underlying grammatical knowledge of Differential Object Marking and null vs overt subjects in Spanish. The authors argued that, the Differential Object Marking phenomenon they tested (i.e. only marking animate objects with the Spanish dative preposition *a*) was at the “internal” syntax-semantics interface because it was determined by syntax and the animacy of the object; however, the interpretation of null vs overt subjects relied on the “external” syntax-pragmatics interface because the involvement of topic shifting and focus. Their results showed that, although the bilinguals were monolingual-like in detecting the violation of Differential Object Marking during on-line processing, accepting the sentences concerning Differential

Object Marking, and interpreting null/overt subject pronouns in off-line tasks, during on-line processing they demonstrated a lower level of sensitivity in detecting the pragmatically infelicitous use of null/overt subjects (i.e. Condition 1 and 4 in Example 2) by not showing any significant differences in the time of reading sentences of all the four conditions. Compared to these bilinguals, the monolinguals were significantly slower in reading the Condition 1 and 4 sentences than in reading the Condition 2 and 3 sentences. This finding suggests that language structures at the "external" interface are more vulnerable to L1 attrition even at the processing level, as well as that off-line tasks alone may not be able to reveal such attrition effects - therefore, combining on-line tasks with off-line ones would be optimum in L1 attrition research.

(2) a. Condition 1: ?Overt/subject match

La	madre	saludó		a	las	chicas	cuando	ella	cruzaba
una	calle	con	mucho	tráfico.					
The	mother	greeted-SG		to	the	girls	when	she	crossed-SG
a	street	with a lot of		traffic					

b. Condition 2: Overt/object match

Las	madres	saludaron		a	la	chica	cuando	ella	cruzaba
una	calle	con	mucho	tráfico.					
The	mothers	greeted-PL		to	the	girl	when	she	crossed-SG

(…)

c. Condition 3: Null/subject match

La madre saludó a las chicas cuando *pro* cruzaba

una calle con mucho tráfico.

The mother greeted-SG to the girls when *pro* crossed-SG

(…)

d. Condition 4: ?Null/object match

Las madres saludaron a la chica cuando *pro* cruzaba

una calle con mucho tráfico.

The mothers greeted-PL to the girl when *pro* crossed-SG

(…)

"The mother/s greeted the girl/s when (she) crossed the street with a lot of traffic."

(Chamorro et al., 2016a, p. 6)

However, not all studies on L1 attrition found evidence in support of the IH, and some scholars argue that Sorace (2011)'s version of the IH is theoretically problematic. For instance, Domínguez (2013) notes that, while Sorace (2011) framed the IH within the Minimalist Program (Chomsky, 1995), it seems that none of the current Minimalist views of the language faculty (e.g. Chomsky, 1995; Jackendoff, 1996, 2003; Reinhart, 2006) presuppose that integrating syntax with discourse/pragmatic information is more difficult than integrating syntax with information from any other linguistic module (e.g. semantics); therefore, the distinction between "internal" and "external" interfaces is not "theoretically justified or sufficiently explanatory" (Domínguez, 2013, p. 97). She also proposes a new definition of interface structures based on Minimalist theoretical grounds, which defines interface structures as “derivations which are selected over other possible well-formed derivations generated by the computational system because they are the only ones which meet the interpretive conditions of the interfaces” (p. 99). According to this definition, L1 attrition is likely to happen with the language structures such as null and overt subject pronouns, as they “require checking for contextual appropriateness in the selection of linguistic outputs” (p. 99), but not with structures like basic word order, which do not require such a “checking” process. The author further argued that, these language structures do not have to be at the “external” interfaces, and those structures at the “internal” interfaces can also be subject to L1 attrition.

Domínguez (2013) also reported an empirical study on L1 attrition in using and comprehending null and postverbal subjects in Spanish. The researcher tested two groups of Spanish-English bilinguals. One group consisted of 20 Cuban Spanish speakers who lived in Miami, USA and had regular access to a mixed variety of Spanish, and the other group consisted of 11 European Spanish speakers who lived in the UK and had little access to Spanish. This study employed both production and comprehension tasks, and the results of the production task showed that both groups' use of null and postverbal subjects differed significantly from their monolingual counterparts in terms of frequency; the Cuban Spanish-English bilinguals produced more null and postverbal subjects under the influence of the Spanish variety spoken in Miami, and the European Spanish-English bilinguals produced less postverbal subjects under the influence of L2 English. More importantly, the results of the comprehension task suggested that, while the Cuban Spanish-English bilinguals demonstrated non-monolingual-like preference for subject-verb inversion in intransitive structures, this deviation could not be explained by pragmatic factors, but by syntactic factors. Specifically, the bilinguals did not differ from the monolinguals under two pragmatically distinctive conditions (i.e. broad vs narrow focus), but differed from the monolinguals when unergative verbs were used. These results demonstrated that the so-called “internal” structures could also be subject to L1 attrition, and that input could affect the outcomes of L1 attrition.

In contrast to the IH approaches, Schmid & Köpke (2017a) proposed that a theory of L2 acquisition, namely the Feature Reassembly Hypothesis (Lardiere, 2005, 2009), could be extended to explaining L1 attrition. The Feature Reassembly Hypothesis assumes that lexical items encode a bundle of specific, abstract grammatical features, and claims that the task of L2 acquisition is “to acquire the entirety of the bundle of grammatical features associated with any particular lexical head of the target grammar and assemble them onto the lexical form” (Schmid & Köpke, 2017a, p. 650). During L2 acquisition, the learners will have to re-assemble the feature bundles encoded in specific lexical heads in their L1s towards the L2 settings. Schmid & Köpke (2017a) argue that, in the context of L1 attrition, we should ask whether extensive L2 exposure could weaken the activation of feature bundles encoded in specific lexical heads in the L1, and eventually lead to a re-assembly of the feature bundles towards the L2 settings.

Since it is based on contrastive analyses of L1s and L2s, the Feature Reassembly Hypothesis may be a good way to explain how the quality of L1/L2 input could lead to L1 attrition in language structures at the level of representation (see Domínguez & Hicks, 2016), as well as why there might be different crosslinguistic effects for bilinguals with the same L1-L2 pair. However, there is a problem if we were to predict L1 attrition based on this hypothesis. In its current form, this theory does not sufficiently explain how much similarity/distinction between L1 and L2 lexical heads is necessary to trigger L1 attrition, or how should we determine the level of similarity between two different lexical heads. For example, for a Mandarin-English bilingual, the bare Mandarin reflexive *ziji* “self” differs from the English reflexives *himself/herself* in that *ziji* allows LD binding (i.e. to be bound to an antecedent outside the local domain), but these reflexives are also similar in that both could be bound to an antecedent within the local domain. It seems difficult to predict whether this similarity/distinction is likely to trigger L1 attrition in reflexive binding among Mandarin-English bilinguals. Therefore, this problem prevents us from making testable predictions, even when a L1 vs L2 contrastive analysis is available.

2.2.2. Other cognitive approaches to L1 attrition

Apart from the reviewed formal approaches to L1 attrition, researchers have attempted to explain L1 attrition from a perspective where usage related to relative activation matters a great deal in the start of and cascading effects of attrition. For example, the Activation Threshold Hypothesis as discussed in Paradis (2007) claims that, the frequency of using a specific linguistic item (e.g. a word, or a certain language structure) determines its activation and availability to a speaker. The more often an item is used, it becomes easier for the speaker to activate this item, and the activation threshold for this item becomes lower. In contrast, the threshold of activation becomes higher if the item is inactive. For bilinguals living in an L2-dominant environment, it is difficult to access a language structure with raised activation threshold, and this could cause L1 attrition effect.

In a study framed within the Activation Threshold Hypothesis (ATH), Gürel (2004) further argues that the activation threshold for a linguistic item can raise if it faces competition from an L2 item which has similar form/meaning/function, because the activation of the L2 item would inhibit the activation of the corresponding L1 item and lower its activation threshold. Therefore, L1 attrition effect is likely to be observed for the L1 structures which have analogous forms in the L2, but not for the L1 structures with no analogous forms in the L2. She tested this hypothesis by studying a group of 24 Turkish-English bilinguals' interpretation of overt, null and nominative pronouns, and found supportive evidence. The participants deviated from the monolingual counterparts in interpreting the overt pronoun, which has analogous form in English, but remained monolingual-like in interpreting the null and nominative pronouns.

It should be noted though, as the ATH seems to define L1 attrition as the inability to access relevant L1 knowledge, it implies that the L1 knowledge is not subject to changes at the level of representation. However, as the studies reviewed in the previous section suggest, changes to an L1 at the level of representation can actually happen, and such evidence challenges the validity of the ATH. Based on the current findings, it may no longer be theoretically justified to frame L1 attrition research within the ATH.

A particular usage-based cognitive model of language acquisition, namely the Unified Competition Model (MacWhinney, 2012), might be suitable for modelling L1 attrition as argued by Schmid & Köpke (2017a). For Schmid & Köpke (2017a), this model assumes that language development (including acquisition and attrition) is determined by cue[3] availability, i.e. "the proportion of times the cue is available over

[3] According to MacWhinney (2004), cues refer to the mapping between form and function, so cues can be any linguistic unit, such as phoneme and word.

the times it is needed” (MacWhinney, 2012, p. 217), and cue reliability, i.e. “the proportion of times the cue is correct over the total number of occurrences of the cue”. Meanwhile, in the context of L1 attrition, this development is also constrained by a number of other factors, including (1) neurobiological entrenchment, which allows bilinguals to preserve previously acquired L1 knowledge; however, this process is reversible, allowing an L2 to affect L1 items; (2) transfer, which allows L2-to-L1 crosslinguistic influence; (3) (social) participation/isolation/identity, which drives a bilingual to actively use or suppress the use of L1; (4) resonance (i.e. the process of linking new experiences to old concepts), which allows bilinguals to consolidate and maintain L1 knowledge through oral and written L1 input, as well as internal use of L1; (5) decoupling, which allows a bilingual to decouple an L2 from its dependence on the L1, and minimize the competition between these two languages (for a detailed discussion about these processes, see MacWhinney, 2004, 2012, 2018).

Compared to the ATH, the Unified Competition Model better explains why most aspects of an L1 are resistant to attrition, and allows us to make specific and interesting predictions. For instance, within this framework, one can measure the availability and reliability of cues in the L1/L2 input, and predicts whether a bilingual receiving such input would show attrition in certain language structure; by introducing the concept of cue reliability, it is possible to explain why certain L1 structures are not subject to attrition despite not being highly available in an L2. Moreover, the decoupling mechanism also allows us to examine an interesting issue: will bilinguals with very high level of L2 proficiency show less L1 attrition, because s/he is better at decoupling the L1 from the L2 and faces less competition between L1 and L2 during processing? Unfortunately, to date there seems to be no L1 attrition study framed within this model, and these questions have not yet been answered.

While it is certainly promising to study L1 attrition within this framework, this study will not attempt to do so, again because it is difficult to formulate predictions and measure the necessary data based on the assumptions of this framework. This framework predicts L1 attrition mainly based on the cue availability and reliability in L1/L2 input received by bilinguals, but it is not clear when and how these two factors will cause L1 attrition effects. Moreover, given the theoretical complexity of this framework, many other factors have to be taken into consideration when formulating a good prediction. The cue availability and reliability in L1/L2 input received by bilinguals can highly vary - for example, a Mandarin teacher and a Mandarin-English interpreter are unlikely to receive the same L1/L2 input, even if both live in the same L2 environment. Furthermore, we have to consider the bilinguals' individual differences in resonance and decoupling, as well as how these factors interact and affect the outcomes of L1 attrition. Measuring each of these factors can be unfeasible for an experimental study like this one, and it might be better examined using other approaches, such as computational modelling.

2.2.3. What is missing in these theoretical frameworks?

The last two sections critically reviewed a number of formal and cognitive theoretical frameworks of L1 attrition. Although the non-IH approaches, such as the Feature Reassembly Hypothesis and the Unified Competition Model, might better explain some issues, such as the roles of L1/L2 input played in L1 attrition, at present it is difficult to formulate predictions within these frameworks particularly for L1 attrition. That is, while they provide useful frameworks to explain observed L1 attrition effects *a posteriori*, they do not yet make clear predictions on the selectivity of L1 attrition itself *a priori*. The IH still seems to be the approach that allows us to make the most testable predictions, therefore the research questions and predictions of this study will be framed within the IH. However, this does not mean that we are in support of the IH as the sole explanatory theory for evidence of L1 attrition. In Chapter 6, we will not only discuss whether the results of this study support the IH or not, but also discuss whether the results are compatible with other frameworks or not.

While each of the frameworks has its own strength or promise in predicting and explaining part of L1 attrition, it seems that none of them has provided sufficient explanations about the high level of variation in L1 attrition among bilinguals. On the one hand, such variation could be due to L1/L2 input, and this has been explicitly explained by the reviewed frameworks. On the other hand, as it is generally assumed that L1 attrition associates with the co-activation of L1 and L2, or the competition between L1 and L2 during processing, such variation could also be partly caused by the individual differences in cognitive abilities, but this possibility has not yet been extensively discussed in the reviewed frameworks, or explored in empirical research, particularly in combination with formal representational theories.

In Sorace (2011), the author speculated that L1 attrition in language structures at the "external" structures might result from that "bilinguals […] have fewer general cognitive resources to deploy on the integration of different types of information in online language comprehension and production" (p. 15). This speculation seems to imply that, depending on the availability of general cognitive resources, bilinguals may show varied levels or types of L1 attrition in the affected structures in language processing and production. In other words, when other factors such as language background, length of L2 exposure, L1/L2 input were matched, a hypothetical bilingual with fewer cognitive resources (e.g. smaller WM capacity) may still differ from another hypothetical bilingual with greater cognitive resources in processing or producing a certain language structure. Therefore, understanding how individual differences in available cognitive resources may affect L1 attrition will enable us to understand why there is a high level of variation in the instantiations of L1 attrition among bilinguals with similar language background, and it will help us to develop a model that could explain and predict how extra-linguistic factors could lead to individual differences in L1 attrition.

However, none of the reviewed theoretical frameworks explicitly explained the relationship between individual differences in cognitive abilities and L1 attrition. This is not surprising, as most of existing studies focused on the relationship between L1 attrition and formal linguistic features or the frequency of L1/L2 input and use. Even in the reviewed L1 attrition studies concerning on-line processing, such as Chamorro et al. (2016a) and Chamorro et al. (2016b), cognitive abilities have not been included as a predictor. In the studies by Kasparian et al. (2017) and Kasparian & Steinhauer (2017), the researchers tested the working memory capacity of their bilinguals, but did not include working memory capacity as a covariate in their analyses. Given the theoretical importance of understanding the relationship between cognitive abilities and L1 attrition, as well as the lack of relevant empirical research, it seems necessary to conduct a study and explore this issue. In order to do so, this study will explore the relationship between individual differences in WM capacity and variation in L1 attrition among bilingual speakers. We decided to look into WM capacity because it is often assumed that individual differences in this cognitive ability associates with variations in language comprehension and processing (Just & Carpenter, 1992; Waters & Caplan, 1996; cf. MacDonald & Christiansen, 2002). In the next section, we will briefly introduce the concept of WM capacity and how this cognitive ability might affect language processing and comprehension. We will also discuss how the investigation into the relationship between WM capacity and L1 attrition could contribute to the field of L1 attrition research.

2.3.Working memory, language comprehension/processing and L1 attrition

Working memory (WM) generally refers to "the temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities" (Baddeley, 2003, p. 189). The influential multi-component WM model (Baddeley, 2003, 2012) posits four components of WM: (1) the central executive system, which controls the attentional focus; (2) the phonological loop, which temporarily stores and processes verbal information; (3) the visualspatial sketchpad, which briefly maintains and manipulates visual and spatial information; (4) the episodic buffer, which integrates information of different types into multidimensional code, holds the multidimensional representation, communicates with the central executive system, and links WM to perception and long-term memory.

In contrast to the multi-component model, the state-based models of WM do not assume separate storage for verbal and visual/spatial information; they view the WM as "the set of items in long-term memory (LTM) that are currently active" (Truscott, 2017, pp. 314–315), and suggest that there are different states of activation of information in the WM. For example, Cowan (2005)'s well-known state-based model distinguishes between two states in the WM: (1) the focus of attention, in which a limited number of items are held for immediate access and manipulation; (2) the activated long-term memory, which holds items that were previously in the focus of attention. Unlike the focus of attention, the activated long-term memory does not have a capacity limit, but items in this state are prone to decay or interference.

While the numerous WM models vary in their assumptions about the structure of the WM system, they all seem to agree on two issues: (1) the WM has an executive function, which is used to "activate, focus, update, switch, and inhibit memory during information processing" (Linck, Osthus, Koeth, & Bunting, 2014, p. 862). This is achieved by the central executive system in the multi-component WM model, and attentional control processes in the state-based models (see D'Esposito & Postle, 2015; Linck et al., 2014); (2) the WM has a limited storage capacity and it can only hold a fixed number of items at one time (assumingly four chunks, see Baddeley, 2012; Cowan, 2005; see also Jonides et al., 2007 for a summary of arguments assuming a "one item" capacity) in the episodic buffer or focus of attention. Moreover, individual differences in WM capacity seems to exist, as measured by the complex span tasks which simultaneously require storage and processing of information (e.g. the reading/listening span task, see Daneman & Carpenter, 1980; cf. Waters & Caplan, 1996, 2004). Such differences indicate that people vary in this cognitive ability, although it is debatable whether they should be interpreted as individual differences in cognitive resource (e.g. WM storage resource) or executive control (e.g. susceptibility to interference) (see Lewis, Vasishth, & Van Dyke, 2006; Montrul & Tanner, 2017; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016; Van Dyke & Johns, 2012; Van Dyke, Johns, & Kukona, 2014).

Given the executive function of the WM system, it is evident that it plays a critical role in language comprehension, because language comprehension requires one to store and process "a sequence of symbols that is produced and perceived over time" (Just & Carpenter, 1992, p. 122). Therefore, it is reasonable to speculate that speakers may vary in language comprehension and on-line processing depending on their WM capacity (cf. MacDonald & Christiansen, 2002), especially when memory-taxing linguistic phenomena, such as the resolution of ambiguity, anaphors, filler-gap dependencies, are concerned. For example, King & Just (1991) reported that speakers with a larger WM capacity were quicker and more accurate in comprehending the gaps in objective relative clauses (e.g. "*The reporter that the senator attacked* **__** *admitted the error*", King & Just, 1991, p. 580). In the meta-analysis by Daneman & Merikle (1996), the authors reported that WM capacity measured by complex span tasks was a good predictor for specific language comprehension, which includes detecting ambiguity and assigning pronominal references. In an eye-tracking study, Cunnings & Felser (2013) also found that only the speakers with smaller WM capacity were susceptible to interference from linearly close but structurally inaccessible antecedents during the on-line processing of English reflexives.

With respect to the bilinguals of interest in this study, it is also reasonable to speculate that WM capacity plays a role in explaining individual differences in L1 attrition. Assuming that there is a co-activation of both the L1 and L2 during L1 comprehension and processing in bilinguals, it is likely that bilinguals would have less WM resource than monolinguals because the activated L2 would occupy part of the resource, or that bilinguals would be more susceptible to interference because they face potential interference from both the L1 and the L2. In either case, individual differences in L1 attrition effect may emerge as a function of variation in WM capacity - the bilinguals with larger WM capacity are likely to show less attrition effects than those with smaller capacity, because the bilinguals with larger WM capacity have more WM resource available or are less susceptible to interference. However, this possibility has not been systematically explored, and this gap motivates the present study.

Furthermore, by investigating the role of WM capacity in L1 attrition, it is possible to gain further insights into the relationship between L1 attrition and length of L2 exposure. Intuitively, longer length of L2 exposure should positively correlate with greater L1 attrition effects, because bilinguals with longer length of L2 exposure tend to receive more input in L2 and/or L1 variations used in the L2-dominant environment, and use less L1 in daily life. However, most studies on L1 attrition failed to find such a correlation (see Köpke & Schmid, 2004; Schmid & Köpke, 2017a for discussions), and this failure might be partially caused by ignoring the confounding variable of individual differences in cognitive abilities, such as WM capacity.

Schmid & Köpke (2017a) argue that, rather than the frequency of L1 use, the mode in which the L1 and L2 are activated may play a critical role in determining whether L1 attrition effects would emerge (p. 656-657). For example, the bilinguals who use their L1s in informal contexts (e.g. family) are more likely to show L1 attrition because of frequent code-switching and co-activation of L1 and L2, while the bilinguals who use their L1s in professional contexts, such as interpreters and translators (see Miličević & Kraš, 2017), are less likely to show L1 attrition due to being capable of inhibiting L2 influence through practice. If this is the case, then bilinguals differing in WM capacity should also vary in the "mode" of L1 and L2 activation - the bilinguals with larger WM capacity may be less affected by the activated L2 because of more available cognitive resource or better resistance to interference, and thus have a more monolingual-like L1 activation, while those with smaller WM capacity may not. This potential effect of WM capacity may interact with length of L2 exposure, and omitting this factor in L1 attrition research may lead to the failure of finding an effect of length of L2 exposure. Therefore, it is worth looking into whether L1 attrition effects would vary according to the interaction between length of L2 exposure and WM capacity, and this enables us to further understand how different factors might affect L1 attrition.

2.4.Interim summary

In the previous sections, we critically reviewed a selection of theoretical frameworks for L1 attrition research, and pointed out that at present it seems best to frame the present study within the IH. We also noted that, while these theoretical frameworks all assume that L1 attrition associates with the co-activation of L1 and L2, and imply that individual differences in cognitive abilities may explain variations in L1 attrition, none of them explicitly explained how individual differences in cognitive abilities may explain and predict the outcomes of L1 attrition. Exploring this issue will expand our knowledge about the relationship between L1 attrition and extra-linguistic factors, and help us to develop a more comprehensive theory of L1 attrition.

We also argued that, in order to explore the above issue, it would be interesting to start with investigating the relationship between WM capacity and L1 attrition. Contemporary views suggest that the WM system has an executive function, and it is crucial for language processing and comprehension. Meanwhile, people can differ in WM capacity, and this individual difference associates with variations in L1 comprehension and processing of memory-taxing language structures, such as ambiguity and anaphoric resolution. In bilinguals, individual differences in WM capacity may also associate with variations in L1 attrition effects, and investigating the role that WM capacity played in L1 attrition will contribute to constructing a theory which explains the high level of variation in bilinguals with similar language background. Meanwhile, we noted that previous studies failed to find an effect of length of L2 exposure might be due to not controlling for individual differences in cognitive abilities, such as WM capacity.

In order to explore the relationship between WM capacity and L1 attrition, this study will investigate the processing, comprehension and production of two language structures, namely aspect marking and reflexive binding in Mandarin. More specifically, this study will look into the interaction between lexical aspect and perfective/durative aspect markers (i.e. *le/zhe*) in simple declarative sentences, and the resolution of the long-distance (LD) binding reflexive *ziji*. We selected Mandarin because relatively few studies have yet looked into the L1 attrition in Chinese morphosyntax from a psycholinguistic perspective or using on-line measures (but see Hui, 2012). In doing so, the present study will provide more empirical evidence for research on L1 attrition in a less investigated language.

Moreover, we decided to investigate the L1 attrition of aspect marking and reflexive binding for two reasons. Firstly, within the IH framework (Sorace, 2011), perfective/durative aspect marking in simple declarative sentences belongs to the "internal" interface, because it only involves the interaction between lexical semantics and syntax. In contrast, reflexive binding in Mandarin belongs to the "external" interface as it does not only involve syntax, but also semantic and pragmatic factors (Huang, 1994). Investigating these two structures will enable us to further examine the IH, and find out whether the results could be reanalyzed within its competing theoretical frameworks. Secondly, these two language structures differ, we claim, in the levels of taxing the WM. The comprehension of aspect marking in simple declarative sentences is not memory-taxing, because it only requires the speakers to read the lexical items in linear order, and integrate the information encoded in the lexical verb and the aspect marker at the end. By contrast, the resolution of LD reflexive binding will be more memory-taxing, as it requires the speakers to keep all the possible antecedents in the WM before encountering the reflexive, and integrate the syntactic, semantic and pragmatic information to resolve to which antecedent the reflexive actually refers (a more detailed discussion is provided in Section 2.6.1). Therefore, looking into these two structures will enable us to find out whether the potential effect of WM capacity would differ depending on the properties of linguistic structures.

In the following sections, we will introduce the aspect marking system and the LD binding reflexive in Mandarin. First we will present an introduction to the lexical and grammatical aspect system of Mandarin, followed by a selective review of existing studies on the acquisition and processing of aspect marking. Then we will turn to the LD binding properties of the reflexive *ziji* in Mandarin, and provide a review of relevant acquisition and processing studies on this reflexive.

2.5.Aspect in Mandarin Chinese

2.5.1. The aspect system of Mandarin

In Mandarin Chinese, temporal information can be conveyed using temporal adverbials, such as *mashang* “soon”, as well as aspect markers. Research on aspect in Mandarin usually adopts a two-component approach to aspect (see Vendler, 1967), and assumes a distinction between lexical aspect and grammatical aspect (or situation aspect vs viewpoint aspect, see Smith, 1997; Xiao & McEnery, 2004). According to Smith (1997), lexical aspect is inherent in verbs and predicates, and concerns the internal temporal features of situations. For instance, *run a mile* encodes a natural end point, so it has an internal feature of [+Telic] in terms of lexical aspect; by contrast, *run* does not encode such an end point and has an internal feature of [-Telic].

Grammatical aspect concerns how a temporal situation is viewed from a speaker's perspective, and is realized differently in languages. In Mandarin, grammatical aspect is realized by aspect markers, such as the perfective marker *le* (e.g. *Ta qu nian xiu le fang zi* "He built a house last year.").

Researchers argue that verbs and predicates can have different types of lexical aspect, depending on what semantic features are presented in the verb/predicate (for discussions, see Klein, Li, & Hendriks, 2000; Peck, Lin, & Sun, 2013; Smith, 1997). Smith (1997) suggests that the values of the following features determine the lexical aspect of a verb/predicate:

Table 2.1. Semantic features in Smith (1997)'s model (adapted from Xiao & McEnery, 2004, pp. 41–47)

Features	**Meaning**
[±Dynamic]	A [+Dynamic] situation involves change over time; A [-Dynamic] situation has no internal phases and involves no change.
[±Durative]	A [+Durative] situation is conceived as lasting for a certain period of time; A [-Durative] situation is not conceived as lasting in time.
[±Telic]	A [+Telic] situation has a natural final point; A [-Telic] situation has an arbitrary final point.

Xiao & McEnery (2004) noted that, while [+Telic] situations have natural final points, not all of them encode a result (compare *win a title* with *win*). Moreover, they noted that Smith (1997) did not specify telicity in terms of space or time, and arguably assigned [-Telic] to Semelfactive verbs/predicates like *tap* and *knock*; for Xiao & McEnery (2004), it is worth questioning how an event without a natural final point could happen multiple times (consider *John is coughing*), and therefore produce iterative readings. They proposed a new model for Mandarin, and redefined [±Telic] by associating it with the presence or absence of a natural final spatial point in a situation. They also introduced [±Result], which is associated with whether a situation encodes a result, and [±Bounded], which is associated with the presence or absence of a final temporal point in a situation. According to this model, there are six types of lexical aspect, as shown in Table 2.2 on the following page.

Table 2.2. Xiao & McEnery (2004)'s system of lexical aspect with examples (p. 59)

Classes	[±Dynamic]	[±Durative]	[±Bounded]	[±Telic]	[±Result]	Example
Activity (ACTs)	+	+	–	–	–	*run*
Semelfactives (SEMs)	+	–	±	–	–	*cough*
Accomplishments (ACCs)	+	+	+	+	–	*write*
Achievements (ACHs)	+	–	+	+	+	*arrive*
Individual-level states (ILSs)	–	+	–	–	–	*resemble*
Stage-level states (SLSs)	±	+	–	–	–	*be busy*

Although Xiao & McEnery (2004) and Smith (1997) were similar in defining Activity verbs as [+Dynamic], [+Durative] and [-Telic], they made different claims about the semantic features encoded in Accomplishments and Achievements. Smith (1997) assumed that both types of lexical aspects were [+Telic] and naturally encoded a result. On the contrary, Xiao & McEnery (2004) argued that Accomplishments were [-Result], as they focused on "the process leading up to but not necessarily achieving the implied result (e.g. *chi* "eat" and *xie* "write")" (p. 56), but when they were combined with quantified arguments (e.g. *chi yiwan fan* "eat a bowl of rice") the semantic property would switch to [+Result] (cf. Peck et al., 2013 for an argument against this definition of Accomplishments). Xiao & McEnery (2004) further argued that Achievements lexically encoded a result and "they focus on the successful achievement of the encoded result with or without profiling the process leading up to the result (e.g. *ying* "win", *daoda* "arrive" and *zhaodao* "find")" (p. 56).

It should be noted that Xiao & McEnery (2004) also distinguishes between two types of state verbs, which are ILSs and SLSs. According to them, "ILS verbs […] do not encode a result in the sense that they are normally predicated of permanent dispositions of an individual", whereas "SLS verbs […] do not encode a result in the sense that they are normally predicated of less permanent stages of an individual" (p. 58). This distinction explains why some state verbs can take the progressive felicitously (e.g. *He is being busy.*), while others cannot (e.g. **They are resembling each other.*).

With respect to grammatical aspect markers, researchers widely acknowledge that there are four frequently used grammatical aspect markers in Mandarin (Klein et al., 2000; Liu, 2015; cf. Wiedenhof, 2015; Xiao & McEnery, 2004), as presented in the Table 2.3:

Table 2.3. Grammatical aspect markers in Mandarin Chinese (adapted from Duff & Li, 2002)

Class	Markers	Meaning	Examples
Perfective	*le*	Bounded, perfective	*Ta qu le Shanghai.*
			"He went to Shanghai."
	guo	Experiential	*Ta qu guo Shanghai.*
			"He has been to Shanghai."
Imperfective	*zai*	Progressive	*Ta zai chi wufan.*
			"He is having lunch."
	zhe	Stative, durative,	*Ta chang zhe ge xizao.*
		progressive situation	"He sang while taking a shower."

As shown in the examples in the table above, lexical aspect encoded in verbs/predicates and grammatical aspect encoded in aspect markers interacts to deliver temporal information in Mandarin (compare *Ta qu le Shanghai* with *Ta qu guo Shanghai* in the table above). In this study, we will focus on potential L1 attrition in the interaction between different lexical aspects and two types of aspect markers, namely the perfective marker *le* and the durative marker *zhe*. It should be noted that this study only concerns the verb-final *le*, as sentence-final *le* does not always function as a perfective marker.

Existing research has found that that some aspect markers tend to co-occur more frequently with verbs/predicates of certain lexical aspects, while not co-occurring with verbs of other lexical aspects. For example, in the corpus study by Xiao & McEnery (2004), the researchers argue that the durative marker *zhe* indicates a situation is viewed as enduring or continuing, so it tends to co-occur with verbs/predicates with a [+Durative] feature, such as ACTs, ACCs, ILSs[4] and SLSs, rather than those with a [-Durative] feature, such as SEMs or ACHs. By contrast, when appearing at verb-final positions, the perfective marker *le* indicates the completion or termination of a situation with reference time in the past, present or future, and it tends to co-occur with ACHs, ACCs, ACTs, but not SEMs, ILSs, or SLSs.

Moreover, Xiao & McEnery (2004) also argue that, while *zhe* is strictly incompatible with ACHs or ILSs indicating personal properties, *zhe* could grammatically co-occur with SEMs and trigger an iterative reading (e.g. *Ta pai zhe wo de jianbang* "He kept patting on my shoulder"). With respect to *le,* they argue that although less frequently, *le* could grammatically co-occur with SEMs, ILSs and SLSs (for a detailed discussion, see Xiao & McEnery, 2004, pp. 100–113 & p. 188-194). As Xiao & McEnery (2004)'s theory of aspect marking is a working one that makes testable predictions about how native Mandarin speakers would behave in perceiving the interaction between lexical and grammatical aspect, and producing aspect markers, it will be adopted in the present study. However, it should be noted that the classification of lexical and grammatical aspects in Mandarin remains a controversial topic, and there are numerous alternative approaches to aspect marking in Mandarin (for example, see Klein et al., 2000; Laws & Yuan, 2010; Liu, 2007; Peck et al., 2013); these alternative approaches may make different predictions about how native

---

[4] Not including "those indicating relations, psychological sensations, and adjectival verbs indicating personal properties (i.e. quality verbs)", as noted in Xiao & McEnery (2004, p. 189).

Mandarin speakers perceive/produce the interaction between lexical and grammatical aspects, and they are certainly worth considering as theoretical frameworks in other future studies on L1 attrition in aspect marking in Mandarin.

In the following two sections, we will first look at the L1 acquisition of *le* and *zhe*, in order to demonstrate that native Mandarin speakers are likely to have acquired these two aspect markers before puberty. We will also review some studies which have implications for research on L1 attrition in aspect marking. Then we will review two existing studies on the L1 processing of aspect marking in Mandarin, which inform us of how the processing of aspect marking in Mandarin could be studied.

2.5.2. L1 acquisition and attrition of *le* and *zhe*

Various studies have investigated the L1 acquisition of aspect marking in Mandarin. Regarding the L1 acquisition of aspect marking, Erbaugh (1992) investigated the emergence of the aspect markers among four children whose ages ranged from 1;9 to 3;9. She reported that these children started to produce *le* more and more frequently after 2;4, but the production of *zhe* stayed rare until 3;0. Similarly, a corpus-based study by Chen & Shirai (2010) found that, the frequencies of *le* started to increase in speech produced by children older than 1;7, whereas the frequencies of *zhe* started to increase in speech by children older than 3;0. It was also found that these children reached an adult-like performance in correctly producing *le* with verbs of different lexical aspects around 3;5. However, although the children made few errors in producing *zhe* with verbs of different lexical aspects since the emergence of *zhe* in their speech, incorrect uses of *zhe* were observed in the speech produced by a child with the age of 3;5.

Another study used two picture telling tasks and examined the elicited production from 30 children whose ages ranged from 4;2 to 10;9, and observed that the children behaved adult-like in producing verb-final *le* and *zhe* with verbs/predicates of different lexical aspect after the age of 7 (Jin & Hendriks, 2003). The results of these three studies suggest that monolingual Mandarin speakers are likely to have acquired the interaction between *le/zhe* and different lexical aspect before the age of 10. Therefore, it is reasonable to assume that any changes to perception/production of *le/zhe* in the Mandarin-English bilinguals in this study should be attributed to L1 attrition, rather than developmental problems.

Until now, few studies have looked into the L1 attrition in aspect marking among bilinguals with L1 Mandarin. Some studies observed that the acquisition of aspect marking in Mandarin could be problematic for early bilinguals with L1 Mandarin, and these bilinguals tended to undersupply aspect markers than their monolingual peers (e.g. Jia & Bayley, 2008; Ming & Tao, 2008; Shi, 2011). However, only one of these studies (Shi, 2011) included late bilinguals as participants. In this study, the researcher studied the oral production of aspect markers by testing six adult Mandarin-Dutch bilinguals in the Netherlands, whose onset age of L2 immersion varied from 2;6 to 8;5. She used a video elicitation task and found that, while the early bilinguals who were exposed to L2 Dutch before or at the age of 4 produced fewer aspect markers and made more errors in their speech, the late bilinguals who were exposed to L2 Dutch after the age of 6 did not have such problems. Although the sample size was very small, this study provided some preliminary evidence that aspect marking in Mandarin may not be susceptible to L1 attrition.

2.5.3. The processing of *le* and *zhe*

Regarding the processing of aspect marking in Chinese, Yap et al. (2009) investigated how lexical aspect and grammatical aspect interact during language processing among adult Cantonese speakers. In their study, the participants were required to listen to a sentence, and then match the sentence with one of the two pictures which appeared immediately after the offset of the sentence. The pictures appeared in pairs with one depicting a completed event, and the other depicting an ongoing event. The sentences were either semantically matched or mismatched in terms of lexical aspect: the semantically matched sentences were formed using ACTs ([-Telic]) plus the progressive marker and ACCs ([+Telic]) plus the perfective marker, whereas the mismatched ones were formed using ACTs plus the perfective marker and ACCs plus the progressive marker. It was found that the participants were faster in matching the semantically matched sentences to the corresponding pictures. Therefore, the authors argue that the match in semantic features between the lexical and the grammatical aspect may facilitate the speed of processing aspect marking.

As the facilitation effect brought by the semantic match between lexical and grammatical aspect requires a speaker to have knowledge about the interaction between these two types of aspect, the method used by Yap et al. (2009) might also be suitable for investigating the processing of *le/zhe* among bilinguals. If a bilingual experienced attrition in aspect marking, s/he might be slower than monolinguals in matching the sentence to the correct picture because of having difficulties in integrating the temporal information encoded in the verbs and the aspect markers, or s/he might not even show such a facilitation effect for processing the semantically-matched sentences due to having lost some knowledge about the semantic features encoded in the verbs/predicates or the aspect markers. If not, a bilingual should behave exactly like monolinguals in this task.

However, it should be noted that, the sentence-picture matching task used by Yap et al. (2009) actually measures the participants' reaction times of making decisions (i.e. off-line comprehension of aspect) rather than on-line processing, and it needs the participants to keep the heard sentences in WM before making decisions (see Marinis, 2010). In other words, this task requires simultaneous processing and storage, so the participants' individual differences in WM capacity may affect their performance. Yap et al. (2009) did not look into the role of WM capacity, but the effect of this factor will be examined in this study. In doing so, we will be able to find out whether the potential bilingual/monolingual differences should be explained by L1 attrition, WM capacity, or the interaction between these two factors.

A study by Zhou, Crain, & Zhan (2014) suggests that monolingual children behaved adult-like in effectively integrating lexical and grammatical aspect during sentence processing. This study used a sentence-picture matching task with eye-tracking to investigate 3 to 5-year-old children's and adults' processing speed of aspect marking in Mandarin. They reported that, in both groups, the proportion of fixation on pictures depicting completed events significantly increased after hearing *le*, and the proportion of fixation on pictures depicting ongoing events significantly increased after hearing *zhe*. Moreover, these two groups did not differ in processing speed, suggesting young children could effectively exploit temporal information encoded in aspect markers during sentence processing. This finding indicates that the potential changes to the processing of aspect marking in the late bilinguals in this study are likely to be caused by L1 attrition, rather than L1 developmental problems.

Currently, few studies have yet investigated the processing of aspect marking in bilinguals with L1 Mandarin. Thus, here, by employing a sentence-picture matching task similar to the one used by Yap et al. (2009), it may be possible to find out whether Mandarin-English bilinguals would also show a monolingual-like facilitation effect for sentence processing when lexical aspect and grammatical aspect are matched, or deviate from the monolinguals in terms of processing speed and/or processing patterns. This study will also look into the role of WM capacity in this task, in order to test the assumption that individual differences in WM capacity constrain the performance of this particular task, but not L1 attrition in the processing of aspect marking; in doing so, we will be able to reach a more comprehensive understanding about the processing of aspect marking in Mandarin by bilingual speakers of Mandarin.

2.6.Long-distance binding in Mandarin Chinese

2.6.1. The long-distance binding property of *ziji* in Mandarin

In Section 2.5, we briefly introduced the aspect marking system in Mandarin. In this section, we introduce the other language structure to be investigated in this study, namely the long-distance binding property of *ziji* in Mandarin. The properties of the bare reflexive in Mandarin, i.e. *ziji*, is different from those of the reflexives in English. In English, the reflexives (e.g. *himself*, *herself*) must be bound to an antecedent within its local domain, and this is consistent with the Principle A of Chomsky's Binding Theory (Chomsky, 1981). However, in Mandarin, the bare reflexive *ziji* "self" can either be bound to a local antecedent, or an LD one. See the following example of Mandarin:

(3)	Xiunv$_i$	shuo	yisheng$_j$	digu	le	ziji$_{i/j}$.
	Nun$_i$	say	doctor$_j$	underestimate	PERF	self$_{i/j}$

"The nun said that the doctor underestimated her/himself."

Unlike in English, where the reflexives contain gender and number information (e.g. *himself*, *themselves*), in Mandarin the bare reflexive *ziji* does not contain such information. Therefore, one cannot interpret *ziji* based on the match in gender/number between *ziji* and the antecedents. In order to interpret *ziji* in ambiguous sentences like Example (3), one has to rely on pragmatic/discourse information - for example, if the exemplar sentence is preceded by another sentence *yisheng bugan zuo zhege shoushu* "the doctor is afraid of performing this surgery", it is more likely that the reflexive *ziji* would be interpreted as referring to the local subject, i.e. *yisheng* "doctor" (see Li & Kaiser, 2009). Moreover, the meaning of the verbs may also affect the interpretation of *ziji* - if *digu* "underestimate" in (3) is replaced by *zhiliao* "treat", the reflexive *ziji* should refer to the matrix subject, i.e. *xiunv* "nun" (see Yuan, 1998; Schumacher, Bisang, & Sun, 2011).

Furthermore, some researchers argue that animacy of the antecedents poses a constraint on the interpretation of *ziji*, since *ziji* should refer to an animate subject in general (Huang & Liu, 2001; Huang, 1994; Xue, Pollard, & Sag, 1994). See the following example:

(4) Xiaoshuo$_i$ shuo laoshi$_j$ hai le ziji$_{*i/j}$.
Novel$_i$ say teacher$_j$ harm PERF self$_{*i/j}$
"The novel said that the teacher harmed himself."

As can be seen from the above examples, resolving the antecedent of *ziji* does no only rely on syntactic information (i.e. whether the antecedent is within the local domain of *ziji*), but also relies on semantic, pragmatic and discourse information. However, studies also suggest that, during the on-line processing of *ziji*, speakers would first rely on structural cues when interpreting *ziji* (Dillon et al., 2014; Dillon, Chow, & Xiang, 2016; Li & Zhou, 2010; cf. Jäger, Engelmann, & Vasishth, 2015). In other words, there is a "locality effect" for the early processing of *ziji*, and this effect is manifested by the fact that monolingual Mandarin speakers will first search for the antecedents of *ziji* within the local domain, and then outside the local domain. This effect is interesting within the scope of this study, as it allows us to test whether and how the IH could be extended to predict L1 attrition in processing by looking into whether early syntactic processing is not vulnerable to L1 attrition while later stage processing is, as well as whether individual differences in cognitive abilities, such as WM capacity, would be associated with variations of potential L1 attrition in this language structure.

As with the discussion on aspect, in the next two sections, we will first review studies on the L1 acquisition of *ziji*, which suggests that the LD binding properties of *ziji* is acquired early in childhood; one study on the L1 attrition of *ziji* will be discussed as well. Then we will review the existing studies on the L1 processing of *ziji*, and discuss how studies on the L1 attrition in the processing of *ziji* could be done.

2.6.2. L1 acquisition and attrition of *ziji*

Chien, Wexler, & Chang (1993) studied the L1 acquisition of LD binding among children aged between 3 and 8. They first presented the children a picture that either depicted a co-referential relationship between the LD antecedent and the reflexive, or a co-referential relationship between the local antecedent and the reflexive; then they presented the children a sentence containing the reflexive, and asked them to judge whether this sentence was true or false. The researchers suggested that the children who were older than 4 had acquired LD binding, since they accepted LD binding at an adult-like level when the pictures and the sentences were matched.

A later study presented the stimuli using dialogues and reported that, compared to the adults, the 4- to 5-year-old children accepted LD binding at a significantly lower level when the dialogue constituted a biased context for interpreting *ziji* as bound to the LD antecedent (Su, 2004). However, although the researcher constituted biased contexts in the dialogues, the dialogues were so long that it was necessary for the participants to remember more than five propositions before exploiting the critical contextual information in resolving the reflexive. Therefore, these young children's performance in resolving *ziji* might be caused by their limited WM capacity (Gathercole, 1998), rather than not realizing the LD binding property of *ziji*. Moreover, while the results of this study suggest that the children within the age range between 4 and 5 might have problems with accepting the LD antecedents, such results did not exclude the possibility that at least some of these children could realize the LD binding property of *ziji*, or that children older than 5 would behave adult-like in interpreting *ziji*.

Surprisingly, apart from these two studies, it seems that few other studies have systematically examined the L1 acquisition of *ziji*. Given the methodological flaw of Su (2004)'s study, in this study we assume that monolingual Mandarin speakers have acquired the LD binding property of *ziji* by the age of 8, as suggested by Chien et al. (1993). However, future studies using more refined methods are necessary to further examine this assumption.

Up to now, very few studies have looked into the L1 attrition of *ziji*. Hui (2012) studied seven late sequential Mandarin-English bilinguals with at least 6 years of L2 exposure, and investigated whether the bilinguals would deviate from the monolinguals in interpreting *ziji*. The researcher used a yes-no judgement task, in which the participants were first presented a sentence containing two antecedents, and then asked to answer three questions, which respectively asked whether the reflexive *ziji* could refer to the local, LD antecedents and any other person not mentioned in the sentence. The results suggest that these bilinguals behaved monolingual-like in realizing that both LD and local binding of *ziji* are allowed, and they did not show any L1 attrition in this respect.

It should be noted, though, that Hui (2012)'s findings can be inconclusive. Firstly, there were only seven participants, and the researcher only used three target sentences to test the participants' acceptance of LD and local binding of *ziji*. Secondly, the judgement task used in this study explicitly required the bilinguals to make metalinguistic judgements, and this requirement might have made the results unnatural; as the participants were explicitly asked to judge whether the sentences could be ambiguous, they might have to revise their first impressions about how *ziji* should be interpreted in the tested sentences, and then make the conclusions that both antecedents were possible. Therefore, this task may not be able to reveal L1 attrition -

while the bilinguals accepted the possibilities for LD and local interpretations of *ziji*, they may still show non-monolingual-like preferences when actually interpreting *ziji* in a more naturalistic task. Lastly, semantic/pragmatic factors were not manipulated, and it remains unknown whether the bilinguals would deviate from the monolinguals in interpreting *ziji* when biased contexts were present.

In the present study, we will attempt to fill in the above gaps by manipulating the semantic/pragmatic constraints on the interpretation of *ziji*, as well as testing the actual interpretation of *ziji* among Mandarin-English bilinguals. More details will be presented in Chapter 3.

2.6.3. The processing of *ziji*

A number of studies have studied the on-line processing of *ziji* among monolingual speakers of Mandarin, and all these studies seem to suggest a locality effect during the processing of *ziji*. Gao, Liu, & Huang (2005) used a cross-modal priming task to investigate whether *ziji* would prime the words semantically related to the local/LD antecedent. The 103 monolingual participants were asked to listen to sentences, and then read out the word appeared after having heard the sentences. It was found that the participants responded the fastest to the words related to the local antecedents, and there was no significant difference between the RTs for the words related to the LD antecedents and the RTs for the words not related to any antecedent. According to the authors, these results suggested that the on-line processing of *ziji* was syntactically constrained by Principle A of Chomsky's Binding Theory.

Liu (2009) studied the on-line processing of *ziji* using a cross-modal priming task among 180 Mandarin monolinguals, with a more refined design. In this task, the researcher manipulated the stimulus onset asynchrony (SOA, i.e. the time interval between the offset of the priming stimulus and the onset of the target), and inserted introductory sentences before the critical sentences. The introductory sentences created contexts that were either biased to a local binding interpretation of *ziji*, or an LD binding interpretation of *ziji*. She observed that, no matter whether the stimulus sentences favoured a local or an LD binding interpretation of *ziji*, a stronger priming effect was observed for the words related to the local antecedents when the SOA equalled 0ms. In contrast, when the SOA equalled 160ms, a stronger priming effect was observed for the words related to the LD antecedents. These findings suggested that monolingual speakers of Mandarin attempted to search for the antecedent of *ziji* within the syntactically constrained local domain during early processing, and only search for the antecedent outside the local domain later. This locality effect has been replicated in various studies using different techniques, including event-related potentials (ERP) (Li & Zhou, 2010), eye-tracking (Jäger et al., 2015), and self-paced reading (Chen, Jäger, & Vasishth, 2012; Dillon et al., 2016; Shuai, Gong, & Wu, 2013).

Unfortunately, there seems to be no study investigating the on-line processing of *ziji* among bilingual speakers with L1 Mandarin. It seems to me that the locality effect for the processing of *ziji* would be suitable for testing L1 attrition in processing among bilinguals with L1 Mandarin. As discussed in previous sections, while the early on-line processing of *ziji* is constrained by syntax, the final resolution of *ziji* depends on the integration of syntactic, semantic and pragmatic information. Therefore, if we follow the IH and assume that only language structures at the "external" interfaces are vulnerable to L1 attrition, it is expected that the bilinguals of our interest would show monolingual-like locality effect during the early processing of *ziji*, but deviate from the monolinguals during later processing stages, as it is assumed that the impact of semantic/pragmatic factors will be instantiated during this time period.

In order to do so, the present study will use a visual-world eye-tracking task to investigate the on-line processing of *ziji* among Mandarin-English bilinguals. To our knowledge, this is the first study to use eye-tracking to measure the on-line processing of *ziji* among bilinguals with L1 Mandarin. Compared to the cross-modal priming and self-paced reading tasks, eye-tracking allows us to be more precise in measuring when and how the bilinguals would deviate from the monolinguals, if they do so. Furthermore, unlike the ERP technique, which has to be done in a lab, eye-tracking tasks could be done using portable eye trackers, and this would enable us to collect data from bilinguals living in different regions. More details of this task will be presented in Chapter 3.

2.7.Summary

In this chapter, we first discussed the definition of L1 attrition, and agree with Schmid & Köpke (2017a)'s approach of defining any bilingualism-caused changes in L1 processing/perception/production among late sequential bilinguals as L1 attrition. Then we reviewed a selection of theoretical frameworks of L1 attrition, and argued that at present the IH seemed to be the best option for formulating testable predictions in L1 attrition. More importantly, we argued that, although the existing theoretical frameworks assume individual differences in cognitive abilities might affect L1 attrition, they did not explicitly explain the relationship between this factor and L1 attrition due to the lack of empirical evidence. Therefore, the present study aims to explore this issue by investigating the L1 attrition of perfective/durative aspect marking and LD binding of the reflexive *ziji* in Mandarin. In doing so, the present study will also add empirical evidence for examining the existing theoretical frameworks.

As introduced in this chapter, monolingual Mandarin speakers acquire both perfective/durative aspect marking and LD binding of *ziji* before puberty. Therefore, the changes to both language structures among late sequential bilinguals with L1 Mandarin are likely to be attributed to L1 attrition, rather than problems with L1 development. We also demonstrated that perfective/durative aspect marking in simple declarative sentences only involves syntax-semantic interaction, whereas the interpretation of *ziji* relies on syntactic, semantic and pragmatic information. Therefore, investigation into these two language structures allows us to further examine the IH and its competing theories. Moreover, as these two language structures differ in the levels of taxing the WM, exploring the relationship between WM capacity and potential L1 attrition in these two structures will also allow us to examine how individual differences in cognitive abilities could affect the L1 attrition of different language structures.

With respect to the potential L1 attrition in processing these two structures, we argued that it would be suitable to test whether the bilinguals would demonstrate a monolingual-like facilitation effect for aspect marking, i.e. be faster in processing sentences containing semantically-matched lexical and grammatical aspect, as well as monolingual-like locality effect, i.e. first search for the local antecedent during the early processing of *ziji*. We would also test whether the bilinguals would deviate from the monolinguals during the processing of *ziji* at later stages. In the following and last section of this chapter, we will present the research questions and hypotheses of the present study.

2.8.Research questions and hypotheses

The present study concerns late sequential Mandarin-English bilinguals who have moved to an English-dominant environment after puberty and lived in this environment for an extended period[5], and it attempts to address the following research questions:

After the bilinguals have been extensively exposed to L2 English for at least 7 years …

(1) Will the bilinguals show L1 attrition in the perception, production and processing of durative and perfective aspect marking? More specifically, will the bilinguals behave like monolinguals in perceiving the interactions between lexical aspect and the durative and perfective aspect markers, producing these two aspect markers, and show a facilitation effect for processing sentences containing semantically-matched lexical and grammatical aspect?

(2) Will the bilinguals show L1 attrition in interpreting and processing the LD binding reflexive *ziji*? More specifically, will the bilinguals behave like monolinguals in interpreting *ziji* when biased contexts were present and absent, and show a locality effect for the early processing of *ziji*, and/or deviation from the monolinguals in the processing of *ziji* at later stages?

---

[5] In line with most studies that observed L1 attrition effects, here an extended period refers to at least 7 years.

(3) If L1 attrition is found in the target language structures, will individual differences in WM capacity explain the L1 attrition of the target language structures?

Following the Interface Hypothesis (Sorace, 2011), which predicts that L1 attrition is likely to be observed for the language structures at the “external” interfaces rather than those at the “internal” interfaces, and the assumption that individual differences in cognitive abilities can explain the variations in L1 attrition, we make the following predictions for each of the research questions:

(1) Based on the assumption that the interaction between lexical and grammatical aspect in simple declarative sentences is at the “internal” interfaces, we predict that these bilinguals should not differ from monolinguals in perceiving/producing/processing perfective and durative aspect marking. Moreover, like the monolinguals, the bilinguals will show a facilitation effect for processing sentences containing semantically-matched lexical and grammatical aspect. Therefore, in this study, we do not expect a significant Group (bilinguals vs monolinguals) effect or a significant interaction between Group and other relevant factors (e.g. WM capacity) for the tasks concerning the perceiving/producing/processing perfective and durative aspect marking;

(2) Based on the assumption that the final resolution of the reflexive *ziji* involves “external” interfaces, we predict that the bilinguals will differ from the monolinguals in interpreting *ziji* and in processing *ziji* at later stages. However, as the early on-line processing of *ziji* seems to be constrained by syntax rather than semantics/pragmatics, we predict that the bilinguals would resemble the monolinguals, and show the locality effect during the early processing of *ziji*. In

other words, we expect a significant Group effect or a significant interaction between Group and other relevant factors for the tasks concerning the interpretation of *ziji* and the processing of *ziji* at later stages, but we do not expect a significant Group effect or a significant interaction between Group and other relevant factors during the early on-line processing of *ziji*;

(3) Based on the assumption that individual differences in cognitive abilities affect and can explain variations in L1 attrition, we assume that, if L1 attrition in the target structures were observed, individual differences in WM capacity will explain the L1 attrition effects. In other words, we expect a significant interaction effect between WM capacity and other relevant factors (e.g. Group, Length of L2 Exposure) for the tasks in which L1 attrition effects were found and WM capacity was supposed to have an impact.

In Chapter 3, we will first describe the profiles of our participants, and introduce the research design. Then we will introduce the tasks, test materials and instruments used to address the research questions, and describe the predictions for each task in more detail. Finally we will describe the procedures of conducting this experimental study, as well as the ethics.

3. The present study

In this chapter, we first present the language background of the participants and the ethical procedures, followed by the research design. Then we introduce the tasks and the materials used to examine the perception, production and processing of perfective and durative aspect marking, as well as those used to examine the interpretation and processing of the LD binding reflexive *ziji*. Finally we describe the procedures of the experiment, and the ethics of this study.

3.1.Participants and ethical procedures

In this study, there were two groups of adult participants. All these participants acquired and used Mandarin as their only L1 during childhood, and they were all literate in Mandarin. These participants all had college/university education.

The first group included 14 Mandarin-English bilinguals who have lived in the UK for an average of 13 years (*SD*=7.46, range: 7-30). The mean age of these bilinguals was 33.07 (*SD*=13.13, range: 19-62), and their mean age of arrival in the UK is 19.14 (*SD*=7.13, range: 11-33). None of these bilinguals had been exposed to English or other languages in an immersion setting before moving to the UK. At the time of this study, all these bilinguals were working as professionals or studying in the UK. All of them reported hearing and speaking English during working/studying (4-8 hours per day), and hearing and speaking Mandarin with their Mandarin-speaking friends and family on a daily basis (4-8 hours per day). They therefore can be assumed to be operating in both languages daily. This group of participants was paid £10 each for taking part in this study.

The second group included 23 Mandarin monolinguals, and this group served as the control group. The mean age of these monolinguals was 32.65 (*SD*=8.03, range:

24-50), and none of them had lived in any non-Chinese speaking country. These monolinguals had learned English in middle and high school classrooms, but had limited English proficiency, and they received minimal input in English after finishing high school. They were recruited and tested in China, and did not receive payment for participating in this study.

For this study, full ethical approval was obtained from the researcher’s department, and consents were sought from each individual participant. The generated data were kept confidential and would be used for research purposes only.

3.2.Research design

This section briefly introduces the instruments used in the present study, with fuller details and justification section by section below.

Firstly, a language background questionnaire was used during a face-to-face interview to obtain information about the participants’ language background, including their previous and current residence, length of exposure to English/Mandarin, age of arrival in the UK and self-rated proficiency in English/Mandarin. The questionnaire is adapted from Montrul (2012). An abridged HSK-3 test (Hanban, 2014) was also used to assess if all these participants had the necessary Mandarin listening and reading skills for this study, because the materials used in this study were designed based on HSK-3 grammar and vocabulary. Furthermore, we used a digits-back recall test to measure each participant’s WM capacity.

Secondly, in order to assess the participants’ linguistic knowledge and processing of durative and perfective aspect marking in Mandarin, this study employed an off-line acceptability judgment (AJT) task, a cloze task, and a sentence-picture matching task. The AJT task was designed based on Xiao & McEnery (2004)'s theory of aspect marking, and it concerned the participants' perception of the interaction between lexical aspect and the durative/perfective aspect markers. The goal of this task was to assess the participants' receptive knowledge about durative and perfective aspect marking. The cloze task assessed the participants' productive knowledge about durative and perfective aspect marking. The sentence-picture matching task is an adaption of the task used by Yap et al. (2009) (see Section 2.5.3), and it looked into the processing of durative and perfective aspect marking.

Lastly, in order to assess the participants' linguistic knowledge and processing of the reflexive *ziji*, a pencil-and-paper interpretation task and a visual world eye tracking task were used. The pencil-and-paper task was adapted from Yuan (1998), and assessed the participants' linguistic knowledge about the LD binding property of *ziji*. More specifically, it examined if the participants were able to realize the LD binding property of *ziji*, and interpret *ziji* based on the neutral/biased contexts created by lexical verbs and syntactic structures. The visual world eye-tracking task consisted of two parts. The first part recorded the participants' eye movements while they were listening to a sentence and observing four pictures at the same time, and it concerned the on-line processing of resolving *ziji*. The second part was another interpretation task, which required the participants to interpret *ziji* based on the heard sentence. This interpretation task assessed whether the participants were sensitive to the animacy constraint in the resolution of *ziji* (see Section 2.6.1).

All the data were collected in individual 1-to-1 settings. All tasks, except for the AJT task and the pencil-and-paper interpretation task, were presented to the participants on a 15.6" laptop. The instruction of each task was orally presented to the participants in Mandarin. The questionnaire was designed and presented using

LimeSurvey (GmbH, 2003), and the rest of the tasks were designed and presented in PsychoPy (Peirce, 2007). The AJT task and the interpretation task were completed using pencil and paper. For the eye-tracking task, a Gazepoint GP3 eye tracker was used for this study due to its portability and sufficient specification (60Hz sampling rate).

The following sections will introduce the rationale of these tasks, as well as some exemplar materials used in the tasks. A complete list of the test materials can be found in Appendix 2 or downloaded from the IRIS repository (Marsden, Mackey, & Plonsky, 2016).

3.3.Language background questionnaire

The language background questionnaire is adapted from Montrul (2012), and it collects the following critical information about the participants' language background: (1) current age; (2) length of residence in the UK/China/Taiwan and/or other regions; (3) onset age of acquiring English and Mandarin, and age of arrival in the UK; (4) self-report general proficiency in L2 English and L1 Mandarin, as well as their self-reported proficiency in listening, reading, writing and speaking English and Mandarin; (5) other known languages and self-report general proficiency in these languages. The self-reported proficiency in English/Mandarin/other languages are measured using a 5-point Likert scale (1 = poor; 2 = needs work; 3 = good; 4 = very good; 5 = native speaker command). The questionnaire was presented in English, and translated orally into Mandarin for the monolinguals.

The first three types of information allow us to confirm the following facts: (1) our participants are adults; (2) our bilingual participants have lived in an L2 dominant environment for more than 7 years; (3) our bilingual participants are late sequential bilinguals, i.e. they did not immerse themselves in the L2 dominant environment in early childhood. The fourth type of information allows us to ensure that, while both groups perceive themselves to have native proficiency in Mandarin, only the bilinguals perceive themselves to have a high level of proficiency in English. The fifth type of information allows us to ensure that the monolinguals are not proficient in any language other than Mandarin, and that the bilinguals do not know an L3/Ln which has an aspect system and/or reflexive binding similar to those in Mandarin. Self-reported proficiency scores were used in this study as other studies (e.g. Shameem, 1998) suggest that such scores strongly correlate with actual language performance, but we are aware that self-reported proficiency may not always accurately reflect language abilities (for discussion, see Bachman, 1990).

3.4.Abridged HSK-3 test

The HSK test is a widely used general proficiency test for examining L2 speakers' proficiency in Mandarin Chinese, and the HSK-3 test is designed for intermediate L2 speakers of Mandarin. The instruction of the HSK-3 test suggests that, in order to pass this test, Mandarin learners should have learned at least 600 Chinese words; the grammar book also indicates that learners at this level should have already learned how to use *le*, *zhe* and *ziji*. As this test is designed for L2 learners, in this study the HSK-3 test was not used to measure the participants' general proficiency in Mandarin, but used to confirm that our participants do not have any difficulty in understanding the Mandarin grammar and vocabulary at this level, as all the tests were designed based on HSK-3 grammar and vocabulary. Therefore, a ceiling effect is expected for this test.

Due to time limitation, this study used an abridged version of the HSK-3 test, which contains 10 items for listening comprehension and 10 items for reading comprehension. The items were chosen from the sample HSK-3 tests published by Hanban. The listening comprehension part requires the participants to listen to short dialogues, and answer questions based on the meaning of the dialogues; each dialogue is played twice. The reading comprehension part requires the participants to read short sentences or dialogues which contain one blank, and fill in the blank using one of the six provided options. We are aware that this abridged HSK-3 test is not as comprehensive as a complete version, but a ceiling effect in this abridged test should already be enough to indicate that a participant has sufficient listening and reading skills for all the other tasks.

3.5.Digits-back recall

As discussed earlier in Chapter 2, individual differences in WM capacity might play a role in explaining variations in L1 attrition, and this factor could affect a participant's performance in the tasks which simultaneously require both storage and processing (e.g. the sentence-picture matching task used in Yap et al., 2009). In order to explore the relationship between individual differences in WM capacity and L1 attrition, as well as to minimize the potential confounding effect of WM capacity on the participants' performance in the tasks requiring simultaneous storage and processing, a widely used digits-back recall task was employed to measure the participants' central executive WM capacity.

The design and the stimuli of this task followed Wright (2010). In this task, the participants were presented a series of sets of random digits, in increasing length of sets, at the rate of one digit per second, and they were required to repeat the digits in reverse order after hearing each set of digits. The participants' response would be recognized as correct only if all the digits were correctly recalled in the right order, otherwise the response would be recorded as incorrect. The length of the sets of digits varied from 3 to 7, and each series consists of three sets of digits. The complete list of these digits can be found in Appendix 2. The participants completed a practice session of one series of 3-digit sets and one series of 4-digit sets before proceeding to the formal test. During the formal test, the test automatically ended if a participant was incorrect in responding to two or more sets of digits in any series.

For both groups of participants, the digits were read aloud in Mandarin at a standard speed of approximately 1 second per digits, with an interval of 1 second between each digit. This task was carried out in Mandarin alone (i.e. the participants' native L1) because previous research suggested that administering this task to bilinguals in an L2 might result in poorer performance (see Olsthoorn, Andringa, & Hulstijn, 2014) and not reflect their actual WM capacity, especially when being compared to monolinguals.

Unlike various other studies concerning WM capacity and language processing/comprehension (e.g. Cunnings & Felser, 2013), which adopted a reading span task (Daneman & Carpenter, 1980) to measure WM capacity, this study used a digits-back span tasks instead. However, as that both tasks are assumed to measure the cognitive ability of simultaneous storing and processing information (see Daneman & Merikle, 1996; Linck et al., 2014), doing so should not cause a problem.

3.6.Acceptability judgement task

In this task, the participants needed to read 36 short sentences which either contain *le* and *zhe*, and judge the acceptability of the sentences using a 5-point Likert scale, with

-2 being completely unacceptable and 2 being completely acceptable. As the purpose of this task is to assess whether the participants have comprehensive monolingual-like knowledge about the interaction between the perfective/durative aspect markers and different lexical aspects, verbs/predicates of all the six lexical aspects, i.e. Achievements (ACHs), Accomplishments (ACCs)[6], Activities (ACTs), Semelfactives (SEMs), Individual-level states (ILSs)[7], and Stage-level states (SLSs), were used to create the stimuli.

The stimuli were created following these procedures. First, three verbs/predicates of each lexical aspect were chosen from the examples in Xiao & McEnery (2004) (pp. 33-50, 89-137, 182-204), so there were 18 verbs/predicates in total. Second, for each verb/predicate, a short sentence is created. Third, for each sentence, a perfective sentence was created by adding *le* at the post-verbal position, and an imperfective sentence was created by adding *zhe* at the post-verbal position. By applying this procedure, the 36 stimulus sentences were created. These sentences were presented to the participants in together with the pencil-and-paper interpretation task (see Section 3.10) and another 9 fillers in a pseudo-randomized order (see Appendix 2 for the full set of stimuli). Examples are given in Table 3.1:

---

[6] In order to circumvent the controversy about Xiao & McEnery (2004)'s definition of this lexical aspect (see Section 2.5.1), we always supplied a quantified argument with each Accomplishment verb.

[7] For the ILS predicates, we selected three ILSs indicating personal properties. These ILSs should be compatible with *le* but not *zhe*.

Table 3.1. Exemplar sentences of the acceptability judgement task

ILS+*le/*zhe*				
Ta	pang	le/zhe.		
He	fat	PERF/DURA		
"He has become fat./He is being fat."				
SLS+*le/zhe*				
Ta	mang	le/zhe.		
He	busy	PERF/DURA		
"He has become busy./He is being busy."				
ACT+*le/zhe*				
Ta	chi	le/zhe	fan.	
He	eat	PERF/DURA	meal	
"He has had a meal./He is having a meal."				
SEM+*le/zhe*				
Ta	ke	le/zhe	sou.	
He	cough	PERF/DURA	cough	
"He coughed./He is coughing."				
ACC+*le/zhe*				
Ta	chi	le/zhe	yiwan	fan.
He	eat	PERF/DURA	one-bowl	meal
"He has eaten a bowl of rice./He is eating a bowl of rice."				
ACH+*le/*zhe*				
Ta	dao	le/zhe	zheli.	
He	arrive	PERF/DURA	here	
"He has arrived./He is arriving."				

As discussed in Chapter 2, Xiao & McEnery (2004) argued that the combinations of ACC+*le/zhe*, ACT+*le/zhe*, SLS+*le/zhe*, SEM+*le/zhe*, ILS+*le* and ACH+*le* would be acceptable, but the ILS+*zhe* and ACH+*zhe* sentences would be strictly unacceptable. Therefore, we expect that the monolinguals would accept the sentences containing the acceptable combinations of lexical and grammatical aspects, and reject the ILS+*zhe* and ACH+*zhe* sentences. As the IH predicted that L1 attrition was unlikely to happen for the interaction between lexical and grammatical aspects in simple declarative sentences, we expect that the bilinguals would not differ from the monolinguals in performing this AJT task.

3.7.Cloze test

This production task required the participants to read 22 short sentences which contained one to three blanks, and fill in the blanks with the appropriate aspect marker (*le* or *zhe*) when they feel necessary. The participants completed this task on a laptop. The verbs/predicates used in this task were of different lexical aspects, and they were either highly compatible with *le* or highly compatible with *zhe*. The blanks either require the perfective marker *le*, the durative marker *zhe*, the locative preposition *zai* "at/in", or nothing. Out of all the 33 blanks, there are 8 blanks requiring *le*, 9 blanks requiring *zhe*, and 20 fillers; the verbs/predicates requiring *le* included four Activities, two Accomplishments and two Achievements, and those requiring *zhe* included one Stage-level state and eight Activities. The filler blanks either require *zai* or nothing. The sentences containing the blanks were presented in a pseudo-randomized order. An example is given below ("n/a" means nothing, and the arrows represent the arrow keys on a laptop keyboard), and a full list of the sentences used in this task can be found in Appendix 2:

(3)	Beijing	xianzai	Xia	( )	yu.
	Beijing	now	Down	( )	rain
	"It is raining in Beijing at the moment."				
	↑ = *zai*	↓ = *le*	← = n/a	→ = *zhe*	

For this task, we expect that both groups would behave similarly in being highly accurate in supplying *le* and *zhe*, because the verbs/predicates we used in this task were highly compatible with one of the aspect markers, and the IH predicted that L1 attrition was unlikely to happen for the interaction between lexical and grammatical aspects in simple declarative sentences.

3.8.Sentence-picture matching

In the sentence-picture matching task, the participants first heard a sentence and were presented with two pictures appearing on the laptop screen at the same time; after hearing the sentence, the participants needed to judge which picture matches the meaning of the sentence by pressing an arrow key on the laptop (see procedure in Figure 3.1). This sentence-picture matching task differs from the one used by Yap et al. (2009) in two respects. Firstly, Yap et al. (2009)'s version presented the pictures at the offset of each sentence, while this version presents the pictures at the onset of each sentence - this modification enables us to obtain more real-time RTs for this task, as the participants will not have to recall the heard sentence after having understood the content of the pictures. Secondly, Yap et al. (2009)'s study included both ACT and ACC verbs, but this task only included ACT verbs, because ACC verbs do not co-occur with *zhe* felicitously (see Xiao & McEnery, 2004, pp. 193–194).

There are two types of target sentences, which are the semantically-matched ACT+*zhe* sentences and the semantically-mismatched ACT+*le* sentences. The examples of these sentences, and a figure illustrating how the participants perform this task are given in Example 4 and Figure 3.1 respectively:

	Semantically-matched ACT+*zhe*			
(4) a.	Ta	xi	Zhe	yifu.
	She	wash	DURA	clothes
	"She is washing clothes."			
	Semantically-mismatched ACT+*le*			
b.	Ta	xi	Le	yifu.
	She	wash	PERF	clothes
	"She has washed clothes."			

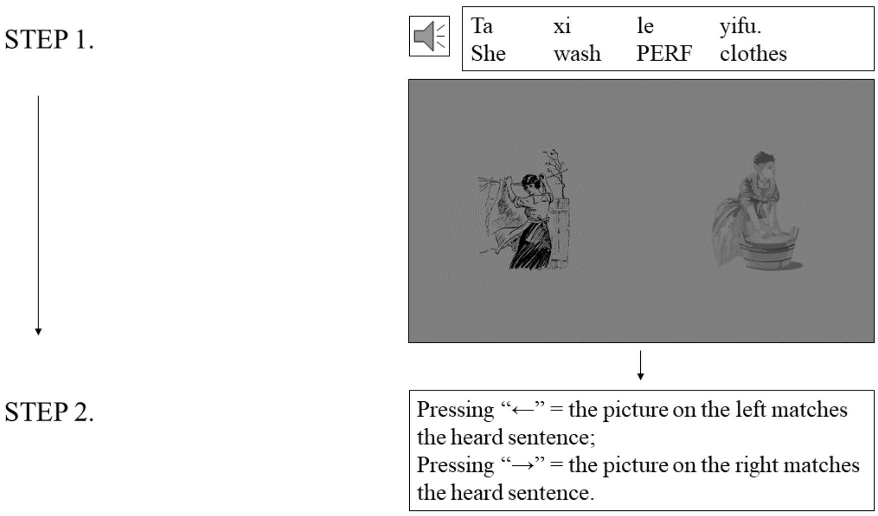


Figure 3.1. Procedures of the sentence-picture matching task

For this task, 10 ACT verbs were chosen, and combined with *le/zhe* for creating the stimulus sentences. All the stimulus sentences were read out by a native Mandarin speaker, and the recordings were edited to ensure that each pair of the *le/zhe* sentences had the same length. The whole stimulus set for this task consists of 5 practice items, 20 ACT+*zhe* sentences, 20 ACT+*le* sentences and 20 fillers. The full list of the stimulus sentences can be found in Appendix 2. The practice items and the fillers were of various structures, and all the participants complete the 5-item practice before proceeding to the formal task. For the purpose of counterbalancing, the stimuli were pseudo-randomized, and two different orders were created; both orders contained the full list of the stimulus sentences. Each participant saw one of the two orders when participating in this study.

The RTs in responding to each sentence were recorded from the onset of each sentence, but the arrow keys were frozen until the offset of the sentence to ensure that the participants did not press the arrow keys too early. As the recordings have been edited to ensure that each pair of the *le/zhe* sentences had the same length, the potential differences in the RTs of responding to the *le/zhe* sentences should be not attributed to the length of the sentences.

In line with the IH, we expect that the bilinguals would not differ from the monolinguals in performing this task in terms of accuracy and RTs. Moreover, we expect that both groups would show a facilitation effect for processing the semantically-matched ACT+*zhe* sentences than processing the semantically-mismatched ACT+*le* sentences. However, as individual differences in WM capacity might constrain the performance in sentence-picture matching tasks (see discussion in Chapter 2), a significant effect of WM Capacity across all participants on the performance of this task is expected, but there should not be a significant interaction between WM Capacity and Group (i.e. bilinguals vs monolinguals).

3.9.Pencil-and-paper interpretation task

This task is a partial replication of Yuan (1998)'s study, which compared how Mandarin monolinguals and L2 learners of Mandarin interpreted *ziji* under different conditions. This task required the participants to read 18 short sentences, and decide to which antecedent *ziji* refers to (the LD or local antecedent, both, or neither). The stimulus sentences were created based on Yuan (1998)'s stimuli using the syntactic structures which were the same to those of Yuan's Type 1-6 sentences (p. 330), but HSK-3 vocabulary were used instead to ensure full comprehensibility. Yuan's study also used another five types of sentences (see pp. 330-331) to investigate whether the interpretation of *ziji* are constrained by the subject orientation property of *ziji*, but these five types of sentences were not included in this study, because we are mainly interested in whether the bilinguals and the monolinguals would show similar preferences for interpreting *ziji* when *ziji* is syntactically ambiguous but semantically/pragmatically neutral or biased to an LD/local interpretation.

In this task, *ziji* appeared in two types of clauses: embedded finite clauses, and infinitive clauses. Both types of syntactic structures allow both LOC and LD binding of *ziji*. For each syntactic structure, three variations (Neutral, LOC favoured, LD favoured) were created by using the same lexical verbs, or verbs of similar meaning, used in Yuan (1998)'s study. For each variation, three different sentences were created, so there were 6 types of sentences, and 18 stimulus sentences in total. As mentioned earlier, these sentences were presented to the participants in together with the AJT task and another 9 fillers in a pseudo-randomized order. The full list of the stimulus sentences is presented in Appendix 2, and examples of the 6 types of sentences are given in Table 3.2:

Table 3.2. Exemplar sentences of the pencil-and-paper interpretation task

Type 1 *ziji* in embedded finite clause (Neutral)

Gao Lin	zhidao	Li Dong	feichang	xiangxin	ziji.	
Gao Lin	know	Li Dong	very	trust	self	

"Gao Lin knows that Li Dong trusts him/himself very much."

Type 2 *ziji* in embedded finite clause (LD favoured)

Wang Ming	bu-gaoxing	de	shuo	Li Dong		
Wang Ming	unhappily	Adv-P	say	Li Dong		
jingchang	bu	xiangxin	ziji.			
often	not	trust	self			

"Wang Ming said unhappily that Li Dong often does not trust him/himself."

Type 3 *ziji* in embedded finite clause (LOC favoured)

Li Dong	jide	Wang Laoshi	diyi-ci	lai		
Li Dong	remember	Teacher Wang	first time	come		
shang-ke	de-shihou	meiyou	jieshao	ziji.		
teach-class	when	not	introduce	self		

"Li Dong remembers that Teacher Wang didn't introduce him/himself when he came to teach the class for the first time."

Type 4 *ziji* in infinitive clause (Neutral)

Wang Ping	rang	Li Dong	buyao	piping	ziji.	
Wang Ping	ask	Li Dong	not	criticize	self	

"Wang Ping asked Li Dong not to criticize him/himself."

Type 5 *ziji* in infinitive clause (LD favoured)

Gao Hong	bu-yuanyi	gen	bieren	jianghua,	suoyi	ta
Gao Hong	not-like	with	others	speak	therefore	she
bu-yuanyi	ta	mama	xiang	bieren	jieshao	ziji.
not-like	her	mother	to	others	introduce	self

"Gao Hong does not like to talk to other people. Therefore, she does not like her mother to introduce her/herself to other people."

Type 6 *ziji* in infinitive clause (LOC favoured)

Li Jiaoshou	rang	Zhang Ping	yange	yaoqiu	ziji,
Professor Li	ask	Zhang Ping	strict	require	self
buyao	zongshi	qu	ti	zuqiu.	
not	always	go	play	football	

"Professor Li asked Zhang Ping to set strict demands on him/himself, and not to play football all the time."

Although Yuan (1998) claimed that the above six types of sentences were manipulated to create neutral, LD favoured or LOC favoured readings, he did not clearly introduce how this was achieved. In order to confirm that these sentences actually have different preferred readings, we piloted these sentences with three Mandarin-English speakers with less than 1 year's exposure to L2 English, and the results suggest that these pilot participants responded differently to sentences of different types. For each variation, there were 18 responses, and the results are summarized in Table 3.3. As it can be seen, the pilot participants showed obvious differences in interpreting *ziji* in sentences of different types. For the LD and LOC favoured sentences, there was a very strong preference for interpreting *ziji* as referring to either the LD or the LOC antecedents. By contrast, although there was a preference for an LD interpretation under the Neutral condition, a higher level of variation can be observed in the responses to these sentences. Therefore, we assume that the distinctions between these sentence types are valid, and will use them in the actual study.

Table 3.3. Pilot data for the pencil-and-paper interpretation task

	BOTH	LD	LOC	Total
Neutral	6	10	2	18
LD favoured	0	18	0	18
LOC favoured	2	1	15	18

As the interpretation of *ziji* in this task clearly involves syntax, semantics and pragmatics, and is assumed to be at the "external" interface by the IH (Sorace, 2011), we expect that the bilinguals would differ from the monolinguals in performing this task. Furthermore, if the bilinguals showed L1 attrition effects as a result of influence from L2 English, they should show a higher level of preference for LOC interpretations than the monolinguals across conditions.

3.10. Visual world eye-tracking task

In this visual world eye-tracking task, the participants first heard a sentence containing *ziji* and saw four pictures at the same time; the four pictures included an object representing the local antecedent, another one representing the LD antecedent, as well as two distractors. After hearing the sentence, an interpretation question appeared on the laptop screen and the participants needed to determine the antecedent of *ziji* (LD antecedent, local antecedent, or both). The participants' eye-movements were recorded when they were listening to the sentences.

Three types of target sentences (Neutral, LD favoured, LOC favoured) were created by manipulating the animacy of the matrix and the local antecedents. All these sentences followed the same syntactic structure of [NP1+V1+[NP2+V2+*ziji*]]+Co-ordinate clause. The co-ordinate clause was used to create a spill-over region, and it was always *zhejian shi shi zhen de* "and this is true". This design allows us to examine if there would be a locality effect during the syntactic processing of *ziji* in our participants by looking into the eye movements, as well whether our participants' final interpretation of *ziji* are affected by the animacy of the antecedents by looking into the results of the interpretation task. The examples of the critical sentences are given in Table 3.4, followed by Figure 3.2 illustrating the procedures of doing this task:

Table 3.4. Exemplar sentences of the visual world eye-tracking

Neutral sentences: animate LD antecedent + animate local antecedent

Laoshi$_i$	shuo	baba$_j$	jieshao	le	ziji$_{i/j}$,	[…]
Teacher$_i$	say	father$_j$	introduce	PERF	self$_{i/j}$	[…]

"The teacher said that the father had introduced him/himself, …"

LD favoured sentences: animate LD antecedent + inanimate local antecedent

Laoshi$_i$	shuo	dianshi$_j$	jieshao	le	ziji$_{i/*j}$,	[…]
Teacher$_i$	say	TV$_j$	introduce	PERF	self$_{i/*j}$	[…]

"The teacher said that the TV programme had introduced him, …"

LOC favoured sentences: inanimate LD antecedent + animate local antecedent

Dianshi$_i$	shuo	laoshi$_j$	jieshao	le	ziji$_{*i/j}$,	[…]
TV$_i$	say	teacher$_j$	introduce	PERF	self$_{*i/j}$	[…]

"The TV programmed said that the teacher had introduced himself, …"

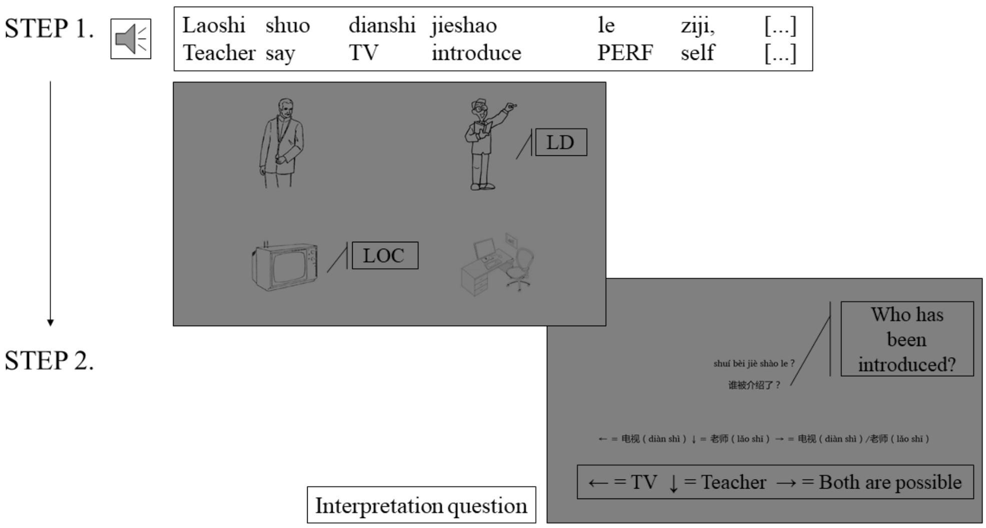


Figure 3.2. Procedures of the visual world eye-tracking task

The stimuli for this task were presented following a Latin-square design. First, 15 sentences were created following the syntactic structure mentioned above. Then, based on each of these sentences, three versions were created (LD preferred, LOC preferred and AMB binding, see the examples above). Following that, three lists are created; for each of the 15 sentences, only one version was presented once in each list. At the end, 5 practice items and 15 fillers of various structures were created and added into each list, and each list was pseudo-randomized. During the experiment, only one of the three lists was presented to each participant, either in the original or a reversed order, so each participant sees 5 LD binding, 5 local binding and 5 ambiguous sentences in a pseudo-randomized order.

Following the IH, we predict that the bilinguals and the monolinguals would not differ in the early syntactic processing of *ziji*, but they may show a difference in on-line processing at later stages, because assumingly these stages would involve the processing of semantics and pragmatics. We also predict that the bilinguals would differ from the monolinguals in performing the interpretation task embedded in this eye-tracking task, because the final resolution of *ziji* depends not only on syntax, but also semantics and pragmatics.

3.11. Procedures

This experiment was conducted in a one-to-one setting. The participants first completed the visual world eye-tracking task and then the sentence-picture matching task on the researcher's laptop. Each task consisted of two blocks, with each block containing half of the testing items. The participant was allowed to have a short break between blocks. The eye-tracker was calibrated before starting each block.

For both tasks, the participants were instructed to listen to each sentence for comprehension, and answer the question after each sentence by pressing the appropriate key on a keyboard. They were also instructed to focus their sight on the cross which appears at the centre of the laptop screen before the start of each sentence. All the participants were asked to keep their head as stable as possible, and to respond to the questions as quickly and accurately as possible.

After finishing these two processing tasks, the participant completed the digits-back recall WM test. They were instructed to memorize the heard digits and speak them out in reversed order, and only look at the cross symbol on the laptop screen during this task. If the digits were correctly recalled, the researcher would press the right arrow on the keyboard and the experiment programme would record this answer as “correct”, and proceed to the play the next set of digits; if the digits were incorrectly recalled, the researcher would press the left arrow key and the programme would record the answer as “incorrect” and proceed to the next set of digits. The practice session was repeated until the participants made less than two errors.

Then, the participant completes the cloze test and the HSK-3 test. The instructions for these two tasks were provided on the laptop before the practice sessions, and both tasks were completed by pressing the appropriate keys on a keyboard. Following that, the researcher asked the participants the questions in the questionnaire, and filled in the questionnaire. At the end, the participant completed the acceptability judgement task and the pencil-and-paper interpretation task. All these tasks were finished within 90 minutes.

4. Results and data analysis: Biodata and aspect marking

In this chapter, we summarize and analyze two types of data. The data were processed and analyzed using *R* (R Core Team, 2016), and the *lme4* and *lmerTest* packages were used for the analyses using linear mixed-effects model (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2016). First, in Section 4.1, we summarize the data collected using the language background questionnaire, abridged HSK-3 test, aiming to confirm that our participants have met the requirements of this study. Then, in Section 4.2, we provide a description of the results of the digits-back recall task, which measured the participants' WM capacity.

Following that, we summarize and analyze the results collected using the tasks concerning durative and perfective aspect marking. In Section 4.3, we look into whether the bilinguals and the monolinguals differed in judging the acceptability of the sentences containing *le/zhe* and verbs/predicates of different lexical aspects, and compare whether both groups had monolingual-like receptive knowledge about durative/perfective aspect marking. In Section 4.4, we compare the bilinguals' performance in the cloze task to the monolinguals', and analyze whether there is any between-group difference in producing *le* and *zhe* in written froms.

Section 4.5 looks into the processing of durative and perfective aspect marking, and it examines if the bilinguals and the monolinguals differed in the accuracy and/or RTs for the sentence-picture matching task, as well as if there was a facilitation effect for both groups. It also examines whether individual differences in WM capacity constrained the performance in this task. As this study also concerns whether length of L2 exposure affects the perception, production and processing of durative and perfective aspect marking, we will also examine if there was a significant Length of

L2 Exposure effect on the bilinguals' performance in the tasks concerning aspect marking.

Section 4.6 summarizes this chapter, and briefly discusses if any observed bilingual/monolingual difference should be taken as a sign of L1 attrition, or other effects brought by extensive L2 exposure. It also points out what issues will be further discussed in this thesis.

4.1.Language background questionnaire and HSK-3 test

The results of the language background questionnaires and the abridged HSK-3 tests are summarized in Table 4.1 and 4.2:

Table 4.1. Means (SDs) of the monolinguals' and the bilinguals' age, onset age of acquiring Mandarin/English, age of arrival in the UK, and length of L2 exposure

	Monolinguals (*n*=23)	**Bilinguals (*n*=14)**
Age	32.65 (8.03)	33.07 (13.13)
Onset age of acquiring Mandarin	0.00 (0.00)	0.00 (0.00)
Onset age of learning English in classrooms	10.86 (1.30)	12.07 (4.46)
Age of arrival in the UK	N/A	19.14 (7.13)
Length of exposure to L2 English in the UK (years)	0.00 (0.00)	13.00 (7.46)

As seen in Table 4.1, the language background of the bilinguals fulfilled the requirements of this study - these bilinguals arrived at the UK after puberty, and had been exposed to L2 English in an English-dominant environment for more than 7 years. Moreover, the bilinguals and the monolinguals were similar in terms of mean age and mean onset age of learning English in classrooms. It should be noted that, although both groups started learning English in classroom settings at similar ages, only the bilinguals have experienced extensive exposure to L2 English in an English-dominant environment. Now we turn to Table 4.2, which summarizes the participants' self-report Mandarin and English proficiency, and their scores for the abridged HSK-3 test.

Table 4.2. Mean scores (SDs) of the monolinguals' and the bilinguals' self-reported Mandarin/English proficiency (min = 1, max = 5) and HSK3 listening/reading/total score

	Monolinguals (*n*=23)	**Bilinguals (*n*=14)**
Self-reported overall Mandarin proficiency	5.00 (0.00)	4.86 (0.36)
Self-reported proficiency in reading Mandarin	5.00 (0.00)	4.86 (0.36)
Self-reported proficiency in speaking Mandarin	5.00 (0.00)	4.86 (0.36)
Self-reported proficiency in listening Mandarin	5.00 (0.00)	4.86 (0.36)
Self-reported proficiency in writing Mandarin	5.00 (0.00)	4.21 (1.05)
Self-reported overall English proficiency	1.70 (0.63)	4.21 (0.43)
Self-reported proficiency in reading English	1.96 (0.88)	3.86 (0.66)
Self-reported proficiency in speaking English	1.74 (0.62)	4.14 (0.77)
Self-reported proficiency in listening English	1.87 (0.76)	4.00 (0.68)
Self-reported proficiency in writing English	1.78 (0.67)	3.93 (0.92)
HSK-3 listening (in percentage)	96.52 (5.73)	94.29 (11.58)
HSK-3 reading (in percentage)	100.00 (0.00)	98.57 (5.35)
HSK-3 total (in percentage)	98.26 (2.86)	96.43 (6.02)

In comparison to the monolinguals, the bilinguals reported slightly lower mean scores for overall Mandarin proficiency and all the four Mandarin skills, but the mean scores were all above 4 (4 = “very good”), and the minimum score for the self-reported proficiency in reading, speaking and listening Mandarin was 4. The bilinguals tended to report a lower score for the proficiency in writing Mandarin (range: 2-5), but the informal talk with the bilinguals during the questionnaire task suggested that these bilinguals were not confident in formal Mandarin writing (e.g. writing business reports or academic essays), rather than producing grammatical sentences in written Mandarin.

As the monolingual group did not show any variance in these measures, we have not been able to use inferential statistical tests to compare the two groups’ self-reported Mandarin proficiency. However, we conducted a series of *t*-tests to compare their HSK-3 listening and total scores, and did not observe any significant difference. This finding suggests that these two groups did not actually differ in their Mandarin listening skills or overall literacy skills. Again, due to that the monolinguals did not show any variance in the HSK-3 reading test, we have not been able to use any inferential statistical tests to compare the bilinguals and the monolinguals. However, as the bilinguals also showed high level of accuracy in performing the reading test, it is reasonable to conclude that they also have sufficient reading skills for this study. Furthermore, we conducted a series of *t*-tests to compare the two groups’ self-report scores for general English proficiency and the four English skills, and found that the bilinguals reported significantly higher scores for all of these measures ($t_{\text{EngProficiency}}$ = 14.43, $t_{\text{EngReading}}$ = 7.46, $t_{\text{EngSpeaking}}$ = 9.89, $t_{\text{EngListening}}$ = 8.85, $t_{\text{EngWriting}}$ = 7.60, all ps < .001).

Among all the participants, no monolingual reported knowing a language other than Mandarin, but five bilinguals reported that they had learned or were learning languages other than Mandarin and English. The languages include French (three), German (one), Spanish and Japanese (one). However, none of these bilinguals regarded themselves as proficient speakers of these languages - two reported that they could have basic conversations in French but rarely used or heard French, and the other three regarded themselves as beginner learners of French, German, or Spanish and Japanese. Due to their relatively low proficiency in these languages, we assume that these languages have minimal influence on their L1 Mandarin.

4.2.Digits-back recall task

The digits-back recall task measures the participants' central executive WM capacity, and the scores for this task are calculated in the following way: for each set of digits, the participants receive a score of 1 for correctly recalling it, and a score of 0 for failing to do so, regardless the length of digits. Following that, each participant's scores are added up and then divided by the total number of digit sets (n = 15), generating the final score for this task.

Conway et al. (2005) argues that, rather than the "all-or-nothing unit scoring" method which we adopted in this study, the "partial-credit scoring" methods might better reflect the participants' WM capacity. When using the "partial-credit" scoring methods, a participant will receive a score for correctly recalling each individual element in a set of digits. Therefore, if a participant correctly recalled three numbers for a trial containing five numbers, s/he will receive a score of 0.6 if the "partial-credit" scoring method is adopted, and a score of 0 if the "all-or-nothing" scoring method is adopted. However, the "partial-credit" scoring methods require the recordings of the participants' responses, and such data are not available in this study; thus we chose the "all-or-nothing" scoring method for this analysis, as Conway et al.

(2005) suggests that this method can also generate scores that highly correlate with the participants' actual WM capacity.

Table 4.3. Mean scores (SDs) of the digits-back recall task

	Monolinguals (*n*=23)	**Bilinguals (*n*=13)**
WM score	0.62 (0.16)	0.67 (0.17)

The mean scores and SDs for the digits-back recall task are summarized in Table 4.3. Due to programme error, one bilingual speaker's score for this task was missing and excluded from this analysis. A *t*-test did not find any significant difference between these two groups. As discussed in Chapter 2, individual differences in WM capacity might play a role in explaining variations in L1 attrition, and should affect performance in tasks requiring simultaneous storage and processing, we will include the digits-back recall task scores as a predictor representing WM capacity in further analyses which we assume WM capacity had played a role, e.g. the processing of aspect marking and the processing of the reflexive *ziji*.

4.3.Acceptability judgement task

In this and the following sections, we will analyze the results from the tasks concerning aspect marking. We will start with the off-line tasks, i.e. the AJT task and the cloze task. Then we will analyze the accuracy and RT results from the sentence-picture matching task, which concerns the processing of perfective and durative aspect marking.

In Section 3.6, we predicted that the bilinguals would resemble the monolinguals in this task, and both groups should accept the ACC+*le/zhe*, ACT+*le/zhe*, SLS+*le/zhe* and SEM+*le/zhe* sentences, and reject the ILS+*le* and ACH+*le* sentences. The mean acceptability scores (rating scale: -2 to 2; -2 = “completely unacceptable”, 2 = “completely acceptable”) and SDs for the sentences containing *le/zhe* and verbs of different lexical aspect are summarized in Table 4.4 on the next page. Among all the 1332 responses from the 23 monolinguals and the 14 bilinguals, 6 responses were missing and excluded from this analysis. As can be seen in Table 4.4, on most occasions, the bilinguals and the monolinguals were similar in judging the acceptability of the target sentences. In general, the performance of both groups was consistent with our prediction, as they all tended to accept the sentences containing ACC+*le/zhe*, ACT+*le/zhe*, SLS+*le/zhe*, SEM+*le/zhe*, ILS+*le* and ACH+*le,* while rejecting the ILS+*zhe* and ACH+*zhe* sentences.

Table 4.4. Mean acceptability scores (SDs) for the target sentences containing le/zhe + verbs of different lexical aspect

	ACC+le	**ACC+zhe**	**ACH+le**	**ACH+zhe**	**ACT+le**	**ACT+zhe**
Bilinguals	2.00 (0.00)	1.17 (0.95)	1.81 (0.31)	-1.88 (0.21)	1.62 (0.89)	1.48 (0.89)
Monolinguals	1.93 (0.28)	0.88 (1.02)	1.84 (0.39)	-1.74 (0.45)	1.87 (0.28)	1.43 (0.87)
	ILS+le	**ILS+zhe**	**SEM+le**	**SEM+zhe**	**SLS+le**	**SLS+zhe**
Bilinguals	0.79 (0.81)	-1.62 (0.54)	0.71 (0.94)	0.79 (0.82)	0.86 (0.50)	1.07 (1.02)
Monolinguals	1.09 (0.82)	-1.00 (1.03)	1.03 (0.63)	0.97 (0.92)	1.04 (0.51)	1.51 (0.66)

ACC = Accomplishment, ACH = Achievement, ACT = Activity,

SEM = Semelfactive, ILS = Individual-level state, SLS = State-level state

For each type of verb/predicate+*le* and verb/predicate+*zhe* combination, we used a series of linear mixed-effects models to examine if there was any statistically significant between-group difference in the acceptability scores. Before constructing the models, the acceptability scores were transformed into *z*-scores. In these models, Group was used as the predictor (bilinguals vs monolinguals; monolinguals coded as 0.5, bilinguals as -0.5), and Subject and Item (i.e. each target sentence) were treated as the random factors. Both Subject and Item had random intercepts, and Item had random slopes for the fixed effect of Group. As this task was self-paced and the performance of this task was unlikely to be affected by WM (Staum Casasanto, Hofmeister, & Sag, 2010), WM Capacity was not included as a predictor.

This analysis only revealed a marginally significant Group effect on the acceptability scores for the ILS+*zhe* sentences (estimate = 0.39, $SE = 0.21$, $t = 1.89$, $p = 0.07$), but no significant Group effect for sentences of other types (see Appendix 1 for full results).

The marginally significant effect of Group on the acceptability scores for the ILS+*zhe* sentences suggests that the bilinguals were more likely to reject the ungrammatical ILS+*zhe* sentences than the monolinguals. However, this effect does not indicate that the bilinguals had non-monolingual-like grammatical knowledge about the interaction between ILSs and *zhe*, because both groups tended to reject the ungrammatical ILS+*zhe* sentences. The high level of variation within the monolingual group (as indicated by the SD) suggests that this group might have treated one of the ILS+*zhe* sentences differently, or that some participants in the monolingual group behaved differently from the others in rating these sentences.

We examined each monolingual participant's responses, and found two participants consistently accepting all the three ILS+*zhe* sentences, as well as three participants accepted two of the three ILS+*zhe* sentences. There are at least two possible explanations for such behaviour. First, it may be the case that some monolinguals mistook *zhe* for an intensifier, such as *zhe ne*, and accepted the ungrammatical ILS+*zhe* sentences (see Li & Thompson, 1989, p. 222 for a discussion about *zhe ne*). Second, it may be caused by the fact that some monolinguals had different dialect background, and tended to treat ILS+*zhe* differently. Unfortunately, since we do not have their dialect background, it is impossible to do any further analysis.

As our analysis did not reveal any significant between-group difference suggesting changes to L1 knowledge about the interaction between *le/zhe* and different lexical aspects among the bilinguals, we assume that the bilinguals did not differ from the monolinguals in terms of receptive knowledge about this language structure. This finding is consistent with the predictions of the IH (see Section 2.2.1 and 2.8), as the structure we examined here was supposed to be at the "internal" interfaces.

In addition to the comparison between bilinguals and monolinguals, we are also interested in whether Length of L2 Exposure would affect the bilinguals' perception about the interaction between lexical aspect and durative/perfective aspect marking (see the discussion in Section 2.3). Therefore, we constructed another series of linear mixed-effects model for each type of verb/predicate+*le* and verb/predicate+*zhe* combination. In these models, Length of L2 Exposure was included as the predictor, and Subject and Item as the random factors. Length of L2 Exposure was centered by subtracting the mean Length of L2 Exposure from each Length of L2 Exposure value.

Both Subject and Item had random intercepts, and Item had random slopes for the fixed effect of Length of L2 Exposure. The ACC+*le* sentences were excluded from this analysis because of no variance in the acceptability score. As the results suggest, the models only found a significant fixed effect of Length of L2 Exposure on the acceptability scores for the ACT+*le* (estimate = -0.04, *SE* = 0.02, t = -2.18, p = 0.05) and a marginally significant effect of Length of L2 Exposure on the SEM+*le* sentences (estimate = -0.06, *SE* = 0.03, t = -1.87, p = 0.08), but not on any other types of verb/predicate+*le/zhe* combination (see Appendix 1 for full results).

The Length of L2 Exposure effect for the ACT+*le* sentences suggests that, as the Length of L2 Exposure increases, the acceptability score for the ACT+*le* decreases (see Figure 4.1). However, this significant fixed effect was caused by an outlier, as a further look into each individual bilingual's data revealed that, only one bilingual with 29 years of exposure to L2 English consistently rejected the grammatical ACT+*le* sentences. After removing this bilingual's data, the fixed effect of Length of L2 Exposure becomes non-significant.

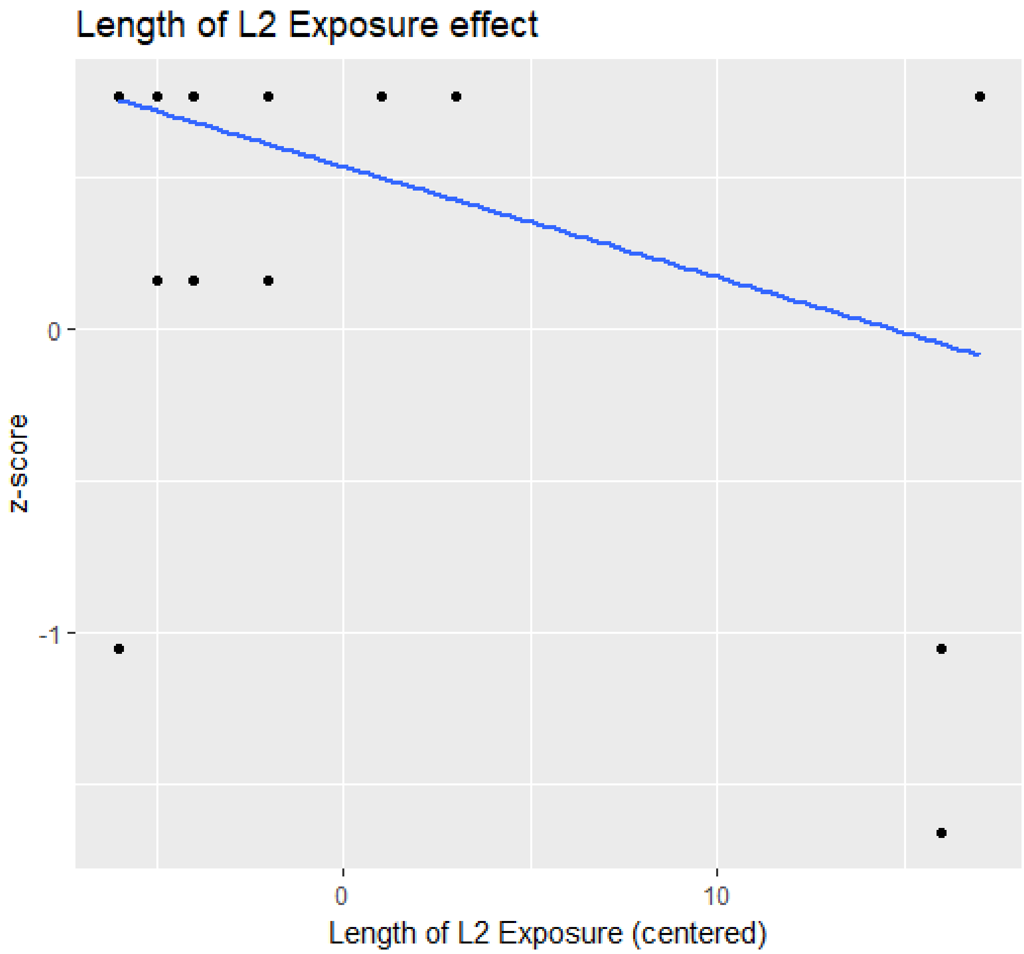


Figure 4.1. Length of L2 Exposure effect on the acceptability score for ACT+le sentences among the bilinguals

The Length of L2 Exposure effect on the acceptability score for SEM+*le* sentences suggests that the bilinguals tended to give lower scores as Length of L2 Exposure increases (see Figure 4.2), but again this effect was caused by the same bilingual who consistently rejected the ACT+*le* sentences. After removing this bilingual's data, this effect becomes non-significant.

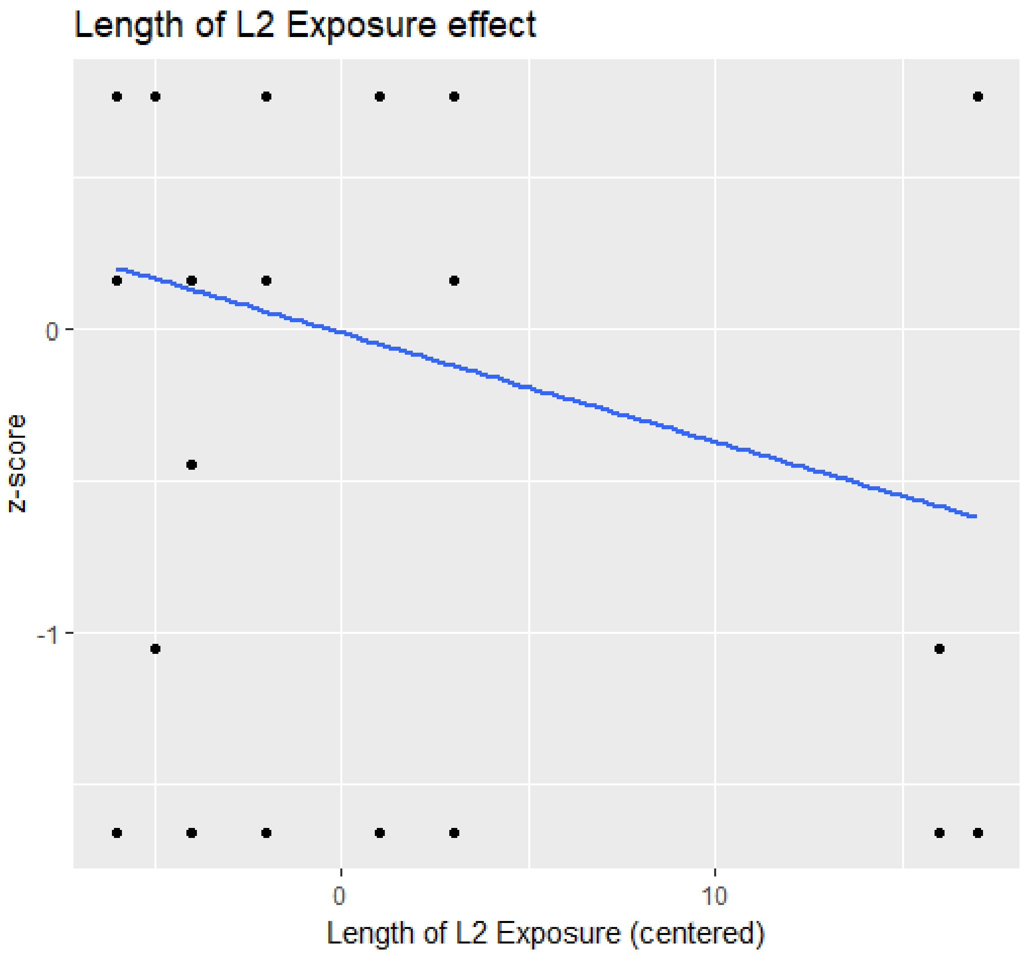


Figure 4.2. Length of L2 Exposure effect on the acceptability score for SEM+le sentences among the bilinguals

As the observed Length of L2 Exposure effect was caused by an individual bilingual who consistently rejected the ACT+*le* and SEM+*le* sentences, we cannot interpret our findings as Length of L2 Exposure had an effect on perceiving the interaction between *le* and these two types of lexical aspects. Although this bilingual had a long period of L2 exposure, we cannot interpret this bilingual's unexpected performance as indicating L1 attrition in perceiving the interaction between *le* and ACT verbs/predicates would emerge after a long period of L2 exposure, because this bilingual did not really have problems with understanding or using ACT+*le* sentences, and s/he reached monolingual-like accuracy in supplying *le* in the cloze test and correctly responding to the ACT+*le* sentences in the sentence-picture matching task. Unfortunately, as we did not further test the production or processing of SEM+*le* sentences, we cannot suggest how this particular bilingual's performance should be interpreted. We will further discuss these issues in Chapter 6.

Overall, our analysis suggests that the bilinguals did not differ from the monolinguals in performing this AJT task. Therefore, we assume that the bilinguals had monolingual-like receptive knowledge about the interaction between *le/zhe* and different lexical aspects, at least when simple declarative sentences were concerned. As noted earlier in this section, this finding is consistent with the IH. Moreover, we did not find evidence supporting the idea that Length of L2 Exposure might explain the variations in perceiving the interaction between *le/zhe* and different lexical aspects within the bilingual group.

4.4.Cloze test

In this section, we analyze the results of the cloze test, which assesses the participants' productive knowledge of *le* and *zhe*. The participants' mean accuracy in supplying *le* and *zhe* for the cloze task is summarized in Table 4.5. Three monolinguals' data were missing due to programme errors, and they were excluded from the analysis. The responses to one blank requiring *zhe* was also excluded from this analysis, as a post-hoc examination found that this blank did not necessarily require *zhe*. As this task is self-paced and unpressured, individual differences in WM capacity should not constrain the performance in this task. Therefore, in this analysis, WM Capacity was not included as a predictor.

Table 4.5. Mean percentage of accuracy (SDs) in le/zhe production

	Monolinguals (*n* = 20)	**Bilinguals (*n* = 14)**
Accuracy for *le* production	92.50 (7.48)	89.29 (6.68)
Accuracy for *zhe* production	90.63 (13.98)	90.18 (11.16)

The descriptive data summarized in Table 4.5 suggest that both groups were highly accurate in supplying both types of aspect markers, and the bilinguals resembled the monolinguals in the accuracy in supplying *le* and *zhe*. A generalized linear mixed-effects model was used to confirm whether these differences are statistically significant. In this model, the responses to the blanks requiring *le/zhe* (incorrect responses coded as 0, and correct ones coded as 1) were treated as the dependent variable, Aspect Marker Type (*le* vs *zhe*; *le* coded as 0.5 and *zhe* as -0.5) and Group (bilinguals vs monolinguals; monolinguals coded as 0.5 and bilinguals as -0.5) are treated as the predictors. The random factors include Subject and Item. Both factors had random intercepts and random slopes - Subject had random slopes for the fixed effect of Aspect Marker Type, and Item had random slopes for the fixed effect of Group. The *t* value, model estimate, SE and *p* value are summarized in Table 4.6. Our model suggests no significant fixed effect of Aspect Marker Type or Group, nor is there a significant interaction between Aspect Marker and Group. Therefore, our data do not suggest any difference between the bilinguals and the monolinguals in producing *le* and *zhe* in written form in unpressured settings, and this is consistent with the IH, which predicts no between-group difference in this respect.

Table 4.6. Summary of the model concerning the Group effect on the accuracy in the cloze task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	2.91	0.40	7.29	<0.001
Type	-0.02	0.73	-0.02	0.98
Group	0.25	0.50	0.49	0.63
Type:Group	0.28	0.76	0.37	0.71

Another generalized linear mixed-effects model was used to examine whether Length of L2 Exposure affects the bilinguals' accuracy of producing *le* and *zhe*. The dependent variable and the random factors were the same as those used in the model which analyzes the effects of Aspect Marker Type and Group; the only difference was that the predictors in this model were Aspect Marker Type and Length of L2 Exposure. In this analysis, Length of L2 Exposure was centered by subtracting the mean Length of L2 Exposure from each Length of L2 Exposure value. As it can be seen in the results summarized in Table 4.7, this model did not find any significant fixed effect of Aspect Marker Type, Length of L2 Exposure, or Aspect Marker Type x Length of L2 Exposure interaction. Therefore, we did not find significant evidence of a Length of L2 Exposure effect on the production of perfective and durative aspect in written forms under an unpressured setting.

Table 4.7. Summary of the model concerning the Length of L2 Exposure effect on the accuracy in the cloze task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	2.54	0.42	6.04	<0.001
Type	-0.63	0.91	-0.09	0.93
L2Exp	-0.00	0.04	0.00	1.00
Type:L2Exp	0.06	0.07	0.86	0.39

Overall, the results of the cloze task do not suggest any significant between-group difference. Therefore, we assume that, under an unpressured setting, the bilinguals did not differ from the monolinguals in producing *le* and *zhe* in written forms when simple declarative sentences were concerned. This finding is also consistent with the IH, which predicts that the interactions between lexical and grammatical aspect marking in this scenario are not vulnerable to L1 attrition. It is also worth noting that, in this task, both groups showed a higher level of variance in supplying *zhe* than in supplying *le*. Such behaviour might be related to the phenomenon called “zero aspect marking”. As the bilinguals and the monolinguals demonstrated a similar level of “zero aspect marking” in this task, we do not consider this observation as a sign of L1 attrition, and assume that this issue is not directly relevant to our research questions. Therefore, we will not further discuss it here, but we will briefly look into this in Chapter 6 as it is related to research on Chinese linguistics. However, it should be noted that, a significant bilingual vs monolingual difference in the level of “zero aspect marking” could suggest L1 attrition, as observed in other studies (e.g. Shi, 2011).

4.5.Sentence-picture matching

This task assesses the participants' processing of the interaction between *le/zhe* and the ACT verbs, aiming to find out whether there was any L1 attrition effects which can only be revealed using processing tasks. In order to find out what factors might affect the processing the ACT+*le/zhe* sentences, we will first analyze the participants' accuracy in performing this task, and then the RTs. The participants' accuracy in responding to the two types of target sentences are summarized in Table 4.8[8]. As discussed earlier in Chapter 2 and 3, this task requires simultaneous storage and processing, so the performance in this task may be constrained by WM capacity. Therefore, WM Capacity was included as a predictor in this analysis.

Table 4.8. Percentages of mean accuracy (SDs) in the sentence-picture matching task

	Monolinguals (*n*=21)	**Bilinguals (*n*=13)**
Accuracy for ACTs+*le*	80.00 (20.25)	91.54 (14.63)
Accuracy for ACTs+*zhe*	98.10 (4.02)	99.23 (2.77)

As Table 4.8 suggests, both groups were less accurate in responding to the semantically-mismatched ACT+*le* sentences than in responding to the semantically-matched ACT+*zhe* sentences. Interestingly, the bilinguals seemed to outperform the

[8] For this task, the data of three outliers, including two monolinguals and one bilingual, are excluded from this analysis due to their extremely low accuracy in responding to the ACT+*le* sentences (below 20%, outside two SDs). This data processing procedure was also applied in the RT analysis of this task, but it did not apply to the other analyses, because all the data in the other tasks were within 2SDs.

monolinguals in accurately responding to sentences of both types. A generalized linear mixed-effects model was used to find out if the differences shown by the descriptive data were statistically significant. In this model, responses to the target sentences (incorrect responses coded as 0, correct ones coded as 1) were the dependent variable, and Aspect Marker Type (*le* vs *zhe*; *le* coded as 0.5 and *zhe* as -0.5), Group (bilinguals vs monolinguals; monolinguals coded as 0.5 and bilinguals as -0.5), and WM Capacity (as measured by the digits-back recall task) were the predictors. Subject and Item were included as the random factors. In this model, both Subject and Item had random intercepts; Subject had random slopes for the fixed effect of Aspect Marker Type, and Item had random slopes for the fixed effects of Group and WM Capacity.

This analysis revealed a significant fixed effect of Aspect Marker Type, as well as a significant fixed effect of WM Capacity, but no significant Group effect or any interaction between these predictors. The model summary can be found in Table 4.9. The fixed effect of Aspect Marker Type suggests that, for both groups, there was a facilitation effect which led to a higher accuracy in responding to the semantically-matched ACT+*zhe* sentences than in responding to the semantically-mismatched ACT+*le* sentences (Yap et al., 2009, see discussion in Chapter 2).

Table 4.9. Summary of the model concerning the Group effect on the accuracy in the sentence-picture matching task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	4.07	0.67	6.04	<0.001
Type	-3.10	1.33	-2.34	0.02
Group	-1.05	1.30	-0.81	0.42
WM	7.42	3.61	2.06	0.04
Type:Group	-0.56	2.56	-0.22	0.83
Type:WM	-6.91	6.91	-1.00	0.32
Group:WM	-2.64	7.15	-0.37	0.71
Type:Group:WM	-3.54	13.62	-0.26	0.80

In addition, the fixed effect of WM Capacity suggests that the participants with larger WM Capacity tended to be more accurate in responding to this task (see Figure 4.3), regardless of Aspect Marker Type and Group, so the observed difference in the descriptive data is explained by the participants' individual differences in WM Capacity, rather than their differences in the status of bilingualism.

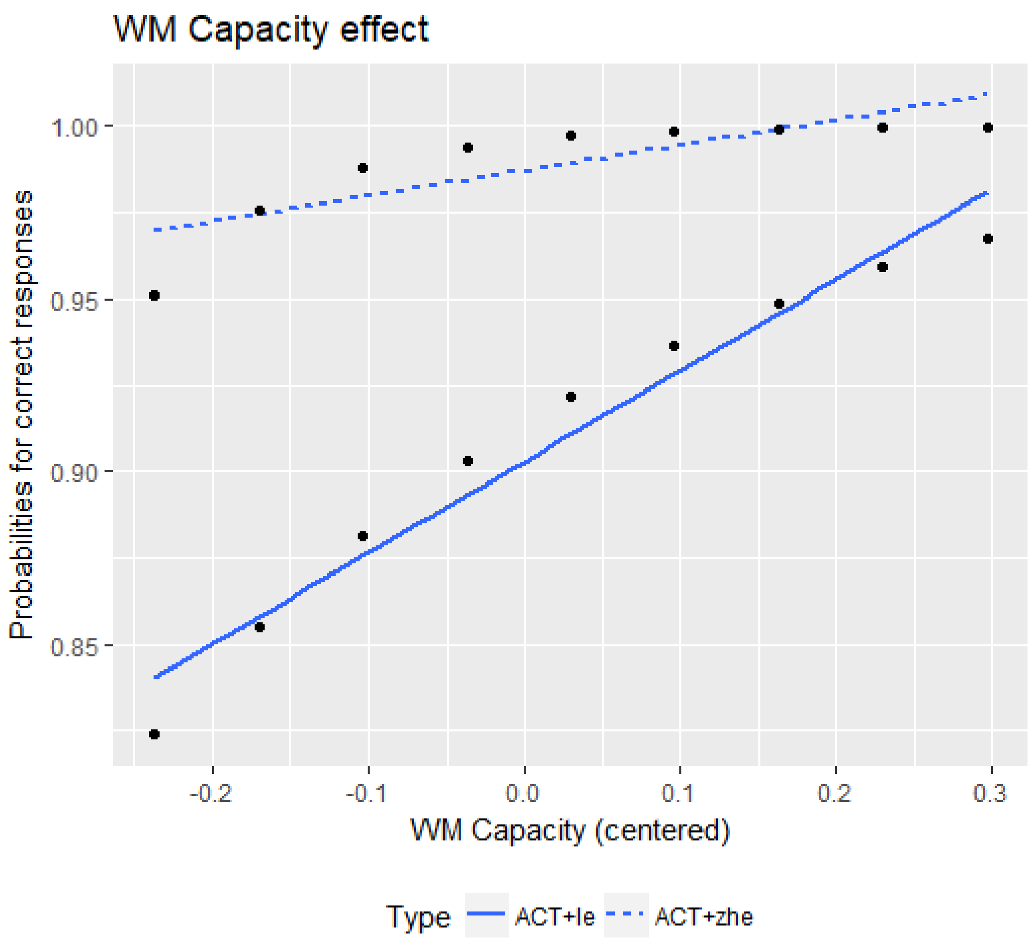


Figure 4.3. WM Capacity effect on the accuracy in responding to the sentence-picture matching task

Now we turn to the participants' RTs in responding to the target sentences. Following Yap et al. (2009)'s analysis, only those correctly responded items are included in this RT analysis. Table 4.10 summarizes the mean RTs and SDs in responding to the ACT+*le/zhe* sentences by Group. As it can be seen from the table, both groups were faster in responding to the semantically-matched ACT+*zhe* sentences than in responding to the semantically-mismatched ACT+*le* sentences. Compared to the monolinguals, the bilinguals seem to be slightly faster in responding to both types of sentences.

Table 4.10. Mean RTs (SDs; in seconds) of the sentence-picture matching task

	Monolinguals (*n*=21)	**Bilinguals (*n*=13)**
RTs for ACTs+*le* (s)	3.28 (0.55)	3.11 (0.30)
RTs for ACTs+*zhe* (s)	2.76 (0.20)	2.71 (0.25)

A linear mixed-effects model with RTs being the dependent variable, and Aspect Marker Type (*le* vs *zhe*), Group (bilinguals vs monolinguals) and WM Capacity being the predictors was used to find out if these differences are statistically significant. Like the previous model concerning accuracy, Subject and Item had random intercepts; Subject had random slopes for the fixed effect of Aspect Marker Type, and Item had random slopes for the fixed effects of Group and WM Capacity. This model, as shown in Table 4.11, suggests a significant fixed effect of Aspect Marker Type, but no significant fixed effect of Group, WM Capacity or any interaction between the predictors. Therefore, our findings suggest that both groups enjoyed a facilitation effect as they were faster in responding to the ACT+*zhe* sentences than to the ACT+*le* sentences, and that these two groups did not significantly differ in the speed of responding to the target sentences.

Table 4.11. Summary of the model concerning the Group effect on the RTs in the sentence-picture matching task

	Estimate	**Std.Error**	***t***	***p***
(Intercept)	2.87	0.06	45.36	<0.001
Type	0.33	0.11	2.98	<0.01
Group	0.08	0.08	1.11	0.28
WM	0.16	0.23	0.70	0.49
Type:Group	-0.04	0.10	-0.44	0.66
Type:WM	-0.18	0.29	-0.63	0.53
Group:WM	-0.47	0.51	-0.92	0.36
Type:Group:WM	-1.14	0.71	-1.65	0.12

Regarding the question of whether Length of L2 Exposure affected the processing of ACT+*le/zhe* sentences in terms of accuracy and RTs, we performed another analysis within the bilingual group. In order to find out if there was any relationship between Length of L2 Exposure and Accuracy, we used a generalized linear mixed-effects model, which included responses to the target sentences (incorrect responses coded as 0 and correct ones as 1) as the dependent variable, and Aspect Marker Type, Length of L2 Exposure and WM Capacity as the predictors. A linear mixed-effects model which included RTs as the dependent variable and the same predictors were used to analyze the effect of Length of L2 Exposure on RTs. In both models, the random factors Subject and Item had random intercepts; Subject had random slopes for the fixed effect of Aspect Marker Type, and Item had random slopes for the fixed effects of Length of L2 Exposure and WM Capacity.

Due to the bilinguals' very high level of accuracy and low level of variance in responding to the ACT+*zhe* sentences, the model concerning accuracy always failed to converge if the responses to the ACT+*zhe* sentence were included as the dependent variable. Therefore, we excluded the responses to the ACT+*zhe* sentences from the dependent variable, and Aspect Marker Type from the predictors, and constructed another generalized linear mixed-effects model to examine if Length of L2 Exposure had any effect on the accuracy in responding to the ACT+*le* sentences. This model, as shown in Table 4.12, revealed no significant fixed effect of Length of L2 Exposure or WM Capacity, or any interaction between these predictors.

Table 4.12. Summary of the model concerning the Length of L2 Exposure effect on the accuracy in the sentence-picture matching task (ACT+le sentences only)

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	3.72	1.04	3.57	<0.001
L2Exp	-0.01	0.10	-0.15	0.88
WM	7.94	5.33	1.49	0.14
L2Exp:WM	0.88	1.18	0.74	0.46

Meanwhile, the model concerning RTs produced a significant fixed effect of Aspect Marker Type, but no significant effect of any other predictor or interaction between the predictors. The effect of Aspect Marker Type suggests that, within the bilingual group, there was still a facilitation effect for the ACT+*zhe* sentences in terms of RTs. However, this model does not provide any evidence suggesting a significant effect of Length of L2 Exposure on the RTs in responding to the target sentences. The summary of this model is presented in Table 4.13.

Table 4.13. Summary of the model concerning the Length of L2 Exposure effect on the RTs in the sentence-picture matching task

	Estimate	**Std.Error**	***t***	***p***
(Intercept)	2.85	0.06	45.06	<0.001
Type	0.38	0.10	3.69	<0.01
L2Exp	0.00	0.01	0.37	0.72
WM	0.59	0.34	1.74	0.11
Type:L2Exp	0.01	0.01	1.17	0.26
Type:WM	0.66	0.47	1.42	0.18
L2Exp:WM	0.12	0.08	1.54	0.15
Type:L2Exp:WM	0.14	0.10	1.36	0.20

In general, the analysis of the sentence-picture matching task did not reveal any between-group difference in performing the sentence-picture matching task. Rather, it suggests that the bilinguals and the monolinguals showed similar pattern and speed in processing the interaction between *le/zhe* and ACT verbs, and the variance observed in this task was explained by individual differences in WM capacity. The similarity between the bilinguals and the monolinguals is in line with the IH, which predicts no L1 attrition effect for the processing examined here. The WM Capacity effect is consistent with our discussion in Chapter 2, and it suggests that the performance in tasks like this one can be constrained by WM capacity. Furthermore, we did not find any evidence suggesting a Length of L2 Exposure effect on the bilinguals' accuracy or RTs in responding to this task.

4.6.Summary

This chapter summarized and analyzed the data collected from the language background questionnaire, the digits-back recall task and the tasks concerning perfective and durative aspect marking. As Section 4.1 and 4.2 have shown, the bilinguals in this study did not differ from the monolinguals in age, WM capacity, self-reported Mandarin proficiency and literacy skills, but they differed from the monolinguals in that they have high self-reported English proficiency and at least 7 years' exposure to L2 English. Furthermore, neither group of participants seem to have been influenced by an L2 other than English. Therefore, the profiles of these two groups suit our purpose of this study, which concerns how extensive exposure of L2 English may affect Mandarin-English bilinguals.

As shown in Section 4.3 and 4.4, we did not find any evidence suggesting L1 attrition in the AJT task or the cloze task. The analyses suggest that, in this study, there was little difference in perceiving or producing *le/zhe* between the monolinguals and the bilinguals. With respect to the processing of durative/perfective aspect marking, our analysis suggests that, for both groups, there was a facilitation effect on processing the semantically-matched ACT+*zhe* sentences in terms of accuracy and RTs, and that the accuracy in responding to the target sentences increases with larger WM Capacity. Meanwhile, our analysis also suggests that the bilinguals and the monolinguals did not differ in the pattern or speed of processing the target sentences. These findings are in line with the IH, which predicts no difference between the bilinguals and the monolinguals in this language structure at the "internal" interface.

It is worth noting that, in the sentence-picture matching task, the descriptive data seemed to suggest that there is a noticeable and unexpected "bilingual advantage" in the accuracy in responding to the semantically-mismatched ACT+*le* sentences between the bilinguals and the monolinguals. Without including WM Capacity as a predictor, the statistical analysis may have found a false significant Group effect and led to wrong conclusions about how bilingualism may affect the processing of aspect marking in Mandarin, rather than the potential effects of individual differences in WM capacity. In Chapter 6, we will further discuss this methodological issue and its implications for L1 attrition.

For all of the tasks concerning aspect marking, we also analyzed whether length of exposure to L2 English affected the perception, production and processing of durative and perfective aspect marking within the bilinguals. As our analysis suggests, in none of these tasks was there a real significant effect of Length of L2 Exposure, and this is consistent with our hypothesis that the perception/production/processing of aspect marking at this level do not necessarily deviate from monolingual norms as the length of L2 exposure increases (see discussion in Chapter 2).

In Chapter 6, we will further discuss why the findings from the tasks concerning aspect marking cannot be interpreted as L1 attrition. We will also discuss these findings could contribute to examining the existing theoretical frameworks of L1 attrition, L1 attrition research method and experimental linguistic research in general.

5. Results and data analysis: Reflexive binding

This chapter summarizes and analyzes the results of the pencil-and-paper interpretation task and the visual world eye-tracking task (see Chapter 3 for a description of these tasks), aiming to address the second research question of whether extensive exposure to L2 English might lead to L1 attrition in the interpretation and processing of the reflexive *ziji* among the Mandarin-English bilinguals, as well as whether the bilinguals' interpretation and processing of *ziji* would vary according to their length of L2 exposure. We also examine whether individual differences in WM capacity played a role in explaining the variations in the eye-tracking task and the interpretation task embedded in the eye-tracking task (see discussion in Section 2.3).

The first two sections concern the questions of whether the bilinguals would show L1 attrition in interpreting the reflexive *ziji*, as predicted by the IH. Section 5.1 reports and analyzes the results of the pencil-and-paper interpretation task, and examines if the bilinguals behaved monolingual-like in interpreting *ziji* when syntactic structures and lexical verbs were controlled. Section 5.2 reports and analyzes the results of the interpretation task embedded in a visual world eye-tracking task, and further examines if the bilinguals behaved monolingual-like in interpreting *ziji* when the animacy of the local and LD antecedents were controlled.

Section 5.3 looks into the eye movements and concerns the bilinguals' on-line processing of *ziji*. More specifically, it focuses the proportions of looks into the areas of interests (AOIs) containing the local antecedents[9], and examines whether the

---

[9] As will be described later, we treated the eye movement data outside the AOIs of local or LD antecedents as missing during data processing. After applying this

bilinguals would show monolingual-like processing pattern during the processing of *ziji*. More specifically, it concerns the issue of whether the bilinguals would resemble the monolinguals in showing a locality effect (i.e. search for the local antecedents regardless of conditions) during the early stage processing, as well as whether the bilinguals would deviate from the monolinguals in later stage processing.

5.1.Paper-and-pencil interpretation task

In this section, we analyze the results of the off-line reflexive interpretation task, and examine if the monolinguals and the bilinguals behaved similarly in interpreting *ziji* in sentences of different syntactic structures and context biases. Out of the 666 responses from the 23 monolinguals and the 14 bilinguals, 2 responses were missing and excluded from the analysis. The participants' responses to this task are summarized in Table 5.1 as percentages. It should be noted that, as none of the participants incorrectly interpreted *ziji* as referring to neither antecedents, the NEITHER option is not presented in the table. Following the predictions of the IH, we predicted that the bilinguals would differ from the monolinguals in this task, as the interpretation of *ziji* involves syntax, semantics and pragmatics (see discussion in Section 2.6 and 3.9).

---

procedure, in the time windows of interest, the proportions of looks into the LD antecedents always raise when the proportions of looks into the local antecedents decrease, and *vice versa*. Therefore, in a given time window of interest, the proportion of looks into the LD antecedents was always 1-(Proportions of looks into the local antecedents). Given this, it seems unnecessary to analyze the proportions of looks into the LD antecedents, as they can always be suggested by the proportions of looks into the local antecedents.

Table 5.1. Percentage of interpretations of coreference between ziji and an indicated antecedent in the off-line task

Condition	**Interpretation**	**Monolinguals (*n* = 23)**	**Bilinguals (*n* = 14)**
Type 1	LD	69.57	52.38
***ziji* in embedded finite clause**	LOC	2.90	0.00
(Neutral)	BOTH	27.54	47.62
Type 2	LD	65.22	65.85
***ziji* in embedded finite clause**	LOC	11.60	0.00
(LD favoured)	BOTH	21.74	34.15
Type 3	LD	7.25	9.52
***ziji* in embedded finite clause**	LOC	66.67	52.38
(LOC favoured)	BOTH	26.09	38.10
Type 4	LD	50.72	42.86
***ziji* in infinitive clause**	LOC	26.09	19.05
(Neutral)	BOTH	23.19	38.09
Type 5	LD	100.00	100.00
***ziji* in infinitive clause**	LOC	0.00	0.00
(LD favoured)	BOTH	0.00	0.00
Type 6	LD	0.00	0.00
***ziji* in infinitive clause**	LOC	98.53	95.24
(LOC favoured)	BOTH	1.47	4.76

As seen in Table 5.1, the bilinguals did not differ much from the monolinguals in interpreting *ziji* in this task. Both groups preferred an LD interpretation of *ziji* when this reflexive appeared in the sentences which favoured a Neutral or LD interpretation (i.e. Type 1, 2, 4 and 5), and a LOC interpretation when *ziji* appeared in the sentences which favoured an LOC interpretation (Type 4 and 6 sentences). Meanwhile, when *ziji* appeared in the infinitive clauses rather than in the embedded clauses, both groups were much more inclined to interpret the *ziji* as referring to the LD or LOC antecedents when biased contexts were present, and both groups were more inclined to interpret *ziji* as referring to LOC antecedents under the Neutral condition. Moreover, it is also worth noting that, compared to the monolinguals, the bilinguals seemed to be more inclined to interpret *ziji* as being able to refer to BOTH antecedents for Type 1-4 sentences.

In order to examine whether both groups showed the same preference in interpreting *ziji* under the same condition, we conducted a series of logistic regression tests for each possible interpretation in each type of sentences. Type 5 sentences were excluded from this analysis, because both groups unanimously interpreted *ziji* as referring to the LD antecedents under this condition. The effect of Group (bilinguals vs monolinguals) on the LOC interpretation in Type 1 and Type 2 sentences, and the LD interpretation in Type 6 sentences was not examined either, because one group or both groups did not choose such an interpretation at all.

Then, for Type 1-4 and 6 sentences, generalized linear mixed-effects models are constructed with possible Interpretations being the dependent variables, and Group (bilinguals vs monolinguals; monolinguals coded as 0.5 and bilinguals as -0.5) being the predictor. The random factors included Subject and Item. Both random factors had random intercepts, and Item had random slopes for the fixed effect of Group. The full models are summarized in Appendix 1.

This analysis only revealed a significant Group effect on the LOC interpretation of *ziji* in Type 6 sentences, which suggests that the bilinguals were less likely to choose a LOC interpretation than the monolinguals under this condition (estimate = 26.78, SE = 6.87, z = 3.90, $p < 0.001$). However, an examination of the raw data found that, under this condition, 67 out of the 68 responses from the monolinguals were LOC, and 40 out of the 42 responses by the bilinguals were LOC. Therefore, this significant effect of Group might be caused by the very low level of variance in the monolingual group, rather than a real between-group difference in interpreting *ziji*. A larger sample is necessary to examine this issue in future studies. Furthermore, although the descriptive data suggests that the bilinguals showed a higher level of preference for a BOTH interpretation of *ziji* in Type 1-4 sentences, our analysis found this difference non-significant.

The absence of significant Group effects on the interpretation of *ziji* in the target sentences indicates that the bilinguals were monolingual-like in interpreting *ziji* regardless of whether biased context was present. This finding does not support the predictions of the IH - the interpretation of *ziji* depends on syntactic, semantic and pragmatic information, so it should be at the "external" interface, and the bilinguals should be likely to show L1 attrition effects for this language structure. However, our failure to find L1 attrition might be due to that off-line tasks like this one were not sensitive enough to reveal any between-group difference, and this is why we implemented the on-line eye-tracking task. We will discuss the results of the on-line tasks later in the next two sections.

As we are also interested in whether Length of L2 Exposure would affect the bilinguals' interpretation of *ziji*, another series of logistic regression tests were conducted for each possible Interpretation in each types of sentences, with Length of L2 Exposure being the predictor. The random factors included Subject and Item. Both random factors had random intercepts, and Item had random slopes for the fixed effect of Length of L2 Exposure. These tests were conducted within the bilingual group, and the sentence types and the interpretations which were excluded from the group analysis were also excluded from this analysis. As the model summary in Appendix 1 suggests, we did not find any significant effect of Length of L2 Exposure for any possible interpretation in any types of sentences. Therefore, we assume that there was no evidence supporting the idea that the bilinguals' interpretation of *ziji* in this task would vary according to their length of L2 exposure.

Overall, the results of this pencil-and-paper interpretation task suggest that the bilinguals had monolingual-like preferences for interpreting *ziji* regardless of whether biased context was present. According to the IH, if there were any L1 attrition among the bilinguals, such attrition effects would be more likely to be observed in interpreting *ziji*, because it involves syntax, semantics and pragmatics and is at the "external" interface. However, as the results of this task did not reveal any L1 attrition effects, they cannot provide evidence for falsifying the IH. In the next section, we will look into the results of the interpretation task embedded in the visual world eye-tracking task, and examine if both groups would show similar preference of interpreting *ziji* when the animacy of the antecedents were controlled.

5.2.Visual world eye-tracking task: interpretation task

This interpretation task was part of the visual world eye-tracking task, and it only included embedded clauses, which had similar syntactic structures of Type 1-3 sentences in the pencil-and-paper interpretation task. However, while the antecedents in the pencil-and-paper interpretation were all animate, the animacy of the antecedents in this task were manipulated, so it allows us to examine if the bilinguals and the monolinguals would show different responses to the animacy constraints on interpreting *ziji*. It is worth noting that, as this task required the participants to recall the heard sentences when answering the interpretation questions, WM capacity might play a role in constraining the performance in this task. Therefore, the potential effect of WM Capacity is also examined here, to ensure that its effect does not confound the effect of Group. The participants' interpretation of *ziji* under the three different conditions are summarized in Table 5.2.

Table 5.2. Percentage of interpretations of coreference between ziji and an indicated antecedent in the visual world eye-tracking task

Condition	**Interpretation**	**Monolinguals (*n* = 23)**	**Bilinguals (*n* = 14)**
Neutral	LD	59.13	51.43
(LD animate, LOC animate)	LOC	28.70	15.71
	BOTH	12.17	32.86
LD favoured	LD	72.17	74.29
(LD animate, LOC inanimate)	LOC	13.91	8.57
	BOTH	13.91	17.14
LOC favoured	LD	45.22	30.00
(LD inanimate, LOC animate)	LOC	46.09	54.29
	BOTH	8.70	15.71

Similar to the results of the pencil-and-paper interpretation task, the two groups did not differ much from each other in interpreting *ziji* in this task. Both groups preferred an LD interpretation of *ziji* under the Neutral and the LD favoured conditions. However, the two groups' preference in interpreting *ziji* under the LOC favoured condition seemed to differ - while the bilinguals showed a preference for the LOC interpretation, the monolinguals demonstrated similar preference for the LD and the LOC interpretation[10]. Meanwhile, in comparison to the monolinguals, the

[10] This result might be caused by the design of our stimulus sentences, in which the verb *shuo* "say" always appeared immediately after the inanimate LD antecedents in

bilinguals seemed to be more inclined to interpret *ziji* as being able to refer to BOTH antecedents, especially for the Neutral sentences.

In order to examine whether the participants' interpretation of *ziji* varied according to Condition (Neutral vs LD favoured vs LOC favoured), Group (bilinguals vs monolinguals, monolinguals coded as 0.5 and bilinguals as -0.5) and WM Capacity, we performed a series of logistic regression tests using generalized linear mixed-effects models for each possible Interpretation (LD vs LOC vs BOTH). For these tests, LOC favoured was set as the reference level of Condition. The random factors included Subject and Item. Both random factors had random intercepts; Subject had random slopes for the fixed effect of Condition, and Item had random slopes for the fixed effects of Group and WM Capacity. In cases where the model failed to converge, the random slope parameter that accounted for the least amount of variance was removed and the model refitted until convergence was achieved.

Now we turn to how the predictors have affected each type of the interpretation of *ziji*. First, with respect to the probability of choosing an LD interpretation of *ziji*, there are significant fixed effects of Condition: Neutral and Condition: LD (see model summary in Table 5.3). This effect suggests that both groups of participants were

---

the matrix clauses. This verb would change the animacy of the LD antecedent - for example, for the phrase *yinhang shuo* "the bank says" to be acceptable, the inanimate NP *yinhang* "bank" has to have some degree of animacy, and this may increase its accessibility as a possible resolution of *ziji*. However, as the following statistical tests will show, this effect brought by *shuo* "say" does not really change our participants' preference for the LOC interpretation of *ziji* under the LOC favoured condition.

significantly more likely to choose an LD interpretation of *ziji* under the Neutral and LD favoured conditions than under the LOC favoured condition. Meanwhile, the absence of significant Group, WM Capacity or interaction between the predictors suggests that the probability of choosing an LD interpretation did not vary according to Group or WM Capacity.

Table 5.3. Summary of the model concerning the effect of Group on the probability of choosing an LD interpretation in the VW interpretation task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	-0.56	0.34	-1.62	0.10
ConditionNeutral	0.92	0.46	1.99	0.05
ConditionLD	1.94	0.50	3.85	<0.001
Group	0.70	0.49	1.43	0.15
WM	-0.24	1.63	-0.15	0.88
ConditionNeutral:Group	-0.35	0.63	-0.55	0.58
ConditionLD:Group	-0.78	0.68	-1.14	0.25
ConditionNeutral:WM	-0.59	2.10	-0.28	0.78
ConditionLD:WM	2.00	2.21	0.90	0.37
Group:WM	4.48	3.47	1.29	0.20
ConditionNeutral:Group:WM	0.12	4.53	0.03	0.98
ConditionLD:Group:WM	-1.06	4.72	-0.22	0.82

Then we looked at how the predictors might affect the probability of choosing a LOC interpretation of *ziji*. The model, as presented in Table 5.4, revealed significant fixed effects of Condition: Neutral and Condition: LD, as well as a significant interaction between the Condition: Neutral and Group, but no significant effect of any other predictors or interactions. The fixed effect of Condition: LD suggests that our participants were significantly less likely to choose an LOC interpretation of *ziji* under the LD favoured conditions than under the LOC favoured condition. Meanwhile, the fixed effect of Condition: Neutral, along with the significant interaction between Condition: Neutral and Group, suggest that, while both groups were less likely to choose a LOC interpretation under the Neutral condition than under the LOC favoured condition, the bilinguals were significantly less likely to choose a LOC interpretation than the monolinguals under the Neutral condition. This interaction seems to indicate an L1 attrition effect, and it will be further discussed in Chapter 6.

Table 5.4. Summary of the model concerning the effect of Group on the probability of choosing a LOC interpretation in the VW interpretation task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	-0.06	0.32	-0.20	0.84
ConditionNeutral	-1.63	0.43	-3.78	<0.001
ConditionLD	-2.25	0.46	-4.92	<0.001
Group	-0.33	0.50	-0.65	0.51
WM	-2.17	1.61	-1.35	0.18
ConditionNeutral:Group	1.44	0.59	2.43	0.02
ConditionLD:Group	0.74	0.65	1.14	0.26
ConditionNeutral:WM	0.36	1.92	0.19	0.85
ConditionLD:WM	0.81	2.10	0.38	0.70
Group:WM	-0.49	3.12	-0.16	0.87
ConditionNeutral:Group:WM	2.20	3.67	0.60	0.55
ConditionLD:Group:WM	-1.94	4.01	-0.48	0.63

Finally we looked at the effects of the predictors on the probability of choosing a BOTH interpretation of *ziji*. As our model summarized in Table 5.5 have shown, there was no significant fixed effect of any predictor or any interaction, so neither groups of participants were more likely to choose a BOTH interpretation under the Neutral or

LD favoured condition than under the LOC favoured condition. Therefore, despite the observed difference shown in Table 5.2, the bilinguals did not actually differ from the monolinguals in the probability of choosing a BOTH interpretation of *ziji*.

Table 5.5. Summary of the model concerning the effect of Group on the probability of choosing a BOTH interpretation in the VW interpretation task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	-3.67	0.91	-4.01	<0.001
ConditionNeutral	1.64	0.91	1.80	0.07
ConditionLD	1.56	0.92	1.69	0.09
Group	-0.18	1.16	-0.16	0.87
WM	4.75	3.46	1.37	0.17
ConditionNeutral:Group	-1.34	0.97	-1.38	0.17
ConditionLD:Group	-0.07	1.07	-0.06	0.95
ConditionNeutral:WM	-3.08	2.80	-1.10	0.27
ConditionLD:WM	-5.70	3.09	-1.84	0.07
Group:WM	-4.81	6.97	-0.69	0.49
ConditionNeutral:Group:WM	-0.79	5.93	-0.13	0.89
ConditionLD:Group:WM	2.00	6.49	0.31	0.76

In order to find out whether and how Length of L2 Exposure might affect the bilinguals' interpretation of *ziji* under different conditions in this task, another series of logistic regression tests using generalized linear mixed-effect models were performed for each type of Interpretation among the bilinguals. In these tests, Condition, Length of L2 Exposure and WM Capacity were included as the predictors; the reference level of Condition was set as LOC favoured. Subject and Item were included as the random factors. Both random factors had random intercepts; Subject had random slopes for the fixed effect of Condition, and Item had random slopes for the fixed effects of Length of L2 Exposure and WM Capacity. In cases where the model failed to converge, the random slope parameter that accounted for the least amount of variance was removed and the model refitted until convergence was achieved.

First, regarding the probability of choosing an LD interpretation of *ziji*, a significant fixed effect of Condition: LD has been observed, suggesting that the bilinguals were more likely to choose an LD interpretation under the LD favoured condition than under the LOC favoured condition. Meanwhile, there was no significant effect of any other predictors or interactions. The absence of significant Length of L2 Exposure or WM Capacity effect suggests that the preference for an LD interpretation of *ziji* might not vary according to these two predictors. The model is presented in Table 5.6.

Table 5.6. Summary of the model concerning the effect of Length of L2 Exposure on the probability of choosing an LD interpretation in the VW interpretation task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	-0.84	0.46	-1.84	0.07
ConditionNeutral	1.20	0.70	1.71	0.09
ConditionLD	2.24	0.72	3.09	<0.01
L2Exp	-0.11	0.07	-1.64	0.10
WM	-1.76	2.51	-0.70	0.48
ConditionNeutral:L2Exp	0.00	0.09	-0.05	0.96
ConditionLD:L2Exp	0.10	0.09	1.05	0.29
ConditionNeutral:WM	-1.69	3.91	-0.43	0.67
ConditionLD:WM	0.74	3.53	0.21	0.83
L2Exp:WM	0.36	0.68	0.53	0.59
ConditionNeutral:L2Exp:WM	-0.06	0.98	-0.06	0.95
ConditionLD:L2Exp:WM	-0.88	0.94	-0.94	0.35

Second, with respect to the probability of choosing a LOC interpretation of *ziji*, our model summarized in Table 5.7 only revealed significant fixed effects of Condition: Neutral and Condition: LD, but no significant effects of any other predictor or interactions. These findings suggest that the bilinguals were less likely to choose a LOC interpretation under the Neutral or LD favoured condition than under the LOC favoured condition, but the probability of choosing a LOC interpretation might not vary according to Length of L2 Exposure or WM Capacity.

Table 5.7. Summary of the model concerning the effect of Length of L2 Exposure on the probability of choosing a LOC interpretation in the VW interpretation task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	0.03	0.61	0.05	0.96
ConditionNeutral	-2.78	0.97	-2.86	<0.01
ConditionLD	-2.77	1.04	-2.66	<0.01
L2Exp	0.12	0.08	1.51	0.13
WM	-2.88	2.96	-0.97	0.33
ConditionNeutral:L2Exp	-0.09	0.10	-0.90	0.37
ConditionLD:L2Exp	-0.08	0.10	-0.76	0.45
ConditionNeutral:WM	-1.53	3.82	-0.40	0.69
ConditionLD:WM	2.48	4.30	0.58	0.56
L2Exp:WM	-1.10	0.80	-1.37	0.17
ConditionNeutral:L2Exp:WM	-0.23	1.18	-0.20	0.84
ConditionLD:L2Exp:WM	0.29	1.19	0.24	0.81

Finally, regarding the probability of choosing a BOTH interpretation of *ziji*, no significant effect of any predictor or interaction between the predictors has been found. Therefore, the bilinguals were not more likely to choose a BOTH interpretation under the Neutral or LD favoured condition than under the LOC favoured condition, and the probability of choosing a BOTH interpretation might not vary according to the bilinguals' Length of L2 Exposure or WM Capacity. The model summary is presented in Table 5.8.

Table 5.8. Summary of the model concerning the effect of Length of L2 Exposure on the probability of choosing a BOTH interpretation in the VW interpretation task

	Estimate	**Std.Error**	***z***	***p***
(Intercept)	-4.16	1.88	-2.21	0.03
ConditionNeutral	2.06	2.06	1.00	0.32
ConditionLD	1.89	1.95	0.97	0.33
L2Exp	-0.15	0.31	-0.48	0.63
WM	14.90	11.12	1.34	0.18
ConditionNeutral:L2Exp	0.30	0.31	0.96	0.34
ConditionLD:L2Exp	0.15	0.31	0.49	0.62
ConditionNeutral:WM	-8.87	11.15	-0.80	0.43
ConditionLD:WM	-13.09	10.92	-1.20	0.23
L2Exp:WM	3.39	3.44	0.99	0.32
ConditionNeutral:L2Exp:WM	-2.96	3.46	-0.86	0.39
ConditionLD:L2Exp:WM	-2.18	3.42	-0.64	0.52

Overall, the above analysis provides evidence that the bilinguals largely resembled the monolinguals in their preferences for interpreting *ziji*, especially under the Neutral and LD favoured conditions. However, we did observe a significant between-group difference, which was that the bilinguals were significantly less likely to choose a LOC interpretation than the monolinguals under the Neutral condition. In Chapter 6, we will argue why this between-group difference should be interpreted as an L1 attrition effect. More importantly, in contrast to our hypothesis that individual differences in WM capacity would play a role in explaining variations in L1 attrition (see Section 2.3 and 2.8), our analysis did not find evidence suggesting that the bilinguals' preference for choosing a LOC interpretation of *ziji* under the Neutral condition was explained by the effect of WM Capacity or WM Capacity x Length of L2 Exposure interaction. This issue will also be further discussed in Chapter 6.

5.3. Visual world eye-tracking task: eye movements

In this section, we focus on the on-line processing of *ziji* and look into the participants' eye movements during the visual world eye-tracking task, and examine if there was any between-group difference which may suggest an L1 attrition effect and which could not be detected using off-line tasks (e.g. the interpretation tasks). The eye movement data were processed following these procedures: (1) the trials with 35% or higher track loss were removed, because these trials did not include enough samples of eye movements for meaningful analyses. As a result, 8.19% of the total samples were removed; (2) the eye movement data which were recorded during the onset of the reflexive *ziji* and the offset of the phrase after *ziji* (i.e. *zhejian shi* "this event") were extracted. The mean length of the time window for each trial is 1733.33ms; (3) the extracted data were segmented into 200ms time bins, and we found that the samples recorded 1400ms after the onset of *ziji* were too few for any meaningful analysis, so only the samples within the 0-1400ms time window were included in this analysis; (4) by examining the recordings to the stimulus sentences, we confirmed that *ziji* lasted from 0ms to 400ms, and the phrase in the spillover region (i.e. *zhejian shi* "this event") lasted from 400ms to 1400ms; (5) all the eye movements that fell outside the AOIs containing the local or LD antecedents were converted into "missing". The eye movement data were processed using the *eyetrackingR* package (Dink & Ferguson, 2015) in *R*, and the graphs were generated using the *ggplot2* package (Wickham, 2009) and the *sjPlot* package (Lüdecke, 2018).

As a number of previous studies suggest, monolingual Mandarin speakers first search for the local antecedent in early stages of processing *ziji*, regardless of the animacy of local antecedent (e.g. Dillon et al., 2014, as reviewed in Chapter 2). Based on these findings, we assumed that, within the concerned time windows, the participants in this study will look more into the area of interest (AOI) containing the local antecedents when hearing *ziji* under all the conditions (see Table 5.2 for the list of the three conditions). In other words, even under the LD favoured condition, in which *ziji* favours an LD interpretation, the participants would look more into the AOI containing the local antecedents rather than the AOI which contains the LD antecedents within this time window. However, it should also be noted that, due to the design of our stimulus sentences, the participants had always heard the local antecedents before hearing *ziji*. Therefore, a sharp increase of looks into the local antecedents during the early processing of *ziji* should not always be expected, because the participants might have already fixated on the local antecedents before the processing of *ziji*.

In the rest of this section, we first visualize the bilinguals' and the monolinguals' proportions of looks into the AOIs containing the local and the LD antecedents. The proportions were calculated based on the looks to the local and the LD antecedents, and those eye movements coded as "missing" were not included in this calculation. Following that, we perform inferential statistical tests to find out whether and how Condition, Group and WM Capacity may affect the proportions of looks into the concerned AOIs, as well as the potential effect of Length of L2 Exposure on the bilinguals' processing *ziji*.

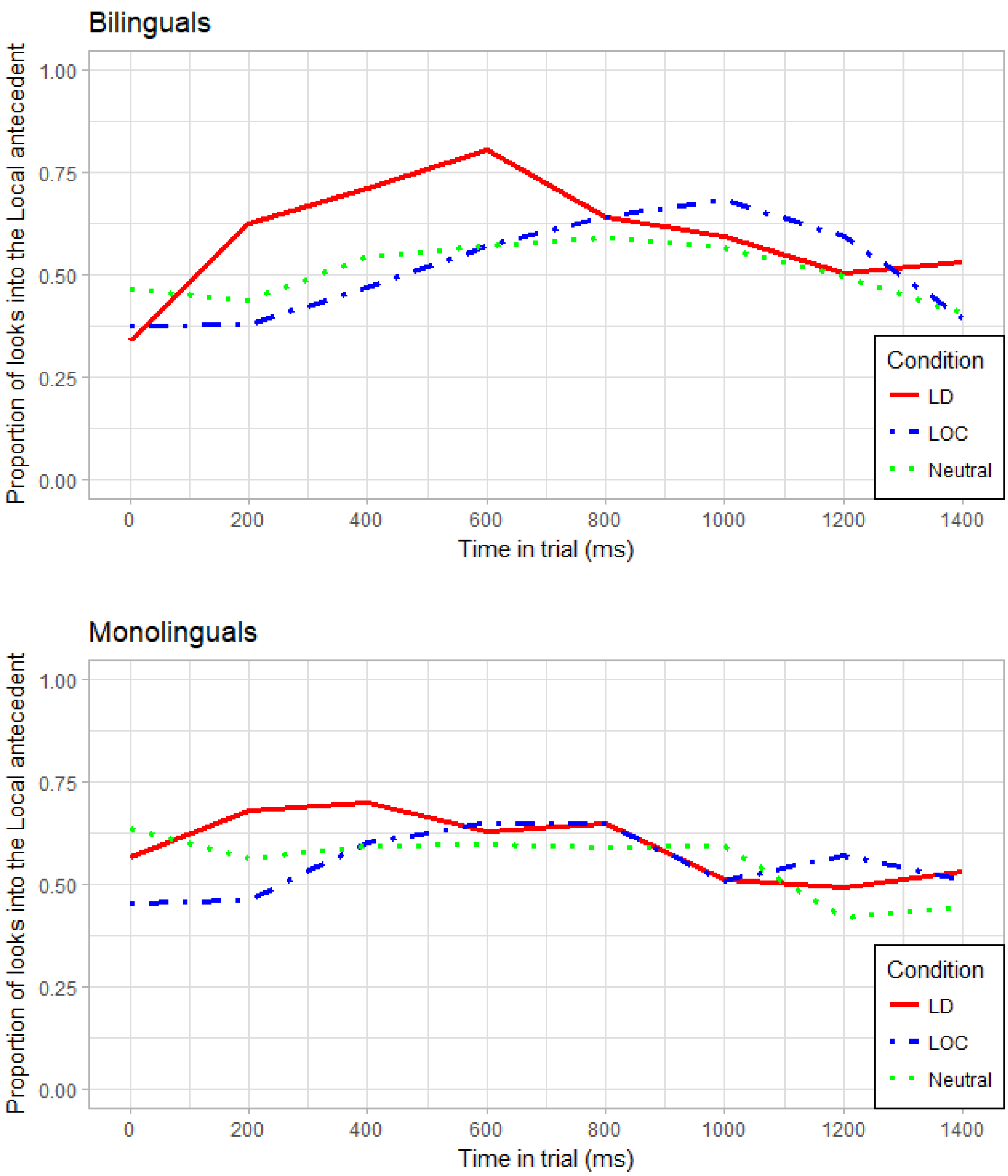


Figure 5.1. Proportions of looks into the AOIs containing local antecedents by Condition and Group

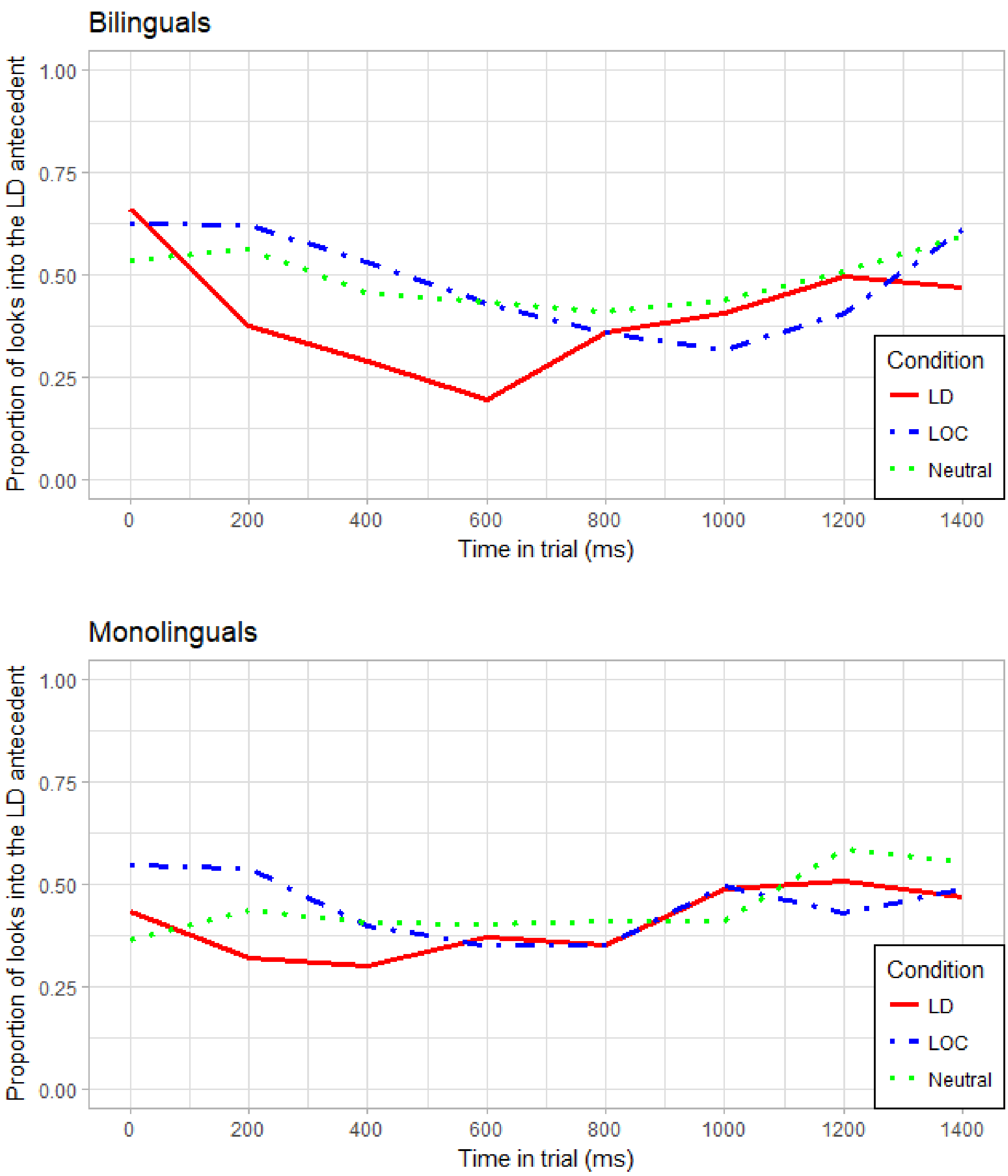


Figure 5.2. Proportions of looks into the AOIs containing LD antecedents by Condition and Group

As seen in Figure 5.1 and 5.2, the two groups showed similar processing patterns within the concerned time windows. For both groups, there was an initial increase of proportions of looks into the local antecedents, followed by a decrease of proportions of looks into the same AOI under the LD favoured and the LOC favoured conditions. Under these two conditions, both groups also showed an initial decrease of proportions of looks into the LD antecedents, followed by an increase of proportions of looks into the same AOI.

Regarding the sentences of Neutral condition, the bilinguals and the monolinguals also demonstrated similar processing pattern. For both groups, the proportions of looks into the local antecedents slightly increased 200ms after hearing *ziji*, and kept at a stable level for approximately 800ms, and finally decreased after that. Meanwhile, the proportions of looks into the LD antecedents showed the reversed processing pattern, beginning with a decrease and ending with an increase.

Aiming to confirm whether the visually observed differences were statistically significant, we conducted a series of analyses. First, we conducted a window analysis on the participants' proportion of looks into the local antecedents for each 200ms time window, so that we could examine if Condition, Group and/or WM Capacity had any effect on the proportion of looks into this AOI within each time window. For this analysis, linear mixed-effects models were used. In these models, the dependent variable was the empirical logit of the proportion of looks into the local antecedents (computed and weighted following the procedures described in Barr, 2008), and the predictors include Condition (Neutral vs LD favoured vs LOC favoured), Group (bilinguals vs monolinguals; monolinguals coded as 0.5 and bilinguals as -0.5) and WM Capacity. The reference level of Condition was set as LOC favoured. The random factors included Subject and Item, and there were random intercepts for both factors - Subject had random slopes for the fixed effect of Condition, and Item had random slopes for the fixed effects of Group and WM Capacity. The model estimates, standard errors, *t* values and *p* values of these models are presented in Table 5.9.

Table 5.9. Model estimates, standard errors, t values and p values of the linear mixed-effects models concerning proportions of looks into the local antecedents during each time window

Time window: 0-200ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.58	0.20	8.08	<0.001
ConditionNeutral	0.03	0.28	0.12	0.90
ConditionLD	-0.07	0.32	-0.22	0.83
Group	-0.20	0.43	-0.46	0.65
WM	-2.19	1.20	-1.83	0.08
ConditionNeutral:Group	0.68	0.60	1.13	0.27
ConditionLD:Group	0.77	0.70	1.11	0.27
ConditionNeutral:WM	2.35	1.67	1.40	0.17
ConditionLD:WM	0.69	2.02	0.34	0.74
Group:WM	5.45	2.60	2.10	0.04
ConditionNeutral:Group:WM	-0.71	3.60	-0.20	0.84
ConditionLD:Group:WM	-3.33	4.25	-0.78	0.44
Time window: 200-400ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.56	0.22	7.05	<0.001
ConditionNeutral	0.07	0.26	0.29	0.78
ConditionLD	0.58	0.22	2.69	0.01
Group	-0.13	0.43	-0.31	0.76
WM	-1.89	1.45	-1.31	0.20
ConditionNeutral:Group	0.35	0.49	0.72	0.48
ConditionLD:Group	0.04	0.40	0.10	0.92
ConditionNeutral:WM	0.84	1.65	0.51	0.62
ConditionLD:WM	-0.16	1.36	-0.12	0.91
Group:WM	3.72	2.98	1.25	0.22
ConditionNeutral:Group:WM	1.01	3.44	0.29	0.77
ConditionLD:Group:WM	-0.01	2.89	0.00	0.99

Time window: 400-600ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.33	0.22	6.01	<0.001
ConditionNeutral	0.22	0.30	0.75	0.46
ConditionLD	0.51	0.28	1.84	0.07
Group	0.23	0.43	0.53	0.60
WM	-1.55	1.50	-1.04	0.31
ConditionNeutral:Group	-0.09	0.57	-0.16	0.88
ConditionLD:Group	-0.38	0.53	-0.73	0.47
ConditionNeutral:WM	1.61	1.95	0.82	0.42
ConditionLD:WM	0.52	1.82	0.29	0.78
Group:WM	-0.67	2.80	-0.24	0.81
ConditionNeutral:Group:WM	1.04	3.62	0.29	0.78
ConditionLD:Group:WM	3.38	3.33	1.02	0.32
Time window: 600-800ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.78	0.19	9.14	<0.001
ConditionNeutral	0.15	0.25	0.60	0.55
ConditionLD	0.49	0.23	2.09	0.04
Group	0.43	0.32	1.33	0.19
WM	-1.95	1.16	-1.68	0.10
ConditionNeutral:Group	-0.65	0.39	-1.66	0.11
ConditionLD:Group	-0.60	0.34	-1.75	0.09
ConditionNeutral:WM	-0.45	1.41	-0.32	0.75
ConditionLD:WM	0.89	1.27	0.70	0.49
Group:WM	0.73	2.48	0.29	0.77
ConditionNeutral:Group:WM	-0.77	3.10	-0.25	0.81
ConditionLD:Group:WM	1.33	2.84	0.47	0.64

Time window: 800-1000ms	Estimate	Std.Error	*t*-value	*p*-value
(Intercept)	1.71	0.22	7.89	<0.001
ConditionNeutral	-0.40	0.31	-1.28	0.21
ConditionLD	0.17	0.28	0.60	0.55
Group	0.28	0.39	0.72	0.47
WM	-1.12	1.11	-1.01	0.32
ConditionNeutral:Group	-0.18	0.56	-0.31	0.76
ConditionLD:Group	-0.56	0.49	-1.14	0.26
ConditionNeutral:WM	-0.18	1.58	-0.11	0.91
ConditionLD:WM	0.03	1.33	0.02	0.98
Group:WM	0.13	2.57	0.05	0.96
ConditionNeutral:Group:WM	-2.09	3.66	-0.57	0.57
ConditionLD:Group:WM	0.34	3.22	0.11	0.92
Time window: 1000-1200ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.37	0.23	6.01	<0.001
ConditionNeutral	0.37	0.26	1.40	0.17
ConditionLD	0.14	0.29	0.49	0.63
Group	-0.57	0.44	-1.32	0.19
WM	-0.43	1.28	-0.34	0.74
ConditionNeutral:Group	0.33	0.49	0.68	0.50
ConditionLD:Group	0.44	0.54	0.81	0.43
ConditionNeutral:WM	0.59	1.38	0.43	0.68
ConditionLD:WM	-0.39	1.54	-0.26	0.80
Group:WM	-6.47	2.41	-2.68	0.01
ConditionNeutral:Group:WM	4.80	2.50	1.92	0.07
ConditionLD:Group:WM	5.28	2.86	1.85	0.08

Time window: 1200-1400ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.36	0.20	6.76	<0.001
ConditionNeutral	0.27	0.28	0.96	0.34
ConditionLD	0.27	0.30	0.91	0.37
Group	0.03	0.34	0.09	0.93
WM	1.70	1.31	1.29	0.20
ConditionNeutral:Group	-0.78	0.48	-1.61	0.12
ConditionLD:Group	-0.70	0.53	-1.33	0.19
ConditionNeutral:WM	-1.36	1.86	-0.73	0.47
ConditionLD:WM	-2.99	1.97	-1.51	0.14
Group:WM	-4.71	2.49	-1.90	0.07
ConditionNeutral:Group:WM	6.44	3.53	1.83	0.08
ConditionLD:Group:WM	5.71	3.77	1.51	0.14

Now we discuss the models for each time window in turn. In the 0-200ms time window, which corresponds to the onset of *ziji*, our model found a significant interaction between Group and WM Capacity. Figure 5.3 shows the Group x WM Capacity effect on the proportion of looks into the local antecedents during this time window. As it can be seen from the graph, the bilinguals with larger WM Capacity were less likely to look into the local antecedents, but this pattern was not so obvious in the monolinguals - the monolinguals tended to look at the local antecedents regardless of their WM Capacity. However, as it normally takes 200ms to program an eye movement (Rayner, Slowiaczek, Clifton, & Bertera, 1983), this finding should be interpreted as that the WM Capacity had different effects on the bilinguals and the monolinguals before the processing of *ziji*, rather than during the early stages of processing *ziji*. Therefore, this finding is not relevant to our research questions, and we will not further discuss this.

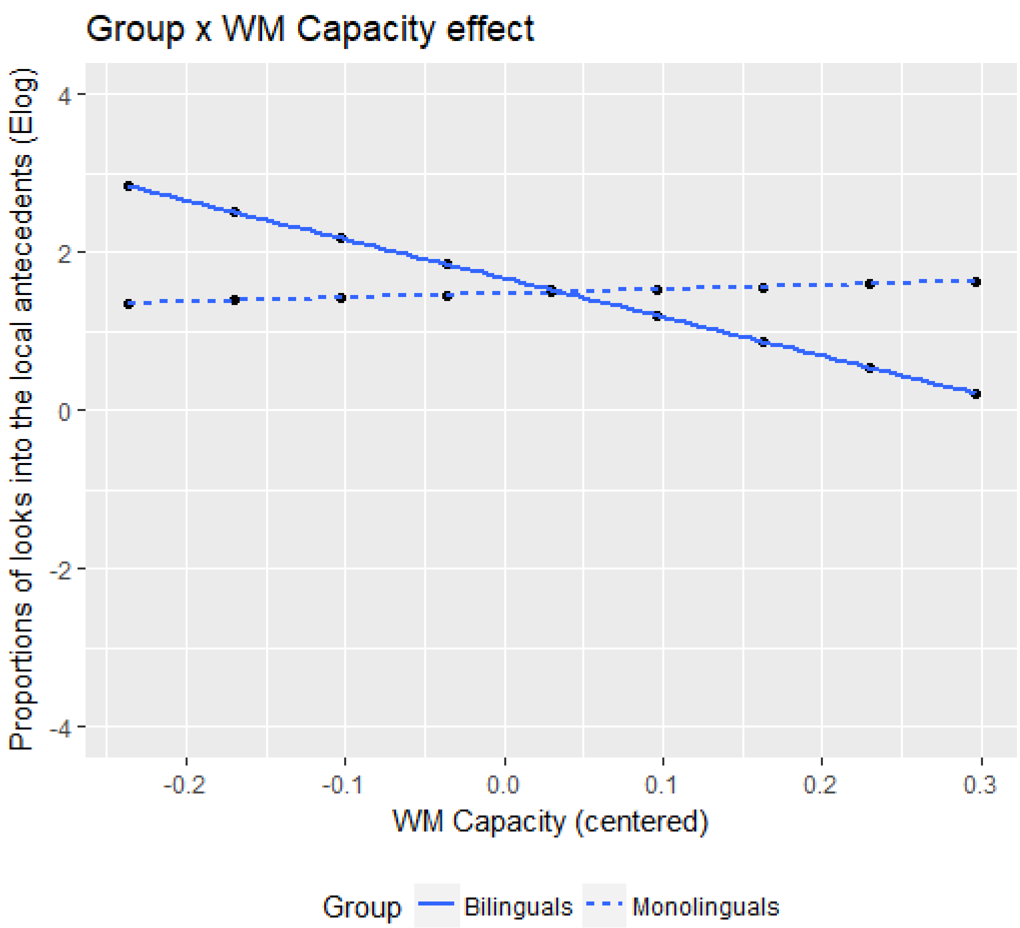


Figure 5.3. Group x WM Capacity effect on the proportion of looks into the local antecedents in the 0-200ms time window

In the 200-400ms time window, which corresponds to the offset of *ziji*, our model found a significant fixed effect of Condition: LD. This finding suggests that, during this time window, both groups of participants tended to look more into the local antecedents under the LD favoured condition than under the LOC favoured condition. However, there was no significant effect of Group, WM or any interaction between the predictors. The absence of such significant effects suggest that the bilinguals and the monolinguals did not differ in processing *ziji* at this early stage, and it is consistent with our prediction that L1 attrition effect was unlikely to emerge during the early syntactic processing of *ziji*.

As we did not find any significant effect of the predictors in the 400-600ms time window, we assume that there was no evidence suggesting any of the predictors had affected the processing of *ziji* during this time window. Therefore, the bilinguals and the monolinguals did not differ in processing *ziji* at this stage either, and this finding is also consistent with our prediction that L1 attrition effect was unlikely to emerge during the early syntactic processing of *ziji*.

In the 600-800ms time window, a significant fixed effect of Condition: LD has been observed. This effect suggests that the participants tended to look more into the local antecedents under the LD favoured condition than under the LOC favoured condition. Again, we failed to observe any significant effect of Group, WM or any interaction between the predictors in this time window. In the 800-1000ms time window, there was no significant effect of any predictors or any interaction either. Therefore, we assume that the bilinguals did not differ from the monolinguals in processing *ziji* in these time windows.

In the 1000-1200ms time window, our model produced a significant interaction between Group and WM Capacity. This interaction effect is presented in Figure 5.4. As seen in the figure, WM Capacity had different effects on the bilinguals' and the monolinguals' proportions of looks into the local antecedents within this time window. The bilinguals with larger WM Capacity were more likely to look into the local antecedents, while the monolinguals showed the reversed pattern. This Group x WM Capacity interaction suggests an L1 attrition effect which was moderated by individual differences in WM Capacity, and we will further discuss this issue in Chapter 6. Finally, our model did not reveal any significant effect of the predictors, or any significant interaction in the 1200-1400ms time window.

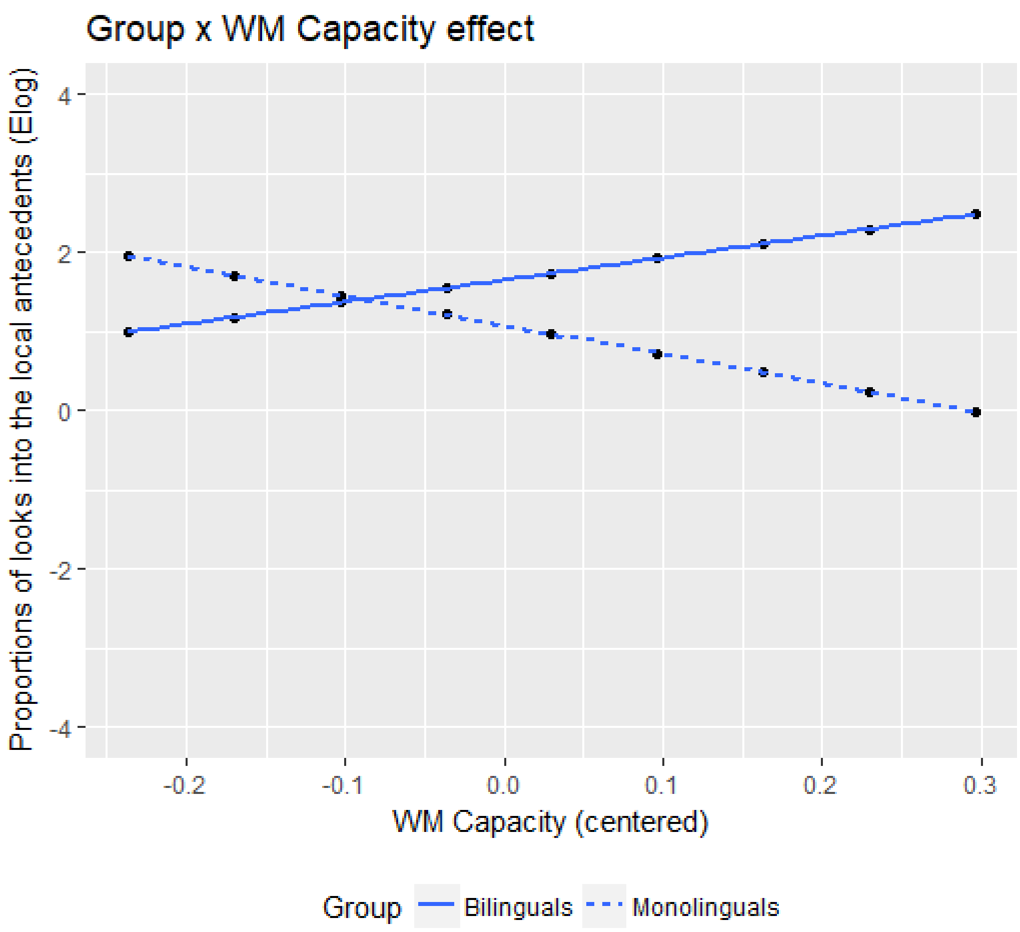


Figure 5.4. Group x WM Capacity effect on the proportion of looks into the local antecedents in the 1000-1200ms time window

As we are also interested in the potential effects of Length of L2 Exposure, as well as how WM Capacity may interact with Length of L2 Exposure and explain variance in the processing of *ziji*, we performed another window analysis within the bilingual group. For each 200ms time bin, we examined the effects of Condition, Length of L2 Exposure and WM Capacity on the empirical logit of proportion of looks to the local antecedents. The empirical logit was weighted during the analysis. Linear mixed-effects models were used, with Subject and Item being the random factors. Both random factors had random intercepts; Subject had random slopes for the fixed effect of Condition, and Item had random slopes for the fixed effects of

Length of L2 Exposure and WM Capacity. In cases where the model failed to converge, the random slope parameter that accounted for the least amount of variance was removed and the model refitted until convergence was achieved. These models are summarized in Table 5.10.

Table 5.10. Model estimates, standard errors, t values and p values of the linear mixed-effects models concerning potential Length of L2 Exposure effect

Time window: 0-200ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.30	0.27	4.72	<0.001
ConditionLD	-0.36	0.44	-0.82	0.42
ConditionNeutral	-0.23	0.42	-0.56	0.58
L2Exp	0.04	0.04	1.05	0.31
WM	-5.53	1.73	-3.19	<0.01
ConditionLD:L2Exp	-0.10	0.06	-1.72	0.10
ConditionNeutral:L2Exp	-0.08	0.06	-1.45	0.17
ConditionLD:WM	4.77	2.88	1.65	0.11
ConditionNeutral:WM	6.13	2.65	2.31	0.04
L2Exp:WM	-0.21	0.36	-0.57	0.58
ConditionLD:L2Exp:WM	1.26	0.65	1.93	0.07
ConditionNeutral:L2Exp:WM	1.26	0.58	2.17	0.05
Time window: 200-400ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.39	0.29	4.76	<0.001
ConditionLD	0.57	0.32	1.76	0.09
ConditionNeutral	-0.33	0.38	-0.87	0.40
L2Exp	0.06	0.04	1.47	0.18
WM	-3.16	2.15	-1.47	0.16
ConditionLD:L2Exp	-0.05	0.05	-1.18	0.28
ConditionNeutral:L2Exp	-0.14	0.05	-2.59	0.03
ConditionLD:WM	-0.01	2.20	0.00	1.00
ConditionNeutral:WM	3.33	2.64	1.26	0.22
L2Exp:WM	0.43	0.46	0.93	0.37
ConditionLD:L2Exp:WM	-0.32	0.49	-0.65	0.52
ConditionNeutral:L2Exp:WM	0.42	0.57	0.73	0.47

Time window: 400-600ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.42	0.28	5.11	<0.001
ConditionLD	0.66	0.31	2.12	0.05
ConditionNeutral	0.51	0.44	1.15	0.28
L2Exp	0.01	0.03	0.38	0.71
WM	-0.52	1.84	-0.28	0.78
ConditionLD:L2Exp	0.04	0.03	1.32	0.20
ConditionNeutral:L2Exp	0.02	0.05	0.33	0.75
ConditionLD:WM	-1.47	1.76	-0.83	0.41
ConditionNeutral:WM	0.88	3.01	0.29	0.78
L2Exp:WM	-0.14	0.38	-0.36	0.73
ConditionLD:L2Exp:WM	0.18	0.36	0.51	0.61
ConditionNeutral:L2Exp:WM	0.32	0.62	0.52	0.62
Time window: 600-800ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.45	0.28	5.21	<0.001
ConditionLD	0.95	0.31	3.09	<0.01
ConditionNeutral	0.79	0.34	2.32	0.03
L2Exp	-0.01	0.04	-0.16	0.87
WM	-4.17	1.98	-2.11	0.05
ConditionLD:L2Exp	0.02	0.04	0.64	0.54
ConditionNeutral:L2Exp	-0.01	0.04	-0.27	0.79
ConditionLD:WM	2.94	2.06	1.43	0.18
ConditionNeutral:WM	3.16	2.37	1.33	0.20
L2Exp:WM	-1.27	0.44	-2.88	<0.01
ConditionLD:L2Exp:WM	1.01	0.46	2.17	0.05
ConditionNeutral:L2Exp:WM	1.35	0.54	2.53	0.02

Time window: 800-1000ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.58	0.26	6.10	<0.001
ConditionLD	0.34	0.34	1.01	0.33
ConditionNeutral	-0.24	0.36	-0.66	0.52
L2Exp	-0.02	0.03	-0.70	0.51
WM	-2.86	1.53	-1.87	0.09
ConditionLD:L2Exp	0.01	0.03	0.35	0.73
ConditionNeutral:L2Exp	-0.01	0.04	-0.28	0.78
ConditionLD:WM	1.29	1.73	0.74	0.47
ConditionNeutral:WM	0.60	2.09	0.29	0.78
L2Exp:WM	-0.73	0.34	-2.16	0.06
ConditionLD:L2Exp:WM	0.42	0.40	1.05	0.31
ConditionNeutral:L2Exp:WM	0.45	0.48	0.93	0.37
Time window: 1000-1200ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.83	0.31	5.99	<0.001
ConditionLD	-0.18	0.34	-0.52	0.61
ConditionNeutral	0.14	0.34	0.41	0.69
L2Exp	-0.04	0.03	-1.04	0.33
WM	0.78	1.74	0.45	0.66
ConditionLD:L2Exp	0.00	0.03	0.07	0.94
ConditionNeutral:L2Exp	0.05	0.03	1.39	0.20
ConditionLD:WM	-1.50	1.46	-1.02	0.32
ConditionNeutral:WM	-0.77	1.44	-0.53	0.60
L2Exp:WM	-0.40	0.43	-0.91	0.38
ConditionLD:L2Exp:WM	0.47	0.40	1.17	0.25
ConditionNeutral:L2Exp:WM	0.34	0.41	0.84	0.40

Time window: 1200-1400ms	**Estimate**	**Std.Error**	***t*-value**	***p*-value**
(Intercept)	1.37	0.21	6.51	<0.001
ConditionLD	0.65	0.39	1.66	0.12
ConditionNeutral	0.73	0.29	2.53	0.02
L2Exp	-0.04	0.03	-1.09	0.29
WM	4.02	1.41	2.85	0.01
ConditionLD:L2Exp	0.01	0.06	0.12	0.90
ConditionNeutral:L2Exp	0.08	0.04	1.83	0.09
ConditionLD:WM	-6.25	2.65	-2.36	0.04
ConditionNeutral:WM	-5.01	1.82	-2.76	0.01
L2Exp:WM	-0.31	0.35	-0.89	0.38
ConditionLD:L2Exp:WM	0.78	0.63	1.23	0.24
ConditionNeutral:L2Exp:WM	-0.12	0.44	-0.26	0.80

As it can be seen in Table 5.10, none of the models concerning Length of L2 Exposure revealed a significant fixed effect of this predictor. However, in a number of time windows, we observed significant interactions between Length of L2 Exposure and Condition/WM Capacity, as well as significant interactions between WM Capacity and Condition. These interactions suggest that the variance in processing *ziji* among the bilinguals should not be explained by Length of L2 Exposure alone, rather, it should be explained by complex interactions between multiple factors. However, it is worth noting that, as opposed to our hypothesis that WM Capacity or WM Capacity x Length of L2 Exposure/Condition interaction may explain the bilinguals' variations in processing *ziji* when L1 attrition effects were observed (i.e. the 1000-1200ms time window), these significant effects/interactions were observed in the time windows where no L1 attrition effects were present. Nevertheless, in this analysis, we will look into each time window and analyze the significant effects/interactions, in order to gain an understanding about how these effects/interactions explain the variances within the bilinguals.

In the 0-200ms time window, our model found a significant fixed effect of WM Capacity, a significant interaction between WM Capacity and Condition: Neutral, and a significant interaction between Condition: Neutral, Length of L2 Exposure and WM Capacity. The fixed effect of WM Capacity suggests that the bilinguals with larger WM Capacity tended to look less into the local antecedents. The two-way interaction between WM Capacity and Condition: Neutral suggests that, the bilinguals with larger WM Capacity tended to look more into the local antecedents under the Neutral condition rather than the LOC favoured condition. More importantly, the three-way interaction suggests that, only under the Neutral condition, the bilinguals with longer length of L2 exposure tended to look less into the local antecedents if their WM Capacity were larger, but the bilinguals with shorter length of L2 exposure showed the reversed pattern. This three-way interaction is presented in Figure 5.5. However, as discussed before in the group comparison, this time window does not correspond to the actual processing of *ziji* and these effects may not be relevant to our research questions, so we will not further discuss them here.

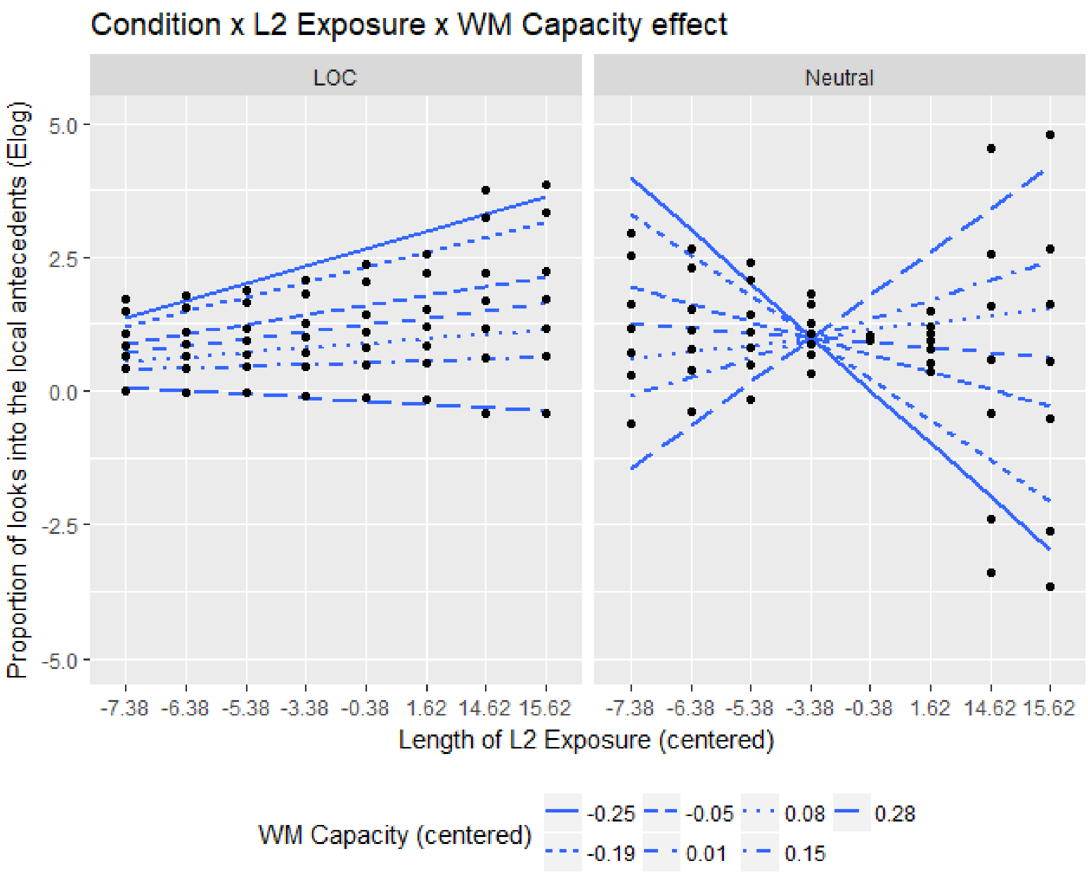


Figure 5.5. Condition x Length of L2 Exposure x WM Capacity effect on the proportion of looks into the local antecedents in the 0-200ms time window

In the 200-400ms time window, there was a significant interaction between Condition: Neutral and Length of L2 Exposure. This interaction is presented in Figure 5.6, and it suggests that, under the Neutral condition, the bilinguals tended to look less into the local antecedents as their Length of L2 Exposure increases. As this time corresponds to the early, syntactic processing of *ziji*, this interaction may indicate that, the Mandarin-English bilinguals with longer L2 exposure may deviate from those bilinguals with shorter L2 exposure in processing *ziji*. However, given that we only had a small number of bilinguals with more than 15 years' exposure to L2 English, and that this effect was not observed in the following time window, which also corresponds to the early processing of *ziji*, it is difficult to confirm whether this interaction should be interpreted as a sign of L1 attrition. We will further discuss this issue in Chapter 6. The model concerning the 400-600ms time window did not reveal any significant fixed effect of or interaction between the predictors, so it will not be discussed here.

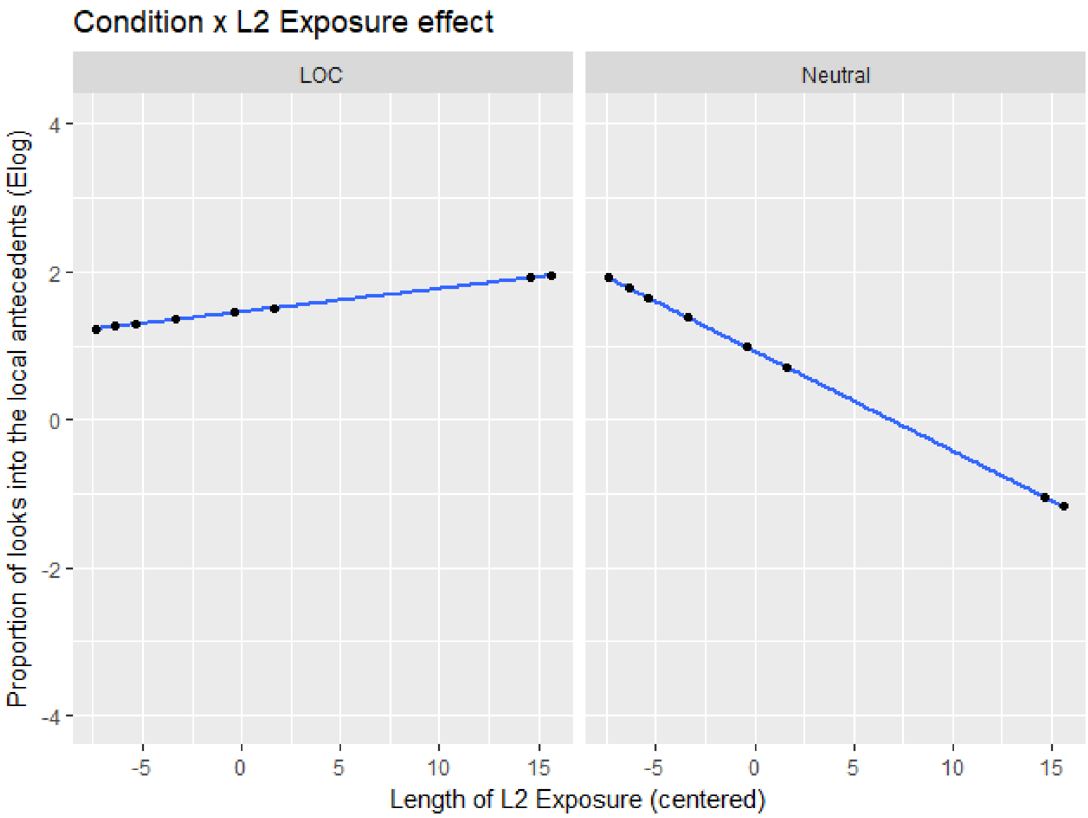


Figure 5.6. Condition x Length of L2 Exposure effect on the proportion of looks into the local antecedents in the 200-400ms time window

In the 600-800ms time window, we found significant fixed effects of Condition: LD, Condition: Neutral, WM Capacity, and significant interactions between Length of L2 Exposure, WM Capacity and Condition. The significant fixed effects of Condition: LD and Condition: Neutral suggest that the bilinguals tended to look more into the local antecedents under the LD favoured and Neutral conditions than under the LOC favoured condition, and the significant WM Capacity effect suggests that the bilinguals with larger WM Capacity tended to look less into the local antecedents as their WM Capacity. However, the significant interactions suggest that these effects varied according to Length of L2 Exposure.

The Length of L2 Exposure x WM Capacity interaction suggests that, while the bilinguals with longer L2 exposure tended to look less into the local antecedents if their WM Capacity were larger, WM Capacity did not have similar effect on those bilinguals with shorter length of L2 exposure. In addition, the three-way interactions between Condition, Length of L2 Exposure and WM Capacity suggest that the Length of L2 Exposure x WM Capacity effect varied according to Conditions - as Figure 5.7 demonstrates, the Length of L2 Exposure x WM Capacity effect was much larger under the LOC favoured condition than under the other two conditions. These interactions seem to support our hypothesis that individual differences in WM Capacity may interact with other predictors, and explain variance in language processing within the bilinguals. This issue will be further interpreted in Chapter 6.

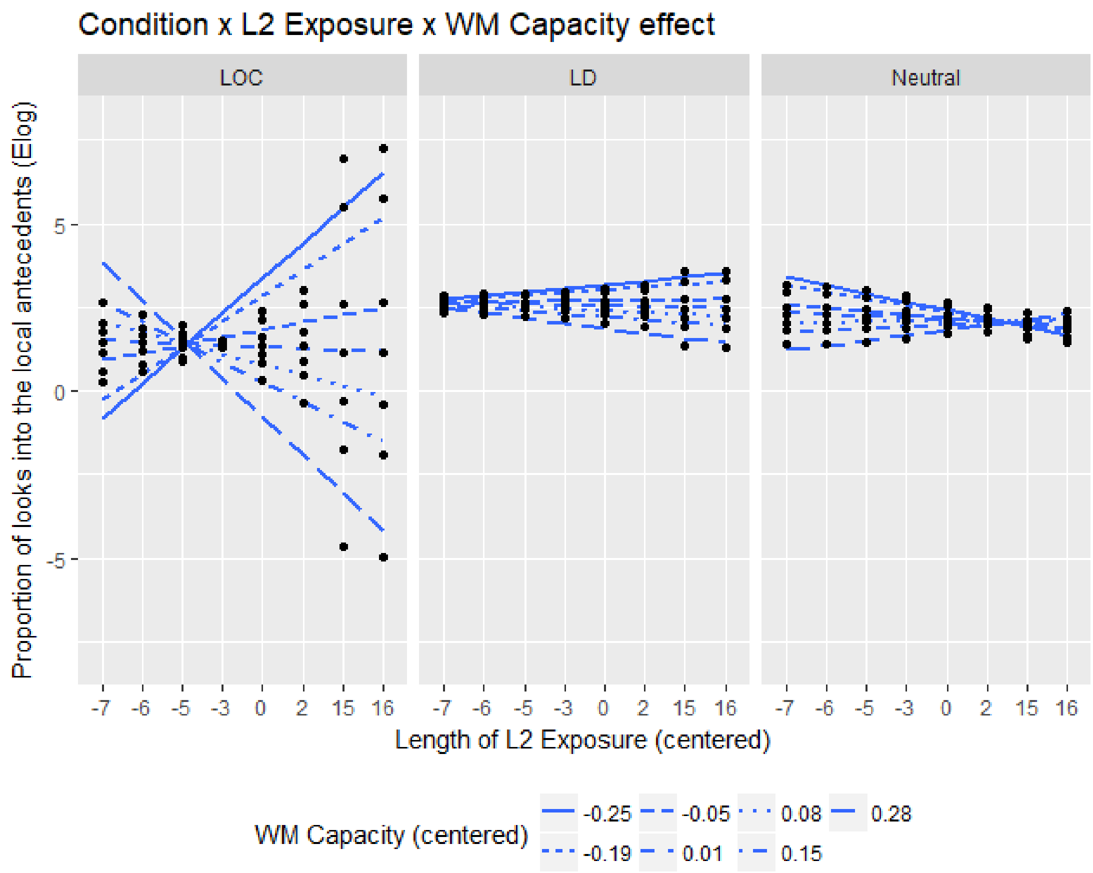


Figure 5.7. Condition x Length of L2 Exposure x WM Capacity effect on the proportion of looks into the local antecedents in the 600-800ms time window

The models concerning the 800-1000ms and 1000-1200ms did not reveal any significant fixed effect of or interaction between the predictors, therefore we will not discuss these two models here. In the final 1200-1400ms time window, the model revealed a significant fixed effect of WM Capacity and significant interactions between WM Capacity and Condition: LD/Condition: Neutral. The fixed effect of WM Capacity suggests that the bilinguals with larger WM Capacity were slightly more likely to look into the local antecedents. However, the significant interactions between WM Capacity and Condition suggest that the WM Capacity effect varied according to Condition. As Figure 5.8 shows, under the Neutral and LD conditions, the bilinguals with larger WM Capacity looked less into the local antecedents.

Meanwhile, under the LOC condition, the bilinguals with larger WM Capacity were more likely to look into the local antecedents.

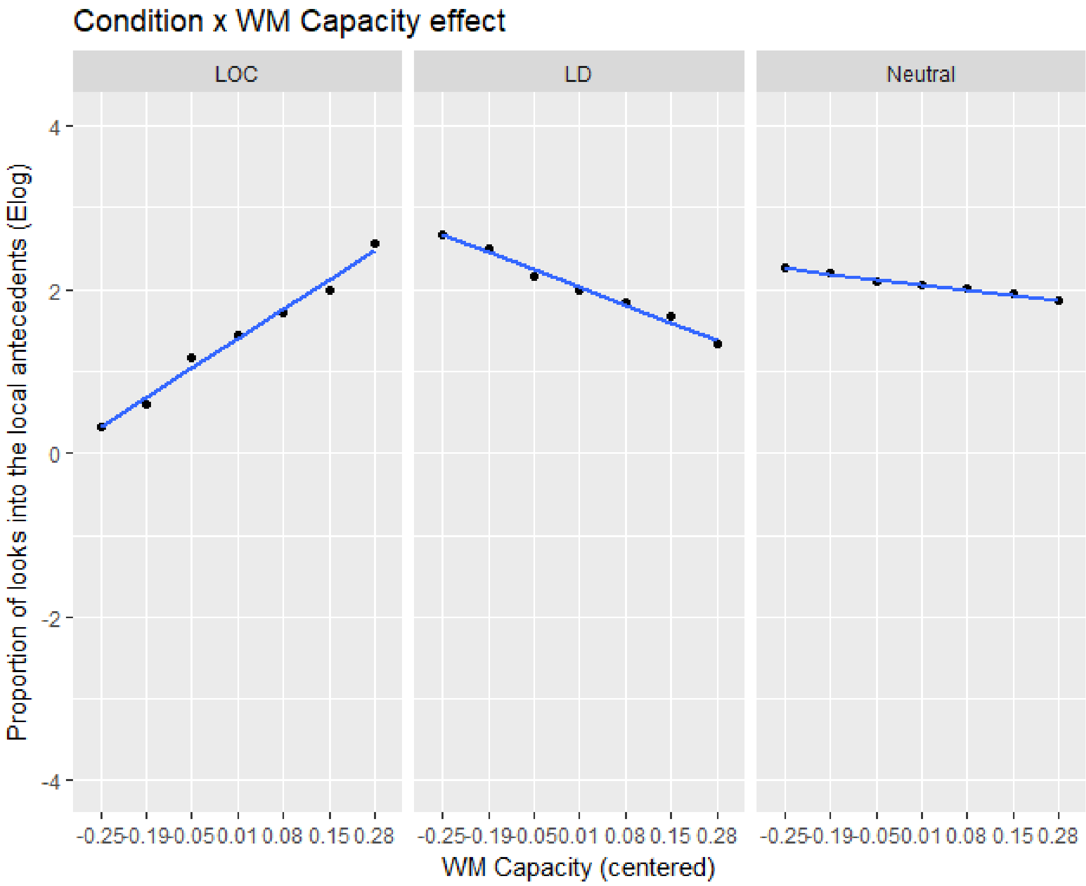


Figure 5.8. Condition x WM Capacity effect on the proportion of looks into the local antecedents in the 1200-1400ms time window

The above analysis concerning the Length of L2 Exposure and WM Capacity effects on the processing of *ziji* among bilinguals revealed various interactions between Length of L2 Exposure/WM Capacity and other predictors (e.g. Condition) in various time windows. These interactions suggest that Length of L2 Exposure and/or WM Capacity interacted with other predictors and affected the on-line processing of *ziji* among the bilinguals, and this finding partially supported our hypothesis that WM Capacity might play a role in explaining variations in L1 processing among the bilinguals. However, our findings did not suggest a clear relationship between these different predictors, and do not support our hypothesis that bilinguals with larger WM capacity would behave more monolingual-like, and those with smaller WM capacity less so. The implications of these findings will be further discussed in Chapter 6.

In summary, the results of this eye-tracking task suggest that, while the bilinguals deviated from the monolinguals in the processing of *ziji* at later stages and thus showed L1 attrition in this respect, such between-group difference was not explained by the status of bilingualism alone, but by the interaction between Group and WM Capacity. Meanwhile, the absence of Group or Length of L2 Exposure in the 200-400ms and 400-600ms time windows seems to support the prediction of the IH, as the bilinguals resemble the monolinguals in the early processing of *ziji*. With respect to the variance within the bilingual group, our analysis found that Length of L2 Exposure interacted with WM Capacity and Condition and affected the bilinguals' on-line processing of *ziji*. On the one hand, it seems that the bilinguals with longer L2 exposure deviated from the bilinguals with shorter L2 exposure in the early processing of *ziji* (i.e. in the 200-400ms time window) under the Neutral condition, but such deviation did not lead to L1 attrition, as the group comparison revealed no Group effect in this time window. On the other hand, the variance in the later stages of processing *ziji* (e.g. in the 600ms-onwards time windows) seemed to be explained by the complex three-way interactions between Length of L2 Exposure, WM Capacity and Condition. These findings partially support our hypothesis about the role of WM Capacity in explaining variance in language processing among bilinguals - individual differences in WM Capacity did play a role, but not in the pattern we speculated. The implications of our findings will be discussed in more detail in Chapter 6.

5.4.Summary

This chapter reported the bilinguals' and the monolinguals' performance in interpreting and processing *ziji*. As Section 5.1 and 5.2 suggest, the bilinguals largely resembled the monolinguals in their preference for interpreting *ziji* under different conditions. The only significant difference we observed is that, in the interpretation task included in the eye-tracking task, the bilinguals were less likely to choose a LOC interpretation of *ziji* than the monolinguals under the Neutral condition. In the next chapter, we will discuss if we should interpret this difference as L1 attrition. Meanwhile, our analysis does not suggest that the bilinguals' preference for interpreting *ziji* varied according to Length of L2 Exposure, or the interaction between Length of L2 Exposure and other predictors.

With respect to the on-line processing of *ziji*, the absence of a Group effect in the earlier time windows (i.e. 200-400ms and 400-600ms time windows) seems to support the IH, which predicts that there would be no between-group difference in the early processing of *ziji* because the on-line processing at this stage only involves syntax. Moreover, in line with the IH, our between-group comparison revealed a between-group difference in the on-line processing of *ziji* at later stages, which are supposed to involve semantics and pragmatics. This difference was explained by the interaction between Group and WM Capacity.

In our analysis of the Length of L2 Exposure and WM Capacity effects on the on-line processing of *ziji*, it was observed that Length of L2 Exposure and/or WM Capacity interacted with Condition and had affected the on-line processing of *ziji* in various time windows. These finding partially support our view that individual differences in cognitive abilities played a role in explaining potential bilingual/monolingual differences in language processing, and the variance in language processing within the bilingual group. However, these findings are not fully consistent with our speculation, as we did not observe a pattern in which the bilinguals with larger WM Capacity would behave more monolingual-like. In Chapter 6, we will further discuss the implications of these findings.

6. Discussion

In this chapter, we discuss a series of issues based on the findings reported in Chapter 4 and 5. In the first two sections, we recapitulate the findings presented in Chapter 4 and 5, and discuss whether these findings should be interpreted as L1 attrition in aspect marking and LD binding. Then, we look into the relationship between individual differences in WM capacity and variation in L1 attrition. Following that, we discuss the theoretical implications of these findings, focusing on whether they could be explained by the IH and/or alternative theoretical frameworks. In the last two sections, we discuss some other general issues not directly relate to the objectives of the present study, but are relevant to Chinese linguistics or general L1 attrition research, as well as the limitations of this study.

6.1.L1 attrition in aspect marking?

In Chapter 4, we compared the Mandarin-English bilinguals' and the Mandarin monolinguals' performance in a series of tasks concerning the perception, production and processing of perfective and durative aspect marking, and we observed a few between-group and within-group differences. Here we argue that these differences should not be interpreted as L1 attrition in perfective/durative aspect marking based on Schmid & Köpke (2017a)'s definition, which views L1 attrition as a continuum and treats all bilingualism-caused changes to L1 processing and representation as L1 attrition.

The results of the AJT task did not reveal any significant Group effect in the AJT task, suggesting that the bilinguals resembled the monolinguals in perceiving the interaction between *le/zhe* and verbs/predicates of different lexical aspects. In this task, both groups' performance was generally consistent with the predictions of Xiao

& McEnery (2004). Both groups tended to accept the sentences containing ACC+*le/zhe*, ACT+*le/zhe*, SLS+*le/zhe*, SEM+*le/zhe*, ILS+*le* and ACH+*le,* while rejecting the ILS+*zhe* and ACH+*zhe* sentences. In the between-group comparison, we found a marginally significant Group effect for the ILS+z*he* sentences, which suggests that the bilinguals were more likely to reject the ungrammatical ILS+*zhe* sentences than the monolinguals. However, as discussed in Section 4.3, this effect was caused by the variance within the monolingual group, and thus did not suggest that the bilinguals showed attrition in this respect and became unable to reject these ungrammatical sentences. Therefore, this Group effect on the ILS+z*he* sentences should not be interpreted as an L1 attrition effect, and the results of the AJT task do not suggest L1 attrition in perceiving the interaction between *le/zhe* and different types of lexical aspects.

Within the bilingual group, we observed a significant effect of Length of L2 Exposure on the acceptability scores for the ACT+*le* sentences, and a marginally significant Length of L2 Exposure effect for the SEM+*le* sentences. However, both effects were caused by one individual bilingual with 29 years' exposure to L2 English, who consistently rejected these grammatical sentences. It should be noted though, this particular bilinguals' performance cannot suggest that L1 attrition effects in aspect marking would emerge after many years of L2 exposure. This particular bilingual performed monolingual-like in all other tasks when ACT+*le* sentences were concerned. In the cloze task, this bilingual was 100% accurate in supplying *le* when ACT verbs were present; in the sentence-picture matching task, this bilingual was also 100% accurate in matching the ACT+*le* sentences to the pictures depicting completed events. Such performance suggests that this bilingual had monolingual-like knowledge about the interaction between ACT and *le*, and the unexpected performance might have resulted from some random factors that we were unable to control.

Unfortunately, as we did not use any SEM verbs/predicates in the cloze or sentence-picture matching task, we were unable to further confirm if this bilingual actually experienced L1 attrition for perceiving the interaction between SEM and *le*. Despite this, based on the fact that this particular bilingual demonstrated a rather random pattern in this AJT task (i.e. accepting the grammatical ACC+*le*, ACH+*le*, ILS+*le* and SLS+*le* sentences, and rejecting all the other types of sentences), and another bilingual with comparable L2 exposure (30 years) performed exactly monolingual-like in this task, it seems more reasonable to conclude that this bilingual had some unknown problems with performing this AJT task, rather than to conclude that long period of L2 exposure may lead to L1 attrition in perceiving the interaction between *le/zhe* and different lexical aspects. This underscores the value of having multiple experimental measures for the same individual, such that we could justify precisely why we take a particular performance to be an anomaly based on a preponderance of evidence.

Regarding the results from the cloze task, our analysis did not reveal any significant Group effect on the accuracy in supplying *le/zhe*, suggesting that both groups performed similarly in this task. Furthermore, within the bilingual group, we did not find evidence for an effect of Length of L2 Exposure on producing *le/zhe* either, as the analysis did not reveal any significant effect of Length of L2 Exposure on the accuracy in responding to the cloze task, or any significant interaction between Length of L2 Exposure and Aspect Marker Type. Based on these findings, we conclude that, for the Mandarin-English bilinguals in this study, there was no L1 attrition in producing *le/zhe* in written form under unpressured settings.

The findings from the sentence-picture matching task did not suggest any L1 attrition in the pattern or speed of processing the interaction between *le/zhe* and ACT verbs either. For this task, we found a significant Aspect Marker Type effect on the accuracy and RTs in responding to the target sentences, and this effect did not interact with the other predictors, i.e. Group and WM Capacity. This finding suggests that both the bilinguals and the monolinguals were more accurate and faster in responding to the semantically-matched ACT+*zhe* sentences than the semantically-mismatched ACT+*le* sentences. And so, both groups enjoyed a similar facilitation effect in terms of accuracy and RTs in this processing task.

It is also worth noting that, although the raw data seemingly suggest the bilinguals were much more accurate in responding to the ACT+*le* sentences, our analysis only revealed a significant effect of WM Capacity on the accuracy in responding to the target sentences, but no significant effect of Group, or any interaction between Group and WM Capacity or Aspect Marker Type. This finding suggests that, regardless of the status of bilingualism, the variations in accurately responding to both the ACT+*le* and the ACT+*zhe* sentences were explained by the participants' individual differences in WM Capacity. However, our analysis did not find a significant effect of WM Capacity on the RTs in responding to the target sentences.

The presence of significant WM Capacity effect on the accuracy in the sentence-picture matching task is not surprising. In Chapter 2 and 3, we argued that, although the processing of interaction between lexical and grammatical aspect in simple declarative sentences does not heavily tax the WM system *per se*, the sentence-picture matching task requires one to simultaneously store and process information before making decisions. Therefore, WM Capacity may constrain the performance in this particular task. The absence of a significant WM Capacity effect on the RTs in the same task might be caused by a speed-accuracy trade-off. When doing this task, the participants were instructed to respond to the sentence as fast (and accurately) as possible, and this instruction might have caused the results observed in this experiment. Future studies using the same task with a different instruction are necessary to examine this speculation about the speed-accuracy trade-off.

The analysis of the bilinguals' performance in the sentence-picture matching task revealed no significant Length of L2 Exposure, WM Capacity effect, or an interaction between these two predictors on the accuracy in responding to the ACT+*le* sentences or the RTs in responding to the target sentences. Therefore, we did not find any evidence suggesting that L1 attrition in processing the interaction between *le/zhe* and ACT verbs would emerge as an effect of Length of L2 Exposure.

Overall, the results from the tasks concerning aspect marking did not suggest that the bilinguals experienced L1 attrition in perceiving, producing or processing the interaction between *le/zhe* and verbs/predicates of different lexical aspects. This is consistent with the IH, which predicts that language structures at the "internal" interfaces, such as the aspect marking phenomenon examined here, are much less likely to be a locus to L1 attrition, if not invulnerable *a priori* to attrition. At least for the bilinguals in this study, who had regular access to their L1, this should be the case (but see Iverson, 2012 for linguistic isolation situations). In Section 6.4, we will further discuss the theoretical implications of these findings.

6.2.L1 attrition in LD binding?

Chapter 5 examined how the bilinguals and the monolinguals interpreted the LD binding reflexive *ziji* in two off-line interpretation tasks, as well as how they processed *ziji* on-line in an eye-tracking task. The analysis revealed significant Group effects or significant interactions involving Group and other predictors for all the three tasks. In this section, we argue that, while the Group effect observed in the pencil-and-paper interpretation task does not necessarily suggest L1 attrition, the other Group effects or interactions involving Group and other predictors do suggest L1 attrition in terms of Schmid & Köpke (2017a)'s definition.

The pencil-and-paper interpretation task was a partial replication of Yuan (1998)'s study, and tested how the participants would interpret *ziji* in embedded clauses and infinitive clauses under Neutral, LD favoured and LOC favoured conditions. The analysis found that the bilinguals resembled the monolinguals in interpreting *ziji* under all but one condition, i.e. infinitive clauses, LOC favoured. Under this condition, the bilinguals were significantly less likely to interpret *ziji* as referring to the local antecedents than the monolinguals. However, an examination of the raw data revealed that this effect was likely to be caused by the very low level of variance in the monolingual group, and may not really indicate a between-group difference. Moreover, given that the bilinguals also dominantly preferred a LOC interpretation of *ziji* and did not choose any LD interpretation under this condition, this Group effect should not be interpreted as a sign of L1 attrition.

The other interpretation task, which was embedded in the eye-tracking task, tested how the participants would interpret *ziji* in embedded clauses when the animacy of the antecedents were manipulated to create Neutral, LD favoured and LOC favoured conditions. The analysis revealed a significant interaction between Group and Condition: Neutral on the probability of choosing a LOC interpretation, which suggests that the bilinguals were significantly less likely to choose a LOC interpretation than the monolinguals under the Neutral condition. This between-group difference indicates a change to the bilinguals' preference in interpreting *ziji* in embedded clauses when both the LD and the local antecedents were animate, so here we treated it as an L1 attrition effect.

While the between-group difference in the probability of choosing a LOC interpretation under the Neutral condition can be regarded as an L1 attrition effect, our data does not suggest that this difference was caused by L2 influence. If the bilinguals' interpretation of *ziji* was influenced by L2 English, they should show a higher level of preference for a LOC interpretation across conditions, as the English reflexives (*himself/herself*) must have a LOC interpretation in embedded clauses. However, this is not supported by our data, which shows that, in comparison to the monolinguals, the bilinguals had a lower level of preference for a LOC interpretation under the Neutral condition, and a similar level of preference for a LOC interpretation under the LD favoured and LOC favoured conditions.

Furthermore, an examination of the raw data also found that, although not statistically significant, the bilinguals differed from the monolinguals in having a higher level of preference for a BOTH interpretation the Neutral condition. While it is possible to explain this difference by speculating that the bilinguals showed indeterminacy in interpreting *ziji* when ambiguity is present and resorted to a "safer" BOTH interpretation, this speculation does not explain why the bilinguals did not always demonstrate a higher level of preference for the BOTH interpretation under the LD and LOC favoured conditions. Moreover, in contrast to our hypothesis, which predicted that a significant interaction between Group and WM Capacity would be present when L1 attrition effects were found, in this task we did not find such significant interaction in the between-group comparison or in the analysis concerning the bilingual group.

The analysis of the results from the eye-tracking task also revealed L1 attrition effects in the on-line processing of the reflexive *ziji* during the later stages. It is significant to highlight this parallelism in light of the calls to bring off-line and on-line methods together in L1 attrition studies (e.g. Iverson & Miller, 2017), precisely because it could have been the case that they did not coincide as has been shown for early bilinguals, or heritage speakers (Puig-Mayenco et al., 2018; Villegas, 2014). In the between-group comparison, we did not observe any Group effect or any interaction between Group and any other predictors in the time windows corresponding to the early processing of *ziji* (i.e. 200-600ms time windows). This finding, along with the descriptive data (see Figure 5.1), suggests that both groups first searched for the local antecedents during the early processing of *ziji*, and thus demonstrated a locality effect. This finding seems to indicate that, even in on-line processing, syntax is not particularly vulnerable to L1 attrition (see also Chamorro et al., 2016b).

However, in the 1000-1200ms time window, which corresponds to the later processing of *ziji*, the analysis found a significant Group x WM Capacity effect. This effect suggests that, unlike the monolinguals, who tended to look less into the local antecedents if their WM capacity were larger, the bilinguals with larger WM capacity tended to look more into the local antecedents during this stage. In line with the IH, which predicts L1 attrition in the language structures at the “external” interfaces, this interaction suggests that the bilinguals showed an L1 attrition effect for the process of resolving *ziji*. More importantly, this interaction also indicates that WM Capacity had different effects on the bilinguals and the monolinguals during the resolution of *ziji*, but not in a straightforward way, as it was not the case that bilinguals with larger WM Capacity would behave more monolingual-like when processing *ziji*.

In the analysis concerning the bilinguals' processing of *ziji*, significant interactions between Length of L2 Exposure and other predictors (i.e. Condition, WM Capacity) were observed for two time windows. Firstly, a significant interaction between Length of L2 Exposure and Condition: Neutral was found for the 200-400ms time window, which corresponds to the early processing of *ziji*. This interaction suggests that, under the Neutral condition, the bilinguals tended to look less into the local antecedents as their length of L2 exposure increases. However, it remains uncertain whether this interaction actually indicates that longer L2 exposure would lead to L1 attrition in the early processing of *ziji*, as the same interaction was not significant in the following 400-600ms time window, which also corresponds to the early processing of *ziji*. Future studies are needed to further examine this issue.

Secondly, a significant interaction between Length of L2 Exposure, WM Capacity and Condition was observed for the 600-800ms time window. This three-way interaction suggests that, while the bilinguals with longer length of L2 exposure tended to look less into the local antecedents if their WM Capacity were larger, WM Capacity did not have such an effect on those bilinguals with shorter length of L2 exposure. Moreover, this effect varied according to Condition, and it was much larger under the LOC favoured condition than under the other two conditions. This interaction indicates that individual differences in WM Capacity played a role in explaining variations in processing *ziji* within the Mandarin-English bilinguals, especially in the bilinguals with longer L2 exposure. To the extent that it is fair to assume length of L2 exposure as a proxy for L2 proficiency and/or opportunity for using the L2, the interaction makes sense in that the bilinguals with longer L2 exposure would have more competition to deal with, and this competition seems to tax the WM system and/or other executive control/functions.

In summary, the results from the tasks concerning the interpretation and processing of *ziji* suggest L1 attrition in the final interpretation and later stage processing of *ziji*, and this is consistent with the predictions of the IH. Moreover, in line with our hypothesis, individual differences in WM Capacity also seemed to have played a role in explaining L1 attrition and the variation within the bilinguals, but the relationship between WM Capacity and variation in L1 attrition does not seem to be as straightforward as we hypothesized. In the next section, we will further discuss this issue.

6.3.Individual differences in WM capacity and variation in L1 attrition

In Chapter 2, we argued that individual differences in cognitive abilities, such as WM capacity, might interact with other factors (e.g. the status of bilingualism, length of L2 exposure) and explain variation in language comprehension and processing among the bilinguals who showed L1 attrition. We also speculated that, as a larger WM capacity suggests more cognitive resources available or better ability to suppress interference during language processing, the bilinguals with larger WM capacity would behave more monolingual-like and show smaller L1 attrition effect than those with smaller WM capacity when comprehending/processing language structures for which L1 attrition effects were observed.

However, as discussed in the previous sections, the findings in this study only partially support our hypothesis. On the one hand, in the eye-tracking task, we observed that WM Capacity interacted with Group during the later stage processing of *ziji*, and had different effects on the bilinguals and the monolinguals. We also observed that WM Capacity interacted with Length of L2 Exposure and Condition, and that WM Capacity had different effects on the bilinguals who differed in their

Length of L2 Exposure. These findings suggest that individual differences in WM Capacity played a role in explaining variation in L1 attrition effects found in the processing of *ziji*, but not in the way we had speculated.

As shown in Figure 5.4, the WM Capacity x Group interaction during the later stage processing of *ziji* suggests that the bilinguals with larger WM capacity tended to behave more like the monolinguals with smaller WM capacity in processing *ziji* at this stage, but it also suggests that the bilinguals with smaller WM capacity tended to behave more like the monolinguals with larger WM capacity. While our speculation is consistent with the former finding, it does not explain the latter one. Due to the design of our stimulus sentences, in which the 1000-1200ms time window may either correspond to the semantic/pragmatic processing of *ziji* or the wrap-up of the whole clause containing *ziji*, it is very difficult to confirm why and how WM Capacity had different effects on each group's processing of *ziji* at this stage. Future studies with refined methodology (e.g. putting more intervention phrases between the reflexive and the word which triggers the wrap-up effect) should look into this issue.

In the analysis concerning the bilinguals' processing of *ziji*, a significant interaction between Length of L2 Exposure, WM Capacity and Condition in the 600-800ms time window suggests that individual differences in WM capacity had greater effects on the processing of *ziji* in LOC favoured conditions among the bilinguals with longer L2 exposure. While this finding supports our hypothesis that individual differences in WM capacity may explain variations in L1 attrition within the bilinguals, the nature of this effect requires further investigation. Firstly, in this study, only three individuals had more than 15 years' exposure to L2 English, so a larger cohort of participants are needed to validate this effect. Secondly, it is worth further examining why this WM Capacity x Length of L2 Exposure interaction had larger impact on the LOC favoured condition rather than the other two conditions - one possible reason is that, under this condition, the bilinguals with larger WM Capacity were able to search for the inanimate LD antecedents with raised level of accessibility, whereas those with smaller WM Capacity were not able to do so at this stage.

On the other hand, some of the findings in this study do not support our hypothesis. For example, in the interpretation task embedded in the eye-tracking task, the bilinguals showed an L1 attrition effect in that they were significantly less likely to choose a LOC interpretation than the monolinguals under the Neutral condition. However, the analysis did not reveal a significant WM Capacity effect or an interaction between WM Capacity and other predictors, despite that the resolution of *ziji* and the nature of this task (i.e. requiring simultaneous storage and processing) are memory taxing. Therefore, individual differences in cognitive abilities may not always be involved in explaining variations in L1 attrition. Unfortunately, based on our findings, it is impossible to further explore the source of this L1 attrition effect.

Within the bilinguals, we also observed that the bilinguals with longer L2 exposure tended to look less into the local antecedents under the Neutral condition, which suggests that the bilinguals with longer L2 exposure might have shown L1 attrition in the early syntactic processing of *ziji* under this condition. The analysis did not reveal a significant WM Capacity effect or an interaction between WM Capacity and other predictors either. We speculate that, this finding suggests that once L1 attrition effects at the syntactic level emerged, individual differences in cognitive abilities, such as WM Capacity, may no longer play a role in explaining the variations among the bilinguals. However, as discussed in the previous section, this attrition effect still needs further validation, and so is this speculation.

Based on the above findings, it is reasonable to conclude that individual differences in WM capacity had played a role in explaining some variations in language processing, but the relationship between WM capacity and variations in L1 attrition was not straightforward. As a preliminary exploration of this relationship, this study has not been able to provide a clear picture, but it has achieved its objective in showing that the important effect of individual differences in cognitive abilities on L1 attrition. In the next section, we discuss the implications of our findings for the theoretical frameworks of L1 attrition.

6.4.Implications for theoretical frameworks of L1 attrition

In Chapter 2, we presented and discussed a selection of theoretical frameworks of L1 attrition, including the IH (Domínguez, 2013; Sorace, 2011), the Feature Reassembly Hypothesis (Lardiere, 2005, 2009) and the Unified Competition Model (MacWhinney, 2012, 2018). In this section, we first argue that the bilinguals' performance in perceiving/interpretation/producing/processing aspect marking and reflexive binding could not always be explained by Sorace's version of the IH. Then we discuss whether the alternative theoretical frameworks of L1 attrition could explain the findings which were not explained by Sorace's version of the IH. Following that, we argue that L1 attrition research should not be limited to a formal/usage-based approach, and it should move towards a multi-dimensional theoretical framework of L1 attrition. Finally, we discuss Schmid & Köpke (2017a)'s continuum approach to defining L1 attrition in terms of our findings.

6.4.1. Sorace (2011)'s version of the Interface Hypothesis

Sorace (2011)'s version of the IH predicts that language structures at the "external" interfaces are more vulnerable to L1 attrition than those at the "internal" interfaces, and our findings are generally in line with this prediction - we observed no L1 attrition effects on the perception, production and processing of perfective and durative aspect marking in simple declarative sentences, but observed L1 attrition effects on the interpretation and processing of *ziji*. However, the IH does not seem capable of providing an explanation for some of our findings. For example, why did the bilinguals only show L1 attrition effects for interpreting and processing *ziji* under the Neutral condition in the eye-tracking task? The IH argues that a potential source of L1 attrition is that the bilinguals may be less efficient in integrating syntactic and other types of information during language processing. If this is true, then we should observe L1 attrition effects for interpreting and processing *ziji* across all conditions, since under every condition the resolution of *ziji* would require the integration of syntactic, semantic and pragmatic information (unless we consider animacy to be a pure lexical-semantic factor, as Chamorro et al., 2016b did for Spanish); yet our findings are against this prediction.

The raw data from the interpretation task embedded in the eye-tracking task suggest that, under the Neutral condition, the bilinguals showed a monolingual-like level of preference for an LD interpretation of *ziji*, as well as a significantly lower level of preference for an LOC interpretation and a non-significant higher level of preference for a BOTH interpretation. Some studies examining the IH tended to interpret such behaviour as indeterminacy of interpreting specific language structures (e.g. Domínguez, 2013; Tsimpli et al., 2004), but our data do not suggest so - why should the bilinguals only show indeterminacy in interpreting *ziji* under the Neutral condition, rather than across all conditions?

Realizing the ambiguity of, or showing non-monolingual-like interpretation of certain language structures, such as the binding of the reflexive *ziji*, does not necessarily mean indeterminacy, as these language structures are essentially ambiguous. Rather than that, such behaviours may indicate that bilinguals have better metalinguistic awareness, or at least a better ability of detecting ambiguity than the monolinguals. L1 attrition studies need to be cautious when interpreting such seemingly “indeterminate” behaviours, and further examine this issue - for example, we could measure the participants’ level of confidence in choosing a BOTH interpretation when interpreting ambiguous sentences, and confirm whether such behaviour was caused by “indeterminacy” or better ability to detect ambiguity; we could also measure the participants’ metalinguistic abilities and analyze the relationship between this factor and the probability of choosing a BOTH interpretation when interpreting ambiguous sentences.

Regarding the findings that the bilinguals only showed L1 attrition effects for interpreting *ziji* under the ambiguous Neutral condition, one may argue that the “external” syntax-discourse interface was involved only when both antecedents were animate, but not so when one of the antecedents was inanimate and the sentences were “disambiguated”. However, we argue that this argument is not supported by our data. As Table 5.2 (see Section 5.2) has shown, even under the LD and LOC favoured conditions, the participants did not unanimously interpret *ziji* as referring to the favoured antecedent. If only “internal” syntax-semantics interface was involved, then why would the participants violate the animacy constraint and choose the unfavoured interpretation? Therefore, we speculate that some “external” pragmatic interface must have been involved when resolving the reflexive *ziji*, even if the sentences were "disambiguated" by animacy.

Moreover, Sorace (2011)'s version of the IH speculated that L1 attrition effects might be caused by bilinguals' reduced efficiency in integrating multiple sources of information when processing language structures at the "external" interfaces, and this speculation leads to the prediction that bilinguals with more cognitive resources available would behave more monolingual-like in comprehending and/or processing language structures at the "external" interfaces. However, our findings were not consistent with this prediction, and the speculation about the (cognitive) source of L1 attrition is not supported.

If we follow Sorace (2011) and assume that more available cognitive resources lead to less L1 attrition effects, in this study we would expect the bilinguals with larger WM capacity to behave more monolingual-like in processing and interpreting *ziji*, and those with smaller WM capacity less so. However, as discussed in Chapter 5 and the last two sections, the Group x WM Capacity interaction observed in the on-line processing of *ziji* at later stages did not conform to such predictions - the bilinguals with smaller WM capacity behaved more like the monolinguals with larger WM capacity in processing *ziji* at this stage. Furthermore, as just discussed above, in the interpretation task where the bilinguals showed L1 attrition, we did not even observe a significant effect of WM Capacity. These findings certainly do not rule out the possibility that some L1 attrition effects were caused by the reduced efficiency in integrating multiple sources of information, but they do suggest that the relationship between cognitive factors and L1 attrition was not as clear as Sorace (2011) has speculated.

In the following section, we will discuss whether the findings that cannot be explained by Sorace's version of the IH could be explained by the other theoretical frameworks discussed in Chapter 2. More specifically, we will discuss whether Domínguez (2013)'s version of the IH and the Feature Reassembly Hypothesis (Lardiere, 2005, 2009) could provide a satisfactory explanation of our findings. We will not discuss whether the Unified Competition Model (MacWhinney, 2012) could explain our findings, because our design was not based on the assumptions of this theoretical framework and did not measure any of the critical factors; however, later we will argue that a holistic theoretical framework like the Unified Competition Model might be more suitable for research into potential links between L1 attrition and multiple factors, e.g. the linguistic properties of a particular language structures and cognitive abilities, as we have investigated in this study.

6.4.2. Alternative theoretical frameworks to Sorace (2011)'s version of the IH

Domínguez (2013)'s modified version of the IH predicts that the language structures which "require checking for contextual appropriateness in the selection of linguistic outputs" (p. 99) are vulnerable to L1 attrition, and these language structures do not have to be at the "external" interfaces. This version of IH also predicts that the resolution of *ziji* is vulnerable to L1 attrition, because a speaker has to check for contextual appropriateness before deciding the antecedent of *ziji*. Like Sorace's version of the IH, this prediction is consistent with our findings.

Moreover, this version of the IH may better explain why the bilinguals only showed L1 attrition effects under the Neutral condition - under the other two conditions, the bilinguals may check for contextual appropriateness based on the information presented in the target sentences and therefore behaved monolingual-like.

In contrast, under the Neutral condition, the bilinguals have to check for contextual appropriateness based on the L1 input (i.e. how *ziji* are normally interpreted in Neutral sentences) they received; given that the L1 input received by the bilinguals may contain different distribution on how *ziji* is interpreted in Neutral sentences in comparison to the L1 input received by the monolinguals, their results of checking for contextual appropriateness might differ, and such differences were reflected in the output (i.e. the final interpretation of *ziji*).

Unlike Sorace's version of the IH, Domínguez (2013)'s version of the IH also predicts L1 attrition in aspect marking. As introduced in Chapter 2, aspect marking is not the only way of expressing temporal information, and it is reasonable to assume that bilinguals with L1 Mandarin would have to check for contextual appropriateness when deciding whether to use aspect marking, temporal adverbials or both ways to express temporal information. Unfortunately, our data cannot suggest whether this prediction is correct or not.

In the tasks concerning perfective and durative aspect marking, we did not observe L1 attrition effects among the bilinguals. However, as we only looked into the interaction between perfective/durative aspect marking and different lexical aspects in simple declarative sentences, and the bilinguals rarely have to check for contextual appropriateness under this setting, we can only conclude that the bilinguals did not show L1 attrition in aspect marking at this very "basic" level. In order to examine whether Domínguez (2013)'s claim applies to aspect marking in Mandarin, future studies should investigate natural speech by the bilinguals with L1 Mandarin, as well as how these bilinguals perceive, produce and process aspect marking in discourses.

In contrast to the IH, the Feature Reassembly Hypothesis predicts L1 attrition based on the similarity/distinction in grammatical features encoded in the L1 and L2 lexical heads. Although this study was not framed within this theoretical framework, some of our findings suggest that much further work has to be done before this theoretical framework can be applied in L1 attrition research, at least equally for all domains. Moreover, it is not clear how the Feature Reassembly Hypothesis is or can be predictive as opposed to provide an *a posteriori* metric to inject a sense of specific formalism into the description of the observed differences.

For instance, in this study, we observed that the bilinguals showed L1 attrition effects in interpreting *ziji*, but such attrition effects do not seem to result from the similarity or distinction in grammatical features encoded in the English and Mandarin reflexives. Theoretically, we could follow the assumptions of the Feature Reassembly Hypothesis, and assume that the Mandarin reflexive *ziji* has a feature of [+LD] in its feature bundle, and that the English reflexive *himself/herself* has a feature of [-LD] in its feature bundle. If the bilinguals' L1 Mandarin was influenced by L2 English, and started to reassemble the feature bundle of *ziji* by changing the [+LD] feature to [-LD], the bilinguals should exhibit a lower level of preference in interpreting *ziji* as bound to the LD antecedents, and a higher level of preference in interpreting this reflexive as bound to the local antecedents; in other words, the features associated with the anaphoric *ziji* should be subject to L1 attrition under the influence from the features associated with the anaphoric pronouns in English, which do not allow for LD binding. However, as shown in Chapter 5, the bilinguals did not show such a pattern in interpreting *ziji*, suggesting that we cannot predict L1 attrition simply based on the similarity/distinction between L1 and L2.

Furthermore, if we were to extend Lardiere (2005; 2009)'s argument that the distinction between the feature bundles encoded in an L1 and an L2 lexical head could facilitate the process of feature re-assembly, we could also argue that the [+LD] encoded in *ziji* and the [-LD] encoded in *himself/herself* were distinct enough, and such distinction could prevent L1 attrition of reflexive binding in Mandarin. If this was the case, then we should not observe any L1 attrition effects in the interpretation task, yet this is against our findings. Therefore, we argue that, in its current form, the Feature Reassembly Hypothesis cannot provide satisfactory explainations of our data, or making useful predictions.

In general, the empirical data collected in this study suggest that the various existing theoretical frameworks as used here which aim to correctly predict L1 attrition effects are likely to be observed for some language structures (e.g. reflexive binding) rather than other structures (e.g. aspect marking), yet none of them are precise enough to predict or explain the specific pattern of L1 attrition we have observed (e.g. L1 attrition effects only observed for interpreting *ziji* under Neutral conditions). More importantly, while we have demonstrated that individual differences in WM capacity interacted with other predictors and played a role in explaining the monolingual-bilingual differences and the variations within the bilinguals, the existing theoretical frameworks either did not correctly predict this relationship (Sorace, 2011) or did not recognize the role of cognitive abilities in L1 attrition (e.g. Domínguez, 2013; Lardiere, 2005, 2009; MacWhinney, 2012).

In the following section, we will argue that, in order to accurately explain and predict L1 attrition phenomena, theoretical frameworks should take a cross-disciplinary approach to L1 attrition, rather than limiting themselves to a formal linguistic or usage-based approach.

6.4.3. Towards a multi-dimensional theoretical framework of L1 attrition

In the previous sections, we have shown that the theoretical frameworks which predicts L1 attrition effects based on the linguistic features of language structures and the similarity/distinction between L1 and L2 cannot always provide satisfactory explanations for our findings. In our study, we have shown that the observed L1 attrition effects in reflexive binding were usually explained by the interaction between multiple factors, and this suggests that L1 attrition was not simply determined by the linguistic features of a particular language structure, the similarity between L1 and L2, or individual differences in cognitive abilities. Given that the existence and magnitude of L1 attrition can vary according to various linguistic and extra-linguistic factors, it seems that any theoretical frameworks only taking a formal or usage-based approach would be insufficient in explaining and predicting L1 attrition.

It seems to us that, in order to capture how different linguistic and extra-linguistic factors determine the outcomes of L1 attrition, holistic theoretical frameworks, such as the Unified Competition Model, might be better than those theoretical frameworks which only based their assumptions on formal linguistics. As discussed in Chapter 2, the Unified Competition Model might be a promising theoretical framework for L1 attrition research, as it does not only consider the roles of linguistic factors (i.e. cue availability, reliability and transfer), but also the roles of neurobiological factors (i.e. entrenchment) and extra-linguistic factors (e.g. resonance, decoupling). Such models do not only allow us to explain and predict the differences between monolinguals and bilinguals, but also the differences within the bilinguals.

However, the Unified Competition Model largely ignored the role of individual differences in cognitive abilities in explaining L1 attrition. In our study, we have shown that individual differences in WM capacity interacted with other linguistic (e.g. animacy of the antecedents) and extra-linguistic factors (e.g. length of L2 exposure), and played a role in explaining the monolingual-bilingual differences and the variations within the bilinguals. Although we have not been able to reveal a clear relationship between these factors in this study, our findings do lead us to think that a competent theoretical framework should explicitly explain the relationship between individual differences in cognitive abilities and the outcomes of L1 attrition.

Furthermore, although not the focus of this study, we think that a rarely discussed dimension - which we term - as "personal sociolinguistic experience" here - should also be included as a factor in any approaches to L1 attrition. Studies like Domínguez (2013), Domínguez & Hicks (2016), Iverson (2012) and Iverson & Miller (2017) highlighted that L1 attrition in certain language structures (e.g. those Sorace, 2011 assumed to be at the "internal" interfaces) can only be revealed when bilinguals with special linguistic experience, such as having regular access to the L1 of a different variety or having no access to the L1 at all, are concerned. Such findings suggest that "personal sociolinguistic experience" can also affect the outcomes of L1 attrition, both at the individual and group level. Furthermore, in our opinion, this factor can also interact with linguistic and cognitive factors, and shape the outcomes of L1 attrition. In the future, it might prove that a theoretical framework recognizing the roles of these different factors will enable us to explain a greater range of L1 attrition data, and precisely predict the outcomes of L1 attrition both at the individual and group level.

It should be noted though, while it might be easy to come up with numerous factors that may affect L1 attrition if a holistic approach to L1 attrition were adopted, it will always be difficult to experimentally examine the interaction between these many factors, and how such complex interactions lead to the observable outcomes of L1 attrition. This methodological problem can be resolved with the advance of techniques (e.g. collecting and analyzing abundant data), and cross-disciplinary methods like combining computational modelling and linguistic experiments may also be helpful. Although the issue of how to resolve such a problem is far beyond the scope of this thesis, it is certainly worth thinking for researchers interested in bilingualism.

6.4.4. Revisiting the definition of L1 attrition

In this thesis, we adopted Schmid & Köpke (2017a)'s continuum approach to defining L1 attrition, which treats all bilingualism-caused changes to L1 processing and representation as L1 attrition. We agree with Schmid & Köpke (2017a) that abandoning the artificial distinction between attrition in "transient" processing and "permanent" representation can better capture some potential properties of L1 attrition, such as that L1 attrition effects can be reversed (see Chamorro et al., 2016a), and this is why we adopted this definition in our study. However, we also recognize that this definition indeed equals L1 attrition to bilingualism, and based on our findings, we question whether "L1 attrition" is always an appropriate term for future research on this topic.

Many of the commentaries on Schmid & Köpke (2017a)'s continuum approach to defining L1 attrition have pointed out that, the term "attrition" naturally entails the meaning that something has been lost or weakened (see Schmid & Köpke, 2017b), and the term "L1 attrition" may not be a suitable term for summarizing the concept proposed by Schmid & Köpke (2017a). This is especially the case when the bilingualism-caused changes to L1 were not disadvantageous, or even advantageous to the bilinguals.

For example, in our study, we have observed that, in comparison to the monolinguals, the bilinguals seemed to be more capable of recognizing the ambiguity of *ziji* when interpreting this reflexive under the Neutral condition, but not under the LD or LOC favoured conditions. This finding can hardly be interpreted as indeterminacy of interpreting *ziji*, and it is more likely to be a bilingual advantage in recognizing ambiguity. However, as we adopted Schmid & Köpke (2017a)'s definition, we still referred this change to L1 as L1 attrition effect. While this is logically coherent, "L1 attrition" may still not be the best term for describing such bilingualism-caused changes to L1 - even some phrase like "bilingual effects on L1" can be more precise.

It seems to me that, Schmid & Köpke (2017a)'s definition of L1 attrition is in fact a challenge to the traditional view that a bilingual has to lose some of the L1 processing ability or knowledge if s/he experienced L1 attrition, and this new definition encourages researchers to look into the different facets of bilingualism effects on a bi-/multilingual's L1. With this new definition of how bilingualism affects an L1, maybe it is necessary for us to find a more precise term for replacing the old "L1 attrition".

6.5.Other issues

In the last four sections, we discussed how our findings addressed the research questions, and the implications for the theoretical frameworks discussed in Chapter 2. In this section, we briefly discuss a few issues not covered in the discussion above. More specifically, we discuss the linguistic and methodological implications of the results found in the tasks concerning aspect marking, and the difficulty in relating language processing to language comprehension in L1 attrition research.

In Chapter 4, we have shown that the bilinguals did not differ from the monolinguals in performing the tasks concerning aspect marking. However, it is worth noting that both groups demonstrated similar variances in the cloze task and the sentence-picture matching task, and such variances may be worth further investigating in experimental linguistic studies on Mandarin Chinese.

In the cloze task, our participants showed a much higher level of variance in the accuracy in supplying *zhe* than in supplying *le*; meanwhile, they showed a much higher level of variance in the accuracy in responding to the ACT+*le* sentences than the ACT+*zhe* sentences in the sentence-picture matching task. We speculate that, in the cloze task, the participants tended to think that the sentences containing [+Dynamic] and [-Telic] verbs could deliver an imperfective reading without being overtly marked with *zhe*, and therefore "zero aspect" marked these sentences in some cases (see the discussion in Xiao & McEnery, 2004, pp. 236–240). This finding may provide supportive evidence for the "zero aspect marking" phenomenon, and it is worthy studying in the future.

In the sentence-picture matching task, we also found that both the bilinguals and the monolinguals showed higher level of variance in the accuracy in responding to the ACT+*le* sentences than in responding to the ACT+*zhe* sentences. The higher level of variance in the accuracy in responding to the ACT+*le* sentences in the sentence-picture matching task may imply that the participants adopted a "good enough" strategy (see Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007) for processing aspect marking under a time pressured setting. These participants might have constructed an imperfective reading for all the target sentences upon hearing the

ACT verbs, and tended to match the sentences to the pictures depicting ongoing events by this time. However, the presence of the perfective *le* would require the participants to revise the original readings for the sentences, whereas the presence of the durative *zhe* would not. Under a time pressured setting, some participants might have not been able to revise the original readings, and made more errors in responding to the ACT+*le* sentences, and this fact has led to the higher level of variance in accuracy.

Furthermore, the observation that individual differences in WM capacity had a significant effect on the participants' accuracy in responding to the ACT+*le/zhe* sentences has a methodological implication for L1 attrition research. Depending on the demand of the tasks, individual differences in cognitive abilities may constrain the performance in these tasks. For example, in this study, the participants' performance in the sentence-picture matching task may be constrained by their WM capacity, and this is supported by our analysis. If we did not include WM Capacity as a predictor when interpreting the results from the sentence-picture matching task, it is likely that we will incorrectly interpret the observed difference between the two groups as a sign of L1 attrition. Therefore, future L1 attrition research should consider the cognitive demands of experimental tasks, and tease apart the influence of individual differences in cognitive abilities and that of L1 attrition.

Now we turn to the difficulty in relating language processing to language comprehension in L1 attrition research. In various literature, researchers suggest that we should investigate both off-line comprehension/production and on-line processing when studying L1 attrition, so that we will be able to uncover more potential L1 attrition effects (e.g. Iverson & Miller, 2017; Schmid & Köpke, 2017a; Sorace, 2012). In the present study, we employed both off-line and on-line tasks as suggested, and we have observed L1 attrition effects in the on-line processing of *ziji*. Such effects cannot be revealed if only traditional off-line tasks were used.

However, while the use of on-line tasks have been successful in unravelling the L1 attrition effects in on-line processing, it still seems difficult to explain how such difference in processing lead to the differences we observed in the off-line interpretation tasks. For instance, we have found that, under the Neutral condition, the bilinguals' proportion of looks into the local antecedents tended to decrease if their WM capacity were larger during the 1000-1200ms time window when processing *ziji*, whereas the monolinguals showed the reversed pattern. In the interpretation task, we also observed that the bilinguals were less likely to choose an LOC interpretation of *ziji* under the Neutral condition. Do these findings suggest that the differences in the processing during the 1000-1200ms time window led to the differences in the interpretation task? We cannot provide an answer to this question, despite that we had both on-line and off-line data about such behaviour.

In order to understand the relationship between on-line processing and off-line comprehension/production, simply combining off-line tasks with on-line tasks in L1 attrition research is far from enough. The tasks we used in this study could reveal some quantitative differences between the bilinguals and the monolinguals, but without on-line processing tasks which look into the qualitative aspects of on-line processing, we may not be able to confirm how such differences in on-line processing lead to the differences in off-line comprehension/production. Perhaps future research on L1 attrition could qualitatively analyze the on-line processing of language structures among bilinguals (e.g. analyzing the scan paths during reading, see von der Malsburg & Vasishth, 2012), and provide more insights into how and why variations in on-line processing may or may not lead to the final, off-line language output.

6.6.Limitations of the present study, and implications for future research

As shown in the previous sections, the present study has made contributions to L1 attrition research by finding empirical evidence that has yet been precisely explained by the existing theoretical frameworks of L1 attrition, as well as exploring the relationship between individual differences in cognitive abilities and variations in L1 attrition. We also argued that future research on L1 attrition should pursue a multi-dimensional theoretical framework of L1 attrition, and consider a new term that could better describe the bilingualism-caused changes to L1. In this section, we point out the limitations of the present study, and suggest how future studies may make amendments to these problems.

First, in this study, we had limited ability to generalize our findings, especially when the effect of Length of L2 Exposure was concerned, because we only had a small sample size of Mandarin-English bilinguals, and their Length of L2 Exposure was not well balanced - most of the bilinguals had less than 15 years' exposure to L2 English, while the other three had 16, 29 and 30 years' L2 exposure. This is due to the difficulty of finding qualified bilinguals. In this study, we assumed that the qualified bilinguals should have at least 7 years' exposure to L2 English and use L2 English on a daily basis. Therefore, we only included bilinguals who were working professionals or students with very limited or no acess to Mandarin when working or studying, and this strict requirement led to our very small sample size. On reflection, other types of study design could have been adopted, which would include participants with potentially shorter length of L2 exposure, less dominant use of L2, less L2 proficiency. This would enable us to further examine whether L1 attrition in processing and representation is a continuum, as claimed by Schmid & Köpke (2017a); future studies should definitely attempt to do so.

Second, in this study we only looked into a limited range of language structures, and this again limited the generalizability of our findings. For instance, this study only concerned the interaction between *le/zhe* and different lexical aspects in simple declarative sentences, so our findings cannot suggest whether the whole aspect marking phenomenon in Mandarin Chinese are vulnerable to L1 attrition or not. Due to this limitation, we have not been able to fully examine the modified version of IH as proposed by Domínguez (2013). Future studies should look into a wider range of language structures (e.g. aspect marking in natural speech and discourse) to overcome this limitation.

Third, the design of some experimental tasks used in this study prevented us from further understanding the relationship between different predictors and the observed L1 attrition. For example, in Section 6.2 and 6.3, we argued that the design of the stimulus sentences prevented us from further exploring whether and how the interaction between WM Capacity and Group had affected the semantic/pragmatic processing of *ziji* and/or the wrap-up of the whole clause at later stages. However, this limitation is relatively more difficult to overcome, as making accurate assumptions about when syntactic/semantic/pragmatic processing happens and how long it lasts largely depend on the findings of psycholinguistic studies focusing on the on-line processing of languages.

Moreover, this study used different methodologies to test both representation and processing of aspect marking and reflexive binding, and therefore aimed to be as complementary as possible; in actuality, the results from the tasks concerning the two target structures were not directly comparable. This was especially the case when we tested the processing of aspect marking and that of reflexive binding – for aspect marking, we used a sentence-picture matching task to examine the RTs of responding to this task, and this task more correctly speaking measured off-line processing of aspect marking (see Section 2.5.3); in contrast, the eye-tracking task used for examining the processing of *ziji* measured more conventionally understood on-line aspects of processing. When designing the tasks, we attempted to make the results as comparable as possible by incorporating eye-tracking into the sentence-picture matching task, so that we could directly examine if the bilinguals would differ from the monolinguals in the speed/pattern during the on-line processing of aspect marking and reflexive binding. Unfortunately, we did not manage to do so, because the stimulus sentences used in the sentence-picture matching tasks ended so soon after the aspect markers were presented that the eye-tracker (with a 60Hz sampling rate) could not collect enough eye movements for any meaningful ananlysis. In future studies, such limitations could be easily overcome by using equipments with better specifications.

And fourth, the design of this study was framed within the IH and its aim was to explore the role of individual differences in cognitive abilities in L1 attrition, so it only examined the roles of a limited range of predictors, such as language structures, WM capacity and length of L2 exposure. However, as we argued in this chapter, future studies should take a more holistic approach and therefore should measure more potentially influential factors, such as the level of decoupling, and that would enable us to achieve a more comprehensive understanding about L1 attrition and bilingualism in general. Given that many different factors may interact with each other and affect the outcomes of L1 attrition, it would also be worthy trying different methods, such as computational modelling, in future studies on L1 attrition.

Finally, in this study we took a capacity approach to WM and language processing (Baddeley, 2012; Cowan, 2005), but recent psycholinguistic research has argued that variations in language processing/comprehension is likely to be constrained by individual differences in the abilities to inhibit interference and/or weighing cues (e.g. Lewis et al., 2006; Cunnings, 2017), rather than individual differences in what we called “WM capacity”. In future L1 attrition research, it is worth discussing whether the aspects of language processing/comprehension can be better explained by interference-based models of language processing, and it is also worth investigating whether individual differences specifically in inhibitory control, rather than in a more general sense of WM capacity, is a better predictor of variation in L1 attrition and bilingual language processing in general.

7. Conclusion

With the objectives of exploring whether two relatively less investigated language structures - i.e. aspect marking and LD binding in Mandarin - are subject to L1 attrition, and whether individual differences in cognitive abilities, such as WM capacity, can explain the potential variation in L1 attrition, the present study investigated the perception, production/interpretation and processing of perfective and durative aspect marking and the LD binding property of *ziji* among a group of 14 adult, late sequential Mandarin-English bilinguals who lived in the UK for an average of 13 years, and looked into whether and how individual differences in WM capacity explained the variation within the bilingual group.

In Chapter 2, we critically reviewed a selection of theoretical frameworks of L1 attrition, including the IH (Domínguez, 2013; Sorace, 2011), Lardiere (2005, 2009)'s Feature Reassembly Hypothesis, Paradis (2007)'s Activation Threshold Hypothesis, and MacWhinney (2012)'s Unified Competition Model. We pointed out that, among all these frameworks, Sorace (2011)'s version of the IH, which predicts that L1 attrition is likely to happen with language structures involving syntax and extra-linguistic domains (e.g. pragmatics, discourse), is the most testable framework, and thus we framed this study within Sorace (2011)'s version of the IH. We also pointed out that, at present, the other frameworks either did not capture the known properties of L1 attrition (i.e. Paradis, 2007), could not provide clear predictions (e.g. Lardiere, 2005, 2009; MacWhinney, 2012), or was difficult to examine (e.g. MacWhinney, 2012).

More importantly, we argued that, although the reviewed theoretical frameworks assumed that L1 attrition was associated with co-activation and competition between

L1 and L2, and therefore implied that individual differences in cognitive abilities might explain the potential variation in L1 attrition, all these frameworks largely failed to explicitly explain this relationship between cognitive abilities and L1 attrition, possibly because of lack of empirical evidence. Therefore, this study attempted to fill in this gap by investigating whether and how individual differences in WM capacity might play a role in explaining variations in L1 attrition.

In line with the IH, we predicted that the aspect marking phenomenon tested in this study is unlikely to be subject to L1 attrition, because it only involves the interaction between syntax and lexical semantics and is therefore at the “internal” interface; in contrast, we predicted that the LD binding property of *ziji* is likely to be subject to L1 attrition, because it involves syntax, semantics and pragmatics and is therefore at the “external” interfaces. We also predicted that individual differences in WM capacity would explain the variance within the bilinguals when L1 attrition was observed, either as a single fixed effect or in interaction with other factors (e.g. Length of L2 Exposure). Regarding how individual differences in WM capacity would affect the outcomes of L1 attrition, we hypothesized that those bilinguals with larger WM capacity would behave more monolingual-like, but those with smaller WM capacity would behave less so.

In Chapter 3, we introduced the methodology of this study. The methodology was relatively novel, as it combined off-line and on-line tasks to investigate the potential L1 attrition effects. We employed a questionnaire (adapted from Montrul, 2012) to collect the Mandarin-English bilinguals’ language background, an abridged HSK-3 test to test if the bilinguals had sufficient Mandarin listening and reading skills for this study, and a digits-back recall task to measure their WM capacity. In order to assess their perception, production and processing of perfective and durative aspect marking, an acceptability judgement task, a cloze task and a sentence-picture matching task (adapted from Yap et al., 2009) were used respectively. With respect to the interpretation and processing of the LD binding reflexive *ziji*, we used a pencil-and-paper interpretation task (adapted from Yuan, 1998), and a visual world eye-tracking task combined with an interpretation task.

The results from the tasks concerning aspect marking did not suggest L1 attrition in perceiving, producing or processing this language structure. This finding was generally consistent with the IH, but we also noted that it was still possible that L1 attrition can happen with aspect marking, especially when aspect marking in discourse was concerned. Future research could look more into this issue, and doing so will enable us to examine alternative theoretical frameworks' predictions about L1 attrition (e.g. Domínguez, 2013).

Meanwhile, in the tasks concerning the LD binding reflexive *ziji*, we observed that the bilinguals showed L1 attrition effects for interpreting *ziji* under the ambiguous Neutral condition, as well as for the later stage of processing *ziji* on-line. This finding was also in line with the IH, as it suggests that language structures at "external" interfaces are likely to be subject to L1 attrition. Furthermore, we also found that, during the later stage processing of *ziji*, the between-group difference varied according to individual differences in WM capacity. This observation partially supports our hypothesis about individual differences in WM capacity and variation in L1 attrition, but the fact that the bilinguals with smaller WM capacity were more similar to the monolinguals with larger WM capacity suggests that the relationship between cognitive abilities and variation in L1 attrition was not as straightforward as we had speculated.

Moreover, in contrast to our hypothesis that individual differences in WM capacity would explain the variance within the bilinguals when L1 attrition was observed, we failed to find any significant WM capacity effect or interaction involving WM capacity in several cases where L1 attrition effect was present (e.g. the interpretation task embedded in the eye-tracking task, see discussion in Section 6.3). Such evidence indicates that, although individual differences in WM capacity may indeed be associated with variation in L1 attrition, much further research on this topic is needed to unravel a clear picture of this complex relationship.

In Chapter 6, we discussed the theoretical implications of our findings. We argued that, although Sorace (2011)'s version of the IH was correct in predicting the general pattern of L1 attrition - that is, LD binding was more vulnerable to L1 attrition than aspect marking as tested in this study, this framework did not sufficiently explain some of our data, such as the bilinguals only showed attrition under the ambiguous Neutral condition when interpreting *ziji*. We also argued that the Feature Reassembly Hypothesis (Lardiere, 2005, 2009) was not suitable for explaining our data, as the observed attrition effects could not be explained by the similarity/distinction between L1 Mandarin and L2 English.

Given that the formal approaches to L1 attrition had limited explanatory coverage of our data, we further argued that a holistic, multidimensional approach might be more suitable for L1 attrition research; a multidimensional theoretical framework will at least need to recognize the roles of linguistic factors (e.g. the linguistic properties of language structures, input), neurobiological factors, sociolinguistic factors and cognitive factors in shaping the outcomes of L1 attrition, and specify the interactions between these factors. Developing and testing such a framework is of course challenging, but it is worth further investigation.

Although we have not been able to reveal a clear relationship between individual differences in cognitive abilities and variations in L1 attrition, this study still achieved its major objectives, and it has made at least two contributions to the field of L1 attrition research. Firstly, this study has added empirical evidence to L1 attrition research by looking into the attrition of two under-researched language structures in this field. Secondly, it has demonstrated that, although not always the case, individual difference in WM capacity could interact with linguistic factors, and affect the outcomes of L1 attrition. This study has also shown that the IH and alternative theoretical frameworks cannot sufficiently explain the observed L1 attrition data, and suggested that a multidimensional framework should be developed. Future research on this topic is welcome, and one day we will eventually be able to solve the mysteries of bilingualism.

References

Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, *10*(1), 79–99.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208.

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29.

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *1*(1).

Bayram, F., Prada, J., Cabo, D. P. y, & Rothman, J. (2018). Why Should Formal Linguistic Approaches to Heritage Language Acquisition Be Linked to Heritage Language Pedagogies? In P. Trifonas & T. Aravossitas (Eds.), *Handbook of Research and Practice in Heritage Language Education* (pp. 187–205). Cham: Springer.

Chamorro, G., Sorace, A., & Sturt, P. (2016a). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism: Language and Cognition*, *19*(3), 520–532.

Chamorro, G., Sturt, P., & Sorace, A. (2016b). Selectivity in L1 Attrition: Differential Object Marking in Spanish Near-Native Speakers of English. *Journal of Psycholinguistic Research*, *45*(3), 697–715.

Chen, J., & Shirai, Y. (2010). The development of aspectual marking in child Mandarin Chinese. *Applied Psycholinguistics*, *31*(01), 1–28.

Chen, Z., Jäger, L., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. In B. Stolterfoht & S. Featherston (Eds.), *Empirical Approaches to Linguistic Theory: Studies in Meaning and Structure* (pp. 43–62). Berlin: De Gruyter Mouton.

Chien, Y.-C., Wexler, K., & Chang, H.-W. (1993). Children's development of long-distance binding in Chinese. *Journal of East Asian Linguistics*, *2*(3), 229–259.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

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*(5), 769–786.

Cowan, N. (2005). *Working Memory Capacity* (1 edition). New York: Psychology Press.

Cunnings, I. (2017). Parsing and Working Memory in Bilingual Sentence Processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*(1-2), 188–219.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433.

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142.

Dillon, B., Chow, W.-Y., Wagers, M., Guo, T., Liu, F., & Phillips, C. (2014). The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology*, *5*. http://doi.org/10.3389/fpsyg.2014.01025

Dillon, B., Chow, W.-Y., & Xiang, M. (2016). The Relationship Between Anaphor Features and Antecedent Retrieval: Comparing Mandarin Ziji and Ta-Ziji. *Frontiers in Psychology*, *6*(January), 1–16. http://doi.org/10.3389/fpsyg.2015.01966

Dink, J. W., & Ferguson, B. (2015). eyetrackingR: An R Library for Eye-tracking Data Analysis.

Domínguez, L. (2013). *Understanding interfaces: Second language acquisition and native language attrition of Spanish subject realization and word order variation*. Amsterdam: John Benjamins.

Domínguez, L., & Hicks, G. (2016). Synchronic change in a multidialectal Spanish community: Evidence from null and postverbal subjects. In A. Cuza, L. Czerwionka, & D. Olson (Eds.), *Issues in Hispanic and Lusophone Linguistics* (Vol. 12, pp. 53–72). Amsterdam: John Benjamins.

Duff, P., & Li, D. (2002). The acquisition and use of perfective aspect in Mandarin. In R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense-aspect morphology* (Vol. 27, pp. 417–454). Amsterdam: John Benjamins.

Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *International Journal of Bilingualism*, *8*(3), 355–371.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in SpanishEnglish bilinguals. *Bilingualism: Language and Cognition*, *10*(1), 101–116.

Erbaugh, M. S. (1992). The acquisition of Mandarin. *The Crosslinguistic Study of Language Acquisition*, *3*, 373–455.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Ferreira, F., & Patson, N. D. (2007). The "Good Enough" Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1-2), 71–83.

Gao, L., Liu, Z., & Huang, Y. (2005). 'Ziji' shi shui? Yueshu yuanze de shiyan yanjiu [Who is ziji? An experimental research on Binding Principle]. *Linguistic Sciences*, *4*(2), 39–50.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *39*(1), 3–27.

GmbH, L. (2003). LimeSurvey: An Open Source survey tool. Hamburg, Germany: LimeSurvey GmbH.

Gürel, A. (2004). Selectivity in L2-induced L1 attrition: A psycholinguistic account. *Journal of Neurolinguistics*, *17*(1), 53–78.

Gürel, A. (2008). Review article: Research on first language attrition of morphosyntax in adult bilinguals. *Second Language Research*, *24*(3), 431–449.

Hanban. (2014). Chinese Tests_HSK.

Huang, C.-T. J., & Liu, C.-S. L. (2001). Logophoricity, attitudes, and ziji at the interface. In P. Cole, C.-T. J. Huang, & G. Hermon (Eds.), *Syntax and Semantics* (Vol. 33, pp. 141–195). San Diego, CA: Emerald Group Publishing Limited.

Huang, Y. (1994). *The syntax and pragmatics of anaphora*. Cambridge: Cambridge University Press.

Hui, S. (2012). *First language attrition of Chinese reflexives* (PhD thesis). The Chinese University of Hong Kong, Hongkong, China.

Iverson, M. (2012). Advanced language attrition of Spanish in contact with Brazilian Portuguese. *Theses and Dissertations*.

Iverson, M., & Miller, D. (2017). Language attrition and maintenance. *Linguistic Approaches to Bilingualism*, *7*(6), 704–708.

Jackendoff, R. (1996). *The Architecture of the Language Faculty*. Cambridge, MA: The MIT Press.

Jackendoff, R. (2003). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, *6*, 617–617. http://doi.org/10.3389/fpsyg.2015.00617

Jia, L., & Bayley, R. (2008). The (re) acquisition of perfective aspect marking by Chinese heritage language learners. In A. W. He & Y. Xiao (Eds.), *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 205–222). Honolulu, HI: University of Hawai'i Press.

Jin, L., & Hendriks, H. (2003). The development of aspect marking in L1 and L2 Chinese. *Working Papers on Applied Linguistics*, *9*, 69–100.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2007). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*(1), 193–224.

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

Kasparian, K., & Steinhauer, K. (2017). When the second language takes the lead: Neurocognitive processing changes in the first language of adult attriters. *Frontiers in Psychology*, *8*. http://doi.org/10.3389/fpsyg.2017.00389

Kasparian, K., Vespignani, F., & Steinhauer, K. (2017). First Language Attrition Induces Changes in Online Morphosyntactic Processing and Re-Analysis: An ERP Study of Number Agreement in Complex Italian Sentences. *Cognitive Science*, *41*(7), 1760–1803.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602.

Klein, W., Li, P., & Hendriks, H. (2000). Aspect and Assertion in Mandarin Chinese. *Natural Language & Linguistic Theory*, *18*(4), 723–770.

Ko, B. (2014, July). *Syntax, interfaces and processing in native language attrition* (PhD Thesis). University of Edinburgh, Edinburgh, UK.

Köpke, B., & Schmid, M. S. (2004). Language attrition: The next phase. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* (pp. 1–43). Amsterdam: John Benjamins.

Kupisch, T., & Rothman, J. (2016). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 1367006916654355.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models.

Lardiere, D. (2005). On Morphological Competence. In L. Dekydtspotter, R. A. Sprouse, & A. Liljestrand (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)* (pp. 178–192). Somerville, MA: Cascadilla Proceedings Project.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, *25*(2), 173–227.

Laws, J., & Yuan, B. (2010). Is the core-peripheral distinction for unaccusative verbs cross-linguistically consistent?: Empirical evidence from Mandarin. *Chinese Language & Discourse*, *1*(2), 220–263.

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

Li, C., & Thompson, S. (1989). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.

Li, D. C.-H., & Kaiser, E. (2009). Overcoming structural preference: Effects of context on the interpretation of the Chinese reflexive ziji. In S. L. Devi, A. Branco, & R. Mitkov (Eds.), *Proceedings of the 7th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC 2009)* (pp. 64–72). Berlin: Springer.

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

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883.

Liu, F.-h. (2007). Auxiliary selection in Chinese. In R. Aranovich (Ed.), *Split auxiliary systems* (pp. 181–205). Amsterdam: John Benjamins.

Liu, M. (2015). Tense and Aspect in Mandarin Chinese. In W. S. Wang & C. Sun (Eds.), *The Oxford Handbook of Chinese Linguistics* (pp. 274-289). Oxford: Oxford University Press.

Liu, Z. (2009). The cognitive process of Chinese reflexive processing. *Journal of Chinese Linguistics*, *37*(1), 1–27.

Lüdecke, D. (2018). *sjPlot: Data Visualization for Statistics in Social Science*.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54; discussion 55–74.

MacWhinney, B. (2004). A Unified Model of Language Acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49–67). New York: Oxford University Press.

MacWhinney, B. (2012). The logic of the unified model. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 211–227). New York: Routledge.

MacWhinney, B. (2018). Attrition and the Competition Model. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 1–15). New York: Oxford University Press.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). Amsterdam: John Benjamins.

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS repository: Advancing research practice and methodology. *Advancing Methodology and Practice: The IRIS Repository of Instruments for Research into Second Languages*, 1–21.

Miličević, M., & Kraš, T. (2017). Translation Between L2 Acquisition and L1 Attrition: Anaphora Resolution in Italian by EnglishItalian Trainee Translators. *Applied Linguistics*, *38*(1), 21–42.

Miller, D. (2017). *Combining behavioral and neurolinguistic methodologies to investigate Spanish scalar indefinites among mono- and bilinguals: An event-related potential study* (PhD Thesis). University of Reading, Reading, UK.

Ming, T., & Tao, H. (2008). Developing a Chinese heritage language corpus: Issues and a preliminary report. In *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 167–187). Honolulu, HI: University of Hawai'i Press.

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.

Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, *121*(4), 591–604.

Montrul, S. (2012). Bilingual background questionnaire for Spanish/English speakers.

Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.

Montrul, S., & Tanner, D. S. (2017). Individual differences and retrieval interference in L2 Processing. *Bilingualism: Language and Cognition*, *20*(4), 704–705.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799.

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2014). Visual and auditory digit-span performance in native and non-native speakers. *International Journal of Bilingualism*, *18*(6), 663–673.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Studies in Bilingualism* (Vol. 33, pp. 121–133). Amsterdam: John Benjamins.

Pascual y Cabo, D., & Rothman, J. (2012). The (Il) logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics*, 1–7.

Peck, J., Lin, J., & Sun, C. (2013). Aspectual Classification of Mandarin Chinese Verbs: A Perspective of Scale Structure. *Language and Linguistics*, *14*(4), 663–700.

Peirce, J. W. (2007). PsychoPyPsychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13.

Perpiñán, S. (2011). Optionality in bilingual native grammars. *Language, Interaction and Acquisition*, *2*(2), 312–341.

Polinsky, M. (2011). Reanalysis in adult heritage language. *Studies in Second Language Acquisition*, *33*(02), 305–328.

Polinsky, M. (2018). *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.

Puig-Mayenco, E., Rothman, J., Cunnings, I., Bayram, F., Miller, D., & Tubau, S. (2018). Language Dominance Affects Early Bilingual Performance and Processing Outcomes in Adulthood. *Frontiers in Psychology*, *9*. http://doi.org/10.3389/fpsyg.2018.01199

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology. Human Perception and Performance*, *9*(6), 912–922.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reinhart, T. (2006). *Interface Strategies: Optimal and Costly Computations*. Cambridge, MA: The MIT Press.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, *13*(2), 155–163.

Rothman, J., & Treffers-Daller, J. (2014). A Prolegomenon to the Construct of the Native Speaker: Heritage Speaker Bilinguals are Natives Too! *Applied Linguistics*, *35*(1), 93–98.

Schmid, M. S. (2011). *Language Attrition*. New York: Cambridge University Press.

Schmid, M. S., & Köpke, B. (2017a). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, *7*(6), 637–667.

Schmid, M. S., & Köpke, B. (2017b). When is a bilingual an attriter? *Linguistic Approaches to Bilingualism*, *7*(6), 763–770.

Schumacher, P. B., Bisang, W., & Sun, L. (2011). Perspective in the Processing of the Chinese Reflexive ziji: ERP Evidence. In I. Hendrickx, S. Lalitha Devi, A. Branco, & R. Mitkov (Eds.), *Anaphora Processing and Applications* (pp. 119–131). Berlin: Springer.

Seliger, H. (1996). Primary language attrition in the context of bilingualism. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 605–626). New York: Academic Press.

Seliger, H. W., & Vago, R. M. (1991). The study of first language attrition: An overview. In H. W. Seliger & R. M. Vago (Eds.), *First Language Attrition* (pp. 3–16). Cambridge: Cambridge University Press.

Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntaxPragmatics interface: Subjects and objects in EnglishItalian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, *7*(3), 183–205.

Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians. *Language Testing*, *15*(1), 86–108.

Shi, M. (2011). *Incomplete Knowledge of Aspect in Heritage Speakers of Mandarin Chinese in the Netherlands* (MA Dissertation). Radboud University, Nijmegen.

Shuai, L., Gong, T., & Wu, Y. (2013). Who Is Who? Interpretation of Multiple Occurrences of the Chinese Reflexive: Evidence from Real-Time Sentence Processing. *PLoS ONE*, *8*(9), 1–11.

Smith, C. S. (1997). *The Parameter of Aspect*. (G. Chierchia, P. Jacobson, & F. J. Pelletier, Eds.) (Vol. 43). Dordrecht: Springer.

Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33.

Sorace, A. (2012). Pinning down the concept of interface in bilingual development: A reply to peer commentaries. *Linguistic Approaches to Bilingualism*, *2*(2), 209–217.

Staum Casasanto, L., Hofmeister, P., & Sag, I. A. (2010). Understanding acceptability judgments: Additivity and working memory effects. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 224–229). Austin, TX: Cognitive Science Society.

Su, Y.-c. (2004). The Development of Long-Distance Binding for Chinese Ziji Revisited. *Chinese Journal of Psychology*, *46*(2/3), 249–249.

Truscott, J. (2017). Modularity, working memory, and second language acquisition: A research program. *Second Language Research*, *33*(3), 313–323.

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, *8*(3), 257–277.

Van Dyke, J. A., & Johns, C. L. (2012). Memory Interference as a Determinant of Language Comprehension. *Language and Linguistics Compass*, *6*(4), 193–211.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403.

Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.

Villegas, Á. (2014). *The role of L2 English immersion in the processing of L1 Spanish sentence complement/relative clause ambiguities* (PhD thesis). Pennsylvania State University, Pennsylvania, US.

von der Malsburg, T., & Vasishth, S. (2012). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1–34.

Waters, G. S., & Caplan, D. (1996). The Measurement of Verbal Working Memory Capacity and Its Relation to Reading Comprehension. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 51–79.

Waters, G. S., & Caplan, D. (2004). Verbal Working Memory and on-Line Syntactic Processing: Evidence from self-Paced Listening. *The Quarterly Journal of Experimental Psychology Section A*, *57*(1), 129–163.

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, *121*(4), 577–590.

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wiedenhof, J. (2015). *A Grammar of Mandarin*. Amsterdam: John Benjamins.

Wright, C. (2010). *Working memory in SLA*. Saarbrucken: VDM Publishing.

Xiao, R., & McEnery, T. (2004). *Aspect in Mandarin Chinese: A corpus-based study*. Amsterdam: John Benjamins.

Xue, P., Pollard, C., & Sag, I. A. (1994). A new perspective on Chinese ziji. In.

Yap, F. H., Chu, P. C. K., Yiu, E. S. M., Wong, S. F., Kwan, S. W. M., Matthews, S., … Shirai, Y. (2009). Aspectual asymmetries in the mental representation of events: Role of lexical and grammatical aspect. *Memory & Cognition*, *37*(5), 587–595.

Yuan, B. (1998). Interpretation of binding and orientation of the Chinese reflexive ziji by English and Japanese speakers. *Second Language Research*, *14*(4), 324–340.

Zhou, P., Crain, S., & Zhan, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition*, *133*(1), 262–276.

Appendix 1. Model summaries

Appx 1.1. Model summaries for Section 4.3

Summary of the models concerning the Group effect on judging the sentences containing aspect markers

	Estimate	**Std.Error**	***t***	***p***
ACC+le				
(Intercept)	0.70	0.02	29.58	<0.001
Group	-0.05	0.05	-0.96	0.34
ACC+zhe				
(Intercept)	0.11	0.20	0.56	0.60
Group	-0.18	0.34	-0.83	0.42
ACH+le				
(Intercept)	0.63	0.03	19.19	<0.001
Group	0.05	0.07	0.82	0.41
ACH+zhe				
(Intercept)	-1.68	0.04	-37.42	<0.001
Group	0.12	0.08	1.46	0.17
ACT+le				
(Intercept)	0.57	0.06	9.05	<0.001
Group	0.15	0.12	1.23	0.23
ACT+zhe				
(Intercept)	0.39	0.12	3.09	0.02
Group	-0.03	0.18	-0.14	0.89
ILS+le				
(Intercept)	0.06	0.29	0.20	0.85
Group	0.19	0.20	0.94	0.38
ILS+zhe				
(Intercept)	-1.35	0.13	-10.45	<0.001
Group	0.39	0.21	1.89	0.07

	Estimate	Std.Error	*t*	*p*
SEM+le				
(Intercept)	0.03	0.42	0.06	0.95
Group	0.21	0.17	1.25	0.23
SEM+zhe				
(Intercept)	0.02	0.29	0.08	0.94
Group	0.12	0.21	0.57	0.58
SLS+le				
(Intercept)	0.08	0.42	0.18	0.87
Group	0.13	0.12	1.13	0.29
SLS+zhe				
(Intercept)	0.28	0.14	2.02	0.09
Group	0.27	0.20	1.38	0.19

Summary of the models concerning the Length of L2 Exposure effect on judging the sentences containing aspect markers

	Estimate	Std.Error	*t*	*p*
ACC+zhe				
(Intercept)	0.26	0.19	1.35	0.23
L2Exp	-0.04	0.02	-1.83	0.11
ACH+le				
(Intercept)	0.65	0.05	13.07	<0.001
L2Exp	-0.01	0.01	-1.41	0.17
ACH+zhe				
(Intercept)	-1.61	0.02	-65.87	<0.001
L2Exp	-0.00	0.00	-0.78	0.44
ACT+le				
(Intercept)	0.54	0.12	4.44	<0.001
L2Exp	-0.04	0.02	-2.18	0.05

	Estimate	Std.Error	*t*	*p*
ACT+zhe				
(Intercept)	0.45	0.16	2.88	0.02
L2Exp	-0.03	0.02	-1.29	0.22
ILS+le				
(Intercept)	0.02	0.34	0.07	0.95
L2Exp	0.00	0.02	0.21	0.85
ILS+zhe				
(Intercept)	-1.43	0.08	-16.92	<0.001
L2Exp	-0.01	0.01	-0.70	0.49
SEM+le				
(Intercept)	-0.01	0.44	-0.03	0.98
L2Exp	-0.06	0.03	-1.87	0.08
SEM+zhe				
(Intercept)	0.03	0.33	0.10	0.93
L2Exp	-0.03	0.02	-1.55	0.15
SLS+le				
(Intercept)	0.07	0.43	0.17	0.88
L2Exp	0.00	0.01	0.32	0.76
SLS+zhe				
(Intercept)	0.20	0.22	0.95	0.38
L2Exp	-0.03	0.02	-1.30	0.22

Appx 1.2. Model summaries for Section 5.1

Summary of the models concerning the effect of Group on the probability of choosing a BOTH interpretation

	Estimate	**Std.Error**	***z***	***p***
Type 1				
(Intercept)	-3.68	2.62	-1.40	0.16
Group	-4.10	2.86	-1.43	0.15
Type 2				
(Intercept)	-2.27	1.11	-2.04	0.04
Group	-1.43	1.51	-0.95	0.34
Type 3				
(Intercept)	-7.98	3.53	-2.26	0.02
Group	0.19	3.47	0.05	0.96
Type 4				
(Intercept)	-1.56	1.09	-1.43	0.15
Group	-1.41	0.98	-1.44	0.15
Type 6				
(Intercept)	-90.60	24.44	-3.71	0.00
Group	-59.94	33.85	-1.77	0.08

Summary of the models concerning the effect of Group on the probability of choosing an LD interpretation

	Estimate	**Std.Error**	***z***	***p***
Type 1				
(Intercept)	1.42	0.93	1.52	0.13
Group	1.90	1.62	1.18	0.24
Type 2	**Estimate**	**Std.Error**	***z***	***p***
(Intercept)	1.32	0.97	1.35	0.18
Group	0.09	1.41	0.06	0.95

	Estimate	**Std.Error**	***z***	***p***
Type 3				
(Intercept)	-22.29	11.10	-2.01	0.04
Group	3.15	14.36	0.22	0.83
Type 4	**Estimate**	**Std.Error**	***z***	***p***
(Intercept)	-0.01	1.39	-0.01	1.00
Group	0.64	0.68	0.94	0.35

Summary of the models concerning the effect of Group on the probability of choosing a LOC interpretation

	Estimate	**Std.Error**	***z***	***p***
Type 3				
(Intercept)	0.98	0.74	1.32	0.19
Group	1.62	1.44	1.13	0.26
Type 4				
(Intercept)	-2.49	1.56	-1.59	0.11
Group	0.51	1.15	0.45	0.66
Type 6				
(Intercept)	54.67	7.21	7.58	<0.001
Group	26.78	6.87	3.90	<0.001

Summary of the model concerning the effect of Length of L2 Exposure on the probability of choosing a BOTH interpretation

	Estimate	Std.Error	*z*	*p*
Type 1				
(Intercept)	-0.46	1.25	-0.37	0.71
L2Exp	-0.09	0.18	-0.53	0.60
Type 2				
(Intercept)	-1.60	1.29	-1.24	0.21
L2Exp	0.02	0.15	0.13	0.90
Type 3				
(Intercept)	-19.63	11.22	-1.75	0.08
L2Exp	-0.52	0.48	-1.09	0.28
Type 4				
(Intercept)	-0.63	0.89	-0.71	0.48
L2Exp	0.00	0.06	-0.01	0.99
Type 6				
(Intercept)	-31.60	75.54	-0.42	0.68
L2Exp	-0.06	39.02	0.00	1.00

Summary of the models concerning the effect of Length of L2 Exposure on the probability of choosing an LD interpretation

	Estimate	Std.Error	*z*	*p*
Type 1				
(Intercept)	0.40	1.14	0.35	0.72
L2Exp	0.08	0.16	0.53	0.60
Type 2				
(Intercept)	1.60	1.29	1.24	0.21
L2Exp	-0.02	0.15	-0.13	0.90

	Estimate	Std.Error	*z*	*p*
Type 3				
(Intercept)	-18.46	9.20	-2.01	0.04
L2Exp	-0.69	0.69	-1.00	0.32
Type 4				
(Intercept)	-0.34	1.31	-0.26	0.79
L2Exp	-0.04	0.06	-0.60	0.55

Summary of the models concerning the effect of Length of L2 Exposure on the probability of choosing a LOC interpretation

	Estimate	Std.Error	*z*	*p*
Type 3				
(Intercept)	0.17	0.99	0.17	0.87
L2Exp	0.00	0.13	-0.03	0.98
Type 4				
(Intercept)	-2.06	1.04	-1.98	0.05*
L2Exp	0.06	0.07	0.84	0.40
Type 6				
(Intercept)	31.59	15.15	2.09	0.04*
L2Exp	0.05	8.72	0.01	1.00

Appendix 2. Test materials

Appx 2.1. Language background, reading and listening skills and working memory capacity

2.1.1. Questionnaire and abridged HSK-3 test

Please download these computerized test material from the IRIS depository (https://www.iris-database.org/iris/app/home/detail?id=york%3a933788&ref=search).

2.1.2. Digits-back recall task

Practice set 1:

1-3-7	6-2-5	9-2-3

Practice set 2:

3-2-1-5	7-2-4-6	5-1-6-7

Set 1:

9-5-1	7-2-3	1-7-4

Set 2:

3-2-4-7	8-4-6-1	3-6-8-1

Set 3:

4-2-5-3-1	2-3-7-5-9	8-5-4-2-1

Set 4:

6-8-5-7-4-1	1-6-8-4-6-2	9-8-7-6-4-3

Set 5:

7-5-4-8-9-6-2	1-5-8-6-9-3-6	5-9-7-3-1-6-4

Appx 2.2. Aspect marking

2.2.1. Acceptability judgement task

ILS+*le/zhe

1. 他胖了/着。

Ta	pang	le/zhe.
He	fat	PERF/DURA

2. 他聪明了/着。

Ta	congming	le/zhe.
He	smart	PERF/DURA

3. 他高了/着。

Ta	gao	le/zhe
He	tall	PERF/DURA

SLS+le/zhe

1. 他忙了/着。

Ta	mang	le/zhe.
He	busy	PERF/DURA

2. 他饿了/着。

Ta	e	le/zhe.
He	hungry	PERF/DURA

3. 他高兴了/着。

Ta	gaoxing	le/zhe.
He	happy	PERF/DURA

ACT+le/zhe

1. 他吃了/着饭。

Ta	chi	le/zhe	fan.
He	eat	PERF/DURA	meal

2. 他写了/着字。

Ta	xie	le/zhe	zi.
He	write	PERF/DURA	character

3. 他喝了/着水。

Ta	he	le/zhe	shui.
He	drink	PERF/DURA	water

SEM+le/zhe

1. 他咳了/着嗽。

Ta	ke	le/zhe	sou.
He	cough	PERF/DURA	cough

2. 他眨了/着眼。

Ta	zha	le/zhe	yan.
He	wink	PERF/DURA	eyes

3. 他敲了/着门。

Ta	qiao	le/zhe	men.
He	knock	PERF/DURA	door

ACC+le/zhe

1. 他吃了/着一碗饭。

Ta	chi	le/zhe	yiwan	fan.
He	eat	PERF/DURA	a bowl	rice

2. 他写了/着一本书。

Ta	xie	le/zhe	yiben	shu.
He	write	PERF/DURA	a	book

3. 他喝了/着一杯水。

Ta	he	le/zhe	yibei	shui.
He	drink	PERF/DURA	a cup	water

ACH+le/*zhe

1. 他到了/学校。

Ta	dao	le/zhe	xuexiao.
He	arrive	PERF/DURA	school

2. 他找到了/着钱。

Ta	zhaodao	le/zhe	qian.
He	find	PERF/DURA	money

3. 他看见了/着你。

Ta	kanjian	le/zhe	ni.
He	see	PERF/DURA	you

2.2.2. Cloze task

Please note that the unfilled blanks are filler blanks which do not require *le* or *zhe*.

1. 昨天下午，老师读（了）书（ ），但是没有读完（ ）。

Zuotian	xiawu,	laoshi	du	(le)	shu,	danshi	meiyou	du-wan ( ).
Yesterday	afternoon,teacher		read	(PERF)	book,but		not	read-done.

2. 爸爸回家的时候（ ），菜还热（着）呢。

Baba	hui	jia	de-shihou( ),		cai	hai	re	(zhe)	ne.
Father	return	home	when	( ),	cuisine	still	hot	(DURA)	Intensifier

3. 昨天，王先生打（了）篮球，还踢（了）足球。

Zuotian,	wang-xiansheng	da	(le)	lanqiu,	hai	ti	(le)	zuqiu.
Yesterday,Mr Wang		play	(PERF)	basketball,and play			(PERF)	football.

4. 王先生学（了）一年汉语。

Wang-xiansheng	xue	(le)	yinian	hanyu.
Mr Wang	learn	(PERF)	one-year	Chinese

5. 医生喜欢走（着）去（ ）上班。

Yisheng	xihuan	zou	(zhe)	qu	( )	shangban.
Doctor	like	walk	(DURA)	go-to	( )	work

6. 那些东西都在书上写（着）呢。

Naxie	dongxi	dou	zai	shu	shang	xie	(zhe)	ne.
Those	things	all	at	book	on	write	(DURA)	Intensifier

7. 老师已经吃（了）饭，但是还没回家（ ）。

Laoshi	yijing	chi	(le)	fan,	danshi	hai	mei	huijia	( ).
Teacher	already	eat	(PERF)	meal,	but	yet	not	go-home	( ).

8. 昨天，李先生坐（着）飞机去中国（了）。

Zuotian,	li-xiansheng	zuo	(zhe)	feiji	qu	zhongguo	(le).
Yesterday,	Mr Li	sit	(DURA)	plane	go-to	China	(PERF)

9. 后面坐（着）看书的是我的爸爸（ ）。

Houmian	zuo	(zhe)	kan	shu	de	shi	wo-de	baba	( ).
Behind	sit	(DURA)	read	book	Relativ	be	my	father	( )

10. 王先生一直看（着）电脑，没有说话（ ）。

Wang-xiansheng	yizhi	kan (zhe)	diannao,	meiyou	shuohua	( ).
Mr Wang	continuously	watch(DURA)	computer,	not	speak	( )

11. 北京现在下（着）雨。

Beijing	xianzai	xia	(zhe)	yu.
Beijing	now	down	(DURA)	rain.

12. 学生们（ ）不在学校，都回家（了）。

Xuesheng-men	( )	bu	zai	xuexiao,	dou	huijia	(le).
Students	( )	not	at	school,	all	go-home	(PERF).

13. 李先生坐（着）出租车，要去（ ）公司。

Li-xiansheng	zuo	(zhe)	chuzuche,	yao	qu	( )	gongsi.
Mr Li	sit	(DURA)	taxi,	want	go-to	( )	company

14. 老师（ ）后面坐（着）看你呢。

Laoshi	( )	houmian	zuo	(zhe)	kan	ni	ne.
Teacher	( )	behind	sit	(DURA)	watch	you	Intensifier

15. 他今天写（了）很多东西。

Ta	jintian	xie	(le)	henduo	dongxi.
He	today	write	(PERF)	many	thing

2.2.3. Sentence-picture matching task

ACT+*le* (semantically mismatched condition):

1. 他跑了步。

Ta	pao	le	bu.
S/he	run	PERF	pace

‘He has run.’

2. 他游了泳。

Ta	you	le	yong.
S/he	swim	PERF	swim

‘He has swum.’

3.　他喝了啤酒。

Ta	he	le	pijiu.
S/he	drink	PERF	beer

'He has drunk some beer.'


4.　他看了书。

Ta	kan	le	shu.
S/he	read	PERF	book

'He has read.'


5. 他写了字。

Ta	xie	le	zi.
S/he	write	PERF	letter

‘He has written something.’

6. 她洗了衣服。

Ta	xi	le	yifu.
S/he	wash	PERF	clothes

‘She has washed the clothes.’

7. 他走了路。

Ta zou le lu.
S/he walk PERF road
‘He has walked.’

8. 他踢了足球。

Ta ti le zuqiu.
S/he kick PERF football
‘He has played football.’

9.　他骑了自行车。

Ta	qi	le	zixingche.
S/he	ride	PERF	bicycle

'He has ridden a bicycle.'

10. 他打了篮球。

Ta	da	le	lanqiu.
S/he	play	PERF	basketball

'He has played basketball.'

ACT+*zhe* (semantically matched condition):

1. 他跑着步。

Ta	pao	zhe	bu.
S/he	run	DURA	pace

‘He is running.’


2. 他游着泳。

Ta	you	zhe	yong.
S/he	swim	DURA	swim

‘He is swimming.’


3. 他喝着啤酒。

Ta	he	zhe	pijiu.
S/he	drink	DURA	beer

‘He is drinking some beer.’


4. 他看着书。

Ta	kan	zhe	shu
S/he	read	DURA	book

‘He is reading.’


5. 他写着字。

Ta	xie	zhe	zi.
S/he	write	DURA	letter

'He is writing something.'

6. 她洗着衣服。

Ta	xi	zhe	yifu.
S/he	wash	DURA	clothes

'She is washing the clothes.'

7. 他走着路。

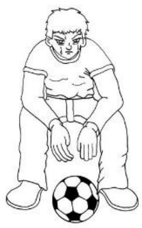
Ta	zou	zhe	lu.
S/he	walk	DURA	road

‘He is walking.’

8. 他踢着足球。

Ta	ti	zhe	zuqiu.
S/he	kick	DURA	football

‘He is playing football.’



9. 他骑着自行车。

Ta	qi	zhe	zixingche.
S/he	ride	DURA	bicycle

‘He is riding a bicycle.’

10. 他打着篮球。

Ta	da	zhe	lanqiu.
S/he	beat	DURA	basketball

‘He is playing basketball.’

Appx 2.3. Long-distance binding property of *ziji*

2.3.1. Pencil-and-paper interpretation task

Please note that the exemplar sentences presented in Section 3.9 were also used as test materials, and those sentences were not present here.

Type 1 *ziji* in embedded finite clause (Neutral)

1. 高林知道李东非常相信自己。

GaoLin zhidao LiDong feichang xiangxin ziji.

GaoLin know LiDong very trust self

‘Gao Lin knows that Li Dong trusts self very much.’

2. 王萍相信李东非常了解自己。

WangPing xiangxin LiDong feichang liaojie ziji.

WangPing believe LiDong very know self

'Wang Ping believes that Li Dong knows her well.'

Type 2 *ziji* in embedded finite clause (LD favoured)

1. 王明不高兴地说李东经常不相信自己。

WangMing bu-gaoxing de shuo LiDong jingchang bu xiangxin ziji.

WangMing unhappily Adv-P say LiDong often not trust self

‘Wang Ming said unhappily that Li Dong often does not trust self.’

2. 李东不高兴地说高林非常不了解自己。

Li Dong bu-gaoxing de shuo GaoLin feichang bu liaojie ziji.

Li Dong unhappily Adv-P say GaoLin feichang not understand self

'Li Dong said unhappily that Gao Lin does not know self well.'

Type 3 *ziji* in embedded finite clause (LOC favoured)

1. 李东记得王老师第一次来上课的时候没有介绍自己。

LiDong jide Wang Laoshi diyi ci lai shang ke de shihou meiyou jieshao ziji.
LiDong rememberWang Teacher first time come teach class when not introduce self
‘Li Dong remembers that Teacher Wang didn’t introduce self when he came to teach the class for the first time.’

2. 高红记得王老师第一次来上课的时候没有吹捧自己。

Gao Hong jide Wang Laoshi diyi ci lai shang ke de shihou meiyou chuipeng ziji.
Gao Hong remember Wang Teacher first time come teach class when not boast self
'Gao Hong remembers that Teacher Wang didn't boast self when he came to teach the class for the first time.'

Type 4 *ziji* in infinitive clause (neutral)

1. 王平让李东不要批评自己。

Wang Ping rang Li Dong buyao piping ziji.
Wang Ping ask Li Dong not criticize self
‘Wang Ping asked Li Dong not to criticize self.’

2. 高红让妈妈不要伤害自己。

Gao Hong rang mama buyao shanghai ziji.
Gao Hong ask mother not harm self
'Gao Hong asked her mother not to harm self'

Type 5 *ziji* in infinitive clause (LD favoured)

1. 高红不愿意跟别人讲话，所以她不愿意她妈妈向别人介绍自己。

Gao Hong bu-yuanyi gen bieren jianghua, suoyi ta bu yuanyi ta mama xiang bieren jieshao ziji.
Gao Hong notlike with others speak therefore she not like her mother to others introduce self
'Gao Hong does not like to talk to other people. Therefore, she does not like her mother to introduce self to other people.'

2. 张萍不愿意跟别人讲话，所以她不愿意她妈妈向别人谈起自己。

Zhang Ping bu-yuanyi gen bieren jianghua, suoyi ta bu yuanyi ta mama xiang bieren tanqi ziji.
Gao Hong notlike with others speak therefore she not like her mother to others talk self
'Gao Hong does not like to talk to other people. Therefore, she does not like her mother to talk about self with other people.'

Type 6 *ziji* in infinitive clause (LOC favoured)

1. 李教授让张萍严格要求自己，不要总是去踢足球。

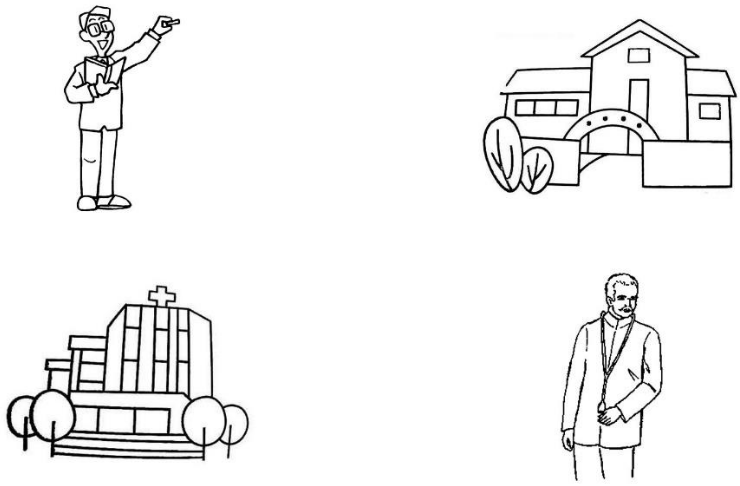
Li Jiaoshou rang Zhang Ping yange yaoqiu ziji, buyao zongshi qu ti zuqiu.
Li Professor ask Zhang Ping strict require self not always go play football
'Professor Li asked Zhang Ping to set strict demands on self, and not to play football all the time.'

2. 王老师让李东严格约束自己，不要总是去打篮球。

Wang Laoshi rang Li Dong yange yueshu ziji, buyao zongshi qu da lanqiu.
Wang Teacher ask Li Dong strict regulate self not always go play basketball
'Teacher Wang asked Li Dong to set strict demands on self, and not to play basketball all the time.'

2.3.2. Visual world eye-tracking task

Set 1:



LD

医生说学校锻炼了自己，这件事是真的。

Yisheng	shuo	xuexiao	duanlian	le	ziji, […]
Doctor	say	school	exercise	PERF	self

‘The doctor said that the school had improved him, …’

LOCAL

学校说医生锻炼了自己，这件事是真的。

Xuexiao	shuo	yisheng	duanlian	le	ziji, […]
School	say	doctor	exercise	PERF	self

‘The school said that the doctor had improved himself, …’

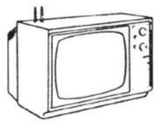
AMB

医生说老师锻炼了自己，这件事是真的。

Yisheng	shuo	laoshi	duanlian	le	ziji, […]
Doctor	say	teacher	exercise	PERF	self

‘The doctor said that the teacher had improved him/himself, …’

Set 2:


LD

老师说电视介绍了自己，这件事是真的。

Laoshi	shuo	dianshi	jieshao	le	ziji, […]
Teacher	say	TV	introduce	PERF	self

‘The teacher said that the TV programme had introduced him, …’

LOCAL

电视说老师介绍了自己，这件事是真的。

Dianshi	shuo	laoshi	jieshao	le	ziji, […]
TV	say	teacher	introduce	PERF	self

‘The TV programme said that the teacher had introduced himself, …’

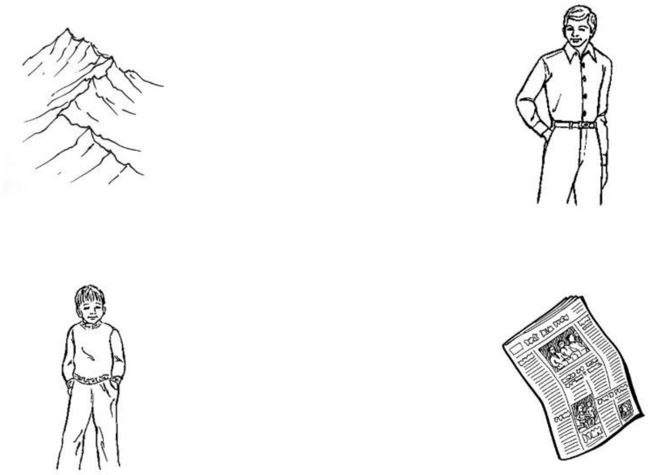
AMB

老师说医生介绍了自己，这件事是真的。

Laoshi	shuo	yisheng	jieshao	le	ziji, […]
Teacher	say	doctor	introduce	PERF	self

‘The teacher said that the doctor had introduced him/himself, …’

Set 3:



LD

爸爸说报纸帮助了自己，这件事是真的。
Baba shuo baozhi bangzhu le ziji, […]
Father say newspaper help PERF self
‘The father said that the newspaper had helped him, …’

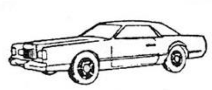
LOCAL

报纸说爸爸帮助了自己，这件事是真的。
Baozhi shuo baba bangzhu le ziji, […]
Newspaper say father help PERF self
‘The newspapaer said that the father had helped himself, …’

AMB

爸爸说男孩帮助了自己，这件事是真的。
Baba shuo nanhai bangzhu le ziji, […]
Father say boy help PERF self
‘The father said that the boy had helped him/himself, …’

Set 4:



LD

妈妈说医院检查了自己，这件事是真的。

Mama	shuo	yiyuan	jiancha	le	ziji, […]
Mother	say	hospital	examine	PERF	self

‘The mother said that the hospital had examined her, …’

LOCAL

医院说妈妈检查了自己，这件事是真的。

Yiyuan	shuo	mama	jiancha	le	ziji, […]
Hospital	say	mother	examine	PERF	self

‘The hospital said that the mother had examined herself, …’

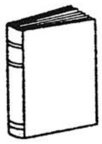
AMB

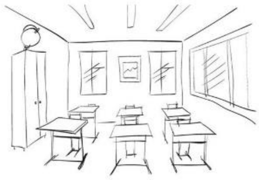
妈妈说女孩检查了自己，这件事是真的。

Mama	shuo	nvhai	jiancha	le	ziji, […]
Mother	say	girl	examine	PERF	self

‘The mother said that the girl had examined her/herself, …’

Set 5:





LD

女孩说那本书影响了自己，这件事是真的。

Nvhai	shuo	na-ben	shu	yingxiang	le	ziji, […]
Girl	say	that	book	impact	PERF	self

‘The girl said that the book had affected her, …’

LOCAL

那本书说女孩影响了自己，这件事是真的。

Na-ben	shu	shuo	nvhai	yingxiang	le	ziji, […]
That-CL	book	say	girl	impact	PERF	self

‘The book said that the girl had affected herself, …’

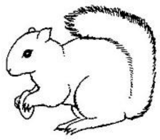
AMB

女孩说妈妈影响了自己，这件事是真的。

Nvhai	shuo	mama	yingxiang	le	ziji, […]
Girl	say	mother	impact	PERF	self

‘The girl said that the mother had affected her/herself, …’

Set 6:



LD

妈妈说出租车在看自己，这件事是真的。
Mama shuo chuzuche zai kan ziji, […]
Mother say taxi PROG see self
'The mother said that the taxi was looking at her, …'

LOCAL

出租车说妈妈在看自己，这件事是真的。
Chuzucheshuo mama zai kan ziji, […]
Taxi say mother PROG see self
'The taxi said that the mother was looking at herself, …'

AMB

妈妈说女孩在看自己，这件事是真的。
Mama shuo nvhai zai kan ziji, […]
Mother say girl PROG see self
'The mother said that the girl was looking at her/herself, …'

Set 7:

LD

爸爸说银行相信自己，这件事是真的。

Baba	shuo	yinhang	xiangxin	ziji, […]
Father	say	bank	trust	self

‘The father said that the bank trusted him, …’

LOCAL

银行说爸爸相信自己，这件事是真的。

Yinhang	shuo	baba	xiangxin	ziji, […]
Bank	say	father	trust	self

‘The bank said that the father trusted himself, …’

AMB

爸爸说男孩相信自己，这件事是真的。

Baba	shuo	nanhai	xiangxin	ziji, […]
Father	say	boy	trust	self

‘The father said the boy trusted him/himself, …’

Set 8:

LD

男孩说商店喜欢自己，这件事是真的。

Nanhai	shuo	shangdian	xihuan	ziji, […]
Boy	say	shop	like	self

‘The boy said that the shop liked him, …’

LOCAL

商店说男孩喜欢自己，这件事是真的。

Shangdian	shuo	nanhai	xihuan	ziji, […]
Shop	say	boy	like	self

‘The shop said that the boy liked himself, …’

AMB

男孩说爸爸喜欢自己，这件事是真的。

Nanhai	shuo	baba	xihuan	ziji, […]
Boy	say	father	like	self

‘The boy said that the father liked him/himself, …’

Set 9:

LD

老师说学校了解自己，这件事是真的。

Laoshi	shuo	xuexiao	liaojie	ziji, […]
Teacher	say	school	understand	self

‘The teacher said that the school understood him well, …’

LOCAL

学校说老师了解自己，这件事是真的。

Xuexiao	shuo	laoshi	liaojie	ziji, […]
School	say	teacher	understand	self

‘The teacher said that the teacher understood himself well, …’

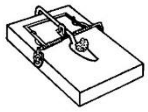
AMB

老师说医生了解自己，这件事是真的。

Laoshi	shuo	yisheng	liaojie	ziji, […]
Teacher	say	doctor	understand	self

‘The teacher said that the doctor understood him/himself well, …’

Set 10:



LD

医生说医院提高了自己，这件事是真的。

Yisheng	shuo	yiyuan	tigao	le	ziji, […]
Doctor	say	hospital	improve	PERF	self

‘The doctor said that the hospital had improved him, …’

LOCAL

医院说医生提高了自己，这件事是真的。

Yiyuan	shuo	yisheng	tigao	le	ziji, […]
Hospital	say	doctor	improve	PERF	self

‘The hospital said that the doctor had improved himself, …’

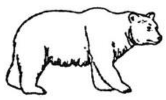
AMB

医生说老师提高了自己，这件事是真的。

Yisheng	shuo	laoshi	tigao	le	ziji, […]
Doctor	say	teacher	improve	PERF	self

‘The doctor said that the teacher had improved him/himself, …’

Set 11:



LD

女孩说出租车看清了自己，这件事是真的。

Nvhai	shuo	chuzuche	kanqing	le	ziji, […]
Girl	say	taxi	see-clear	PERF	self

‘The girl said that the taxi had finally seen her clearly, …’

LOCAL

出租车说女孩看清了自己，这件事是真的。

Chuzucheshuo	nvhai	kanqing	le	ziji, […]
Taxi say	girl	see-clear	PERF	self

‘The taxi said that the girl had finally seen herself clearly, …’

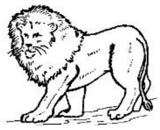
AMB

女孩说妈妈看清了自己，这件事是真的。

Nvhai	shuo	mama	kanqing	le	ziji, […]
Girl	say	mother	see-clear	PERF	self

‘The girl said that the mother had finally seen her/herself clearly, …’

Set 12:



LD

男孩说商店在担心自己，这件事是真的。

Nanhai	shuo	shangdian	zai	danxin	ziji, […]
Boy	say	shop	PROG	worry	self

‘The boy said that the shop was worried about him, …’

LOCAL

商店说男孩在担心自己，这件事是真的。

Shangdian	shuo	nanhai	zai	danxin	ziji, […]
Shop	say	boy	PROG	worry	self

‘The shop said that the boy was worried about himself, …’

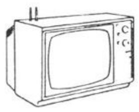
AMB

男孩说爸爸在担心自己，这件事是真的。

Nanhai	shuo	baba	zai	danxin	ziji, […]
Boy	say	father	PROG	worry	self

‘The boy said that the father was worried about him/himself, …’

Set 13:



LD

爸爸说电视讨厌自己，这件事是真的。

Baba	shuo	dianshi	taoyan	ziji, […]
Father	say	TV	dislike	self

'The father said that the TV programme did not like him, …'

LOCAL

电视说爸爸讨厌自己，这件事是真的。

Dianshi	shuo	baba	taoyan	ziji, […]
TV	say	father	dislike	self

'The TV programme said that the father did not like himself, …'

AMB

爸爸说男孩讨厌自己，这件事是真的。

Baba	shuo	nanhai	taoyan	ziji, […]
Father	say	boy	dislike	self

'The father said that the boy did not like him/himself, …'

Set 14:

LD

妈妈说银行累着了自己，这件事是真的。

Mama	shuo	yinhang	lei-zhao	le	ziji, […]
Mother	say	bank	make-tired	PERF	self

‘The mother said that the bank had made her very tired, …’

LOCAL

银行说妈妈累着了自己，这件事是真的。

Yinhang	shuo	mama	lei-zhao	le	ziji, […]
Bank	say	mother	make-tired	PERF	self

‘The bank said that the mother had made herself very tired, …’

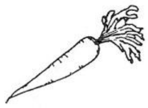
AMB

妈妈说女孩累着了自己，这件事是真的。

Mama	shuo	nvhai	lei-zhao	le	ziji, […]
Mother	say	girl	make-tired	PERF	self

‘The mother said that the girl had made her/herself very tired, …’

Set 15:



LD

老师说报纸想到了自己，这件事是真的。

Laoshi	shuo	baozhi	xiang-dao	le	ziji, […]
Teacher	say	newspaper	think-arrive	PERF	self

'The teacher said that the newspaper had thought about him, …'

LOCAL

报纸说老师想到了自己，这件事是真的。

Baozhi	shuo	laoshi	xiang-dao	le	ziji, […]
Newspaper	say	teacher	think-arrive	PERF	self

'The newspaper said that the teacher had thought about himself, …'

AMB

老师说医生想到了自己，这件事是真的。

Laoshi	shuo	yisheng	xiang-dao	le	ziji, […]
Teacher	say	doctor	think-arrive	PERF	self

'The teacher said that the doctor had thought about him/himself, …'