THE ACQUISITION OF MANDARIN BY HERITAGE SPEAKERS AND SECOND LANGUAGE LEARNERS

BY

CHUNG-YU CHEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
with a concentration in Second Language Acquisition and Teacher Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Tania Ionin, Chair and Director of Research
Professor Silvina Montrul
Professor Chilin Shih
Professor James Hye-Suk Yoon

Abstract

This dissertation investigates whether English-dominant heritage speakers (HSs) of Mandarin have selective advantages over proficiency-matched adult second language learners (L2ers) in several Mandarin linguistic phenomena, given the early age of acquisition (AoA) by HSs. Previous studies have found that HSs have an advantage over L2ers in phonology and core aspects of syntax, which develop before age three, but not in the domains of the lexicon, semantics, inflectional morphology, and syntax-discourse interface (see Montrul, 2012, 2016, for reviews). However, few studies have directly compared proficiency-matched HSs and L2ers across multiple linguistic domains, and none have done so for Mandarin. Thus, the goal of this study is to examine whether, when both Mandarin proficiency and the dominant language (English) are held constant, early AoA confers an advantage to HSs over L2ers of Mandarin, and whether this depends on the linguistic domain. Two broad research questions were asked: (1) Can HSs and L2ers of Mandarin whose dominant language is English fully acquire the properties of Mandarin that are different from or absent in English? (2) Do HSs have selective advantages over proficiency-matched L2ers, and does this vary by linguistic domain?

To answer these questions, four Mandarin phenomena (tone 3 sandhi, aspect marking, relative clauses, and long-distance reflexives) were chosen because they are either absent or differently encoded in English, have different AoAs in monolingual Mandarin-speaking children, and are in different linguistic domains (phonology, morpho-semantics, syntax, and syntax-semantics interface). Three offline tasks were used for testing: a Tone Identification Task, an Acceptability Judgement Task, and a picture-based Truth Value Judgement Task.

It is hypothesized that HSs will have acquired tone 3 sandhi, aspect marking, and relative clauses, as they are acquired by age five in monolingual children, but not long-distance reflexives, which are acquired after age eight when child HSs no longer receive extensive exposure to Mandarin. L2ers are hypothesized to have acquired Mandarin head-final relative clauses given that word order is relatively easy for L2ers, despite different headedness in English. However, L2ers may have difficulty with tone 3 sandhi because phonology is known to be difficult for them (e.g., Granena & Long, 2013), despite ample exposure to tones and tone 3 sandhi. Aspect marking may also be difficult for L2ers given that morphology presents a particularly challenging area (a 'bottleneck') for adult L2ers (Slabakova, 2008, 2014) and because there are cross-linguistic differences between English and Mandarin on aspect marking. Additionally, L2esr might find aspect marking difficult because predicates are not always marked with aspect markers, and there are restrictions on whether aspect markers can be combined with certain lexical predicates. Long-distance reflexives are expected to be difficult for L2ers due to a number of reasons, including English transfer, interface properties, presumably low frequency, and processing considerations.

The results show that HSs were more native-like than L2ers in some domains, giving HSs a slight advantage. For tone 3 sandhi, HSs were slightly more native-like than L2ers in choosing more target-like T2T3 sequences for the T3T3 conditions. However, the advantage in tone 3 sandhi was not as robust as expected, likely due to undesired task effects that led even some Mandarin native speakers to not perform at ceiling in the Tone Identification Task. For aspect marking, HSs were more native-like than L2ers in that the latter seemed to be more subject to dominant language transfer from English progressive *-ing*, thus incorrectly allowing Mandarin progressive *zai* with achievement predicates. While relative clauses were expected to be easy for L2ers, HSs were slightly more native-like than L2ers in some but not all conditions (as in interpreting relative

clauses and judging their headedness). Unlike Mandarin native speakers who accepted both long-distance and local readings of simplex reflexives *ziji* (though not at ceiling), neither HSs nor L2ers had acquired long-distance reading of *ziji.*

Taken together, the acquisition of these different phenomena is influenced by multiple considerations, including transfer from English and domain vulnerability, with HSs having an advantage in those language phenomena that are acquired early in monolingual children. Frequency and processing considerations are briefly discussed.

Acknowledgments

I would like to thank my committee chair and members for their guidance and support. First and foremost, I would like to thank my adviser, Dr. Tania Ionin, for being so patient, caring, and supportive at every stage of my graduate studies at the University of Illinois at Urbana-Champaign (UIUC). I am thankful for her insightful and detailed comments on the design of the experimental tasks and on my dissertation manuscript. I am grateful for Drs. Silvina Montrul, Chilin Shih, and James Hye Suk Yoon for being my committee members and providing much valuable feedback. Dr. Montrul has helped enormously with my research on heritage speakers and second language learners in general. Dr. Yoon helped me to clarify the theoretical division between syntax and morphology, and the theoretical approaches to long-distance reflexives. I am especially grateful to Dr. Shih for helping me design the tone task and clarifying some theoretical concepts of third tone sandhi. I would also like to thank her for providing me with some materials that became the Chinese proficiency test used in the dissertation. Additionally, I want to thank Dr. Marissa Barlaz for statistical consulting and Amy Gorgone for careful proofreading. All remaining errors are mine.

In addition, many special thanks to the undergraduate research assistants (especially Joy Jen, Watson Chen, Chenyu Zhao, Shi-Yun Chen, and Yizhen Ma) who helped create the stimuli, build the experimental surveys, and/or code the data and to my fellow Chinese linguists (especially Yan Sun, Shuju Shi, Yihan Zhou, Yuyun Lei, Kailu Guan, Yinglun Sun, and Mien-Jen Wu) whose linguistic intuitions I probed and/or whose expertise I often consulted. My gratitude also goes to the participants who spent their time doing the experimental tasks and providing post-task feedback. I also thank many professors, teaching assistants, and Chinese instructors, including but not limited to Drs. Dong-Bo Hsu, Jinsong Zhang, Ji Young Kim, Hongyin Tao, Liancheng Chief,

Yu-An Lu, Karen Chung, Li-May Sung, and Wei H. Kao, for their help with advertising the study, and/or even allowing me to use their space to test the participants.

Lastly, I extend my warmest thanks to my family in Taiwan, my extended family in the United States, and my friends in Taiwan and the United States for their continued support.

A shorter and preliminary version of Chapters 6 and 7 was presented in 2019 at the Boston University Conference on Language Development (BUCLD) and published as a proceeding in 2020 (C. Chen, 2020a). This research was made possible by the NSF Doctoral Dissertation Research Grant (BCS-1844447) (PI: Tania Ionin, co-PI: Chung-yu Chen), UIUC Second Language Acquisition and Teacher Education (SLATE) Doctoral Research Award, UIUC Beckman Institute Cognitive Science / Artificial Intelligence Award, and the UIUC Graduate College Dissertation Travel Grant.

Table of Contents

Chapter 1. **Introduction**

1.1. Goals / research objectives

Chinese is the most-spoken language in the world, and the third most-spoken language in the United States after English and Spanish. As China gains visibility on the world stage, the number of adult (post-puberty) second language learners (L2ers) of Mandarin is growing, as is the number of Mandarin heritage speakers (HSs), individuals who were exposed to Mandarin at home from their parents since birth but as adults are dominant in English. However, while linguistic research on L2-Chinese has been steadily increasing (for reviews, see Mai, 2016; Y. Zhao, 2011), research on heritage Chinese is still scant.

The goal of this dissertation is to examine whether HSs have selective advantages over proficiency-matched L2ers by investigating how HSs and L2ers of Mandarin acquire **tone sandhi, aspect marking, relative clauses, and long-distance reflexives**. These four phenomena were chosen because (i) they represent different subdomains in linguistics (**phonology, morpho-semantics, syntax, and syntax-semantics interface**), and (ii) they differ with regard to the age of acquisition (AoA) in monolingual children. While there are many studies on HSs (see Montrul, 2016; Polinsky, 2018b), most of them have tested HSs in only one linguistic domain. Only a few studies have tested HSs **across domains** (Cantonese: Nagy, 2015; Kan, 2018; French: Kupisch et al., 2014; Swedish: Håkansson, 1995); even fewer have compared the same group of HSs and L2ers **across domains** (Spanish: Au, Knightly, Jun, & Oh, 2002; Au, Oh, Knightly, Jun, & Romo, 2008; Knightly, Jun, Oh, & Au, 2003; Korean: Lee, Moon, & Long, 2009, Lee-Ellis, 2012), and there is only one study on Mandarin within morphosyntax (Mai & Deng, 2019). To fill this gap, this dissertation examines, when both Mandarin proficiency and the dominant language (English)

are held constant, whether early AoA confers an advantage to HSs and whether this depends on the linguistic domain.

While this study is on Mandarin only, the outcomes will provide broader insights into language acquisition theory in general, as well as have potential pedagogical implications, given the rapidly growing HS populations in many English-speaking college classrooms (Polinsky & Kagan, 2007; for heritage Chinese in the classroom, see D. Li & Duff, 2008, 2018; Y. Xiao, 2014).

1.2. Theoretical approaches to HS/L2 comparisons

While HSs have long been studied in educational or sociolinguistic research, only in the last decade have HSs been studied by experimental linguists, with particular focus from language acquisition and psycholinguistics (e.g., Benmamoun, Montrul, & Polinsky, 2013; Montrul, 2010b; Polinsky, 2011). More recently, studies have explicitly linked linguistic theories and heritage language acquisition more closely (e.g., Polinsky, 2018b; Polinsky & Scontras, 2020; Scontras, Fuchs, & Polinsky, 2015).

Unlike typically-developing monolingual children, who ultimately reach native-like competence in their first language (L1) and/or native language by adulthood, adult HSs often do not reach native-like competence and age-appropriate language skills. While HSs are native speakers (NSs) of their heritage language (e.g., Montrul, 2013; Rothman & Treffers-Daller, 2014), their ultimate attainment is often not comparable to that of NSs who grew up in predominantly monolingual environments "due to insufficient input leading to incomplete acquisition and attrition rather than to qualitative changes in the environment" (e.g., Montrul, 2016, p. 248).[1, 2]

[1] Montrul (2016, Chapter 7) points out that HSs in the United States seem to have a lower chance of being native-like than HSs in many European countries or Canada, where multilingualism is more valued.
[2] While the term "incomplete" acquisition (e.g., Montrul, 2008; Polinsky, 2006) has generated much debate (e.g., Pascual y Cabo & Rothman, 2012; see also Domínguez, Hicks, & Slabakova, 2019, and commentaries), the idea is

HSs and L2ers also differ in that HSs heard their L1s since birth at home, while adult L2ers typically learned the L2 in a classroom setting with both written and aural exposure. Despite many similarities (including transfer, simplification, and fossilization; see Montrul, 2016, Chapter 8), HSs often outperform adult L2ers, who acquire the target language after puberty when their L1s are already fully in place.

Comparing HSs and L2ers provides researchers with a good test case to re-examine critical factors contributing to language acquisition, mainly the effects of AoA, input, and dominant language transfer (Montrul, 2008, 2016). Indeed, these factors are not mutually exclusive and the relative contribution of these factors depends on the linguistic domain. Below I will consider various explanations and accounts of how and why HSs and L2ers compare.

1.2.1. HS/L2 comparisons: Age effects

The well-known Critical Period Hypothesis (see Mayberry & Kluender, 2018, for an overview) has argued that AoA contributes to the different outcomes of L1 and L2 acquisition.[3] The Critical Period Hypothesis in L2 acquisition predicts that pre-puberty learners can achieve native-like proficiency in an L2 while post-puberty learners cannot, due to decreasing brain plasticity from biological maturation. Though the existence of a critical period in L2 acquisition is debatable (for reviews, see Long, 2005; DeKeyser & Larson-Hall, 2005; Mayberry & Kluender, 2018), researchers agree that age effects exist and affect phonetics and phonology more than morphosyntax (e.g., Granena & Long, 2013; also see the discussion of vulnerable domains below).

---

that HSs (typically) do not achieve "complete" mastery of all aspects of their heritage language. See the distinction among emergence, acquisition, and mastery discussed in Montrul (2016, Chapter 4). Many HSs show acquisition without mastery of some aspects in their grammar.

[3] For the original formulation, see Lenneberg (1967). In L1 acquisition, the Critical Period Hypothesis is better supported, e.g., by studies of delayed L1 acquisition among the deaf population (see Mayberry, 2010; Mayberry & Kluender, 2018).

If so, HSs should acquire the language like monolingually-raised NSs and outperform adult L2ers because they acquired their home language early in life, even though the input is reduced (because another societal language co-exists, e.g., English in the United States). To be specific, the heritage advantage is expected only with early-acquired aspects of language and not those acquired later by children and/or are infrequent in spoken varieties (Montrul, 2008, p. 218).

1.2.2. HS/L2 comparisons: Input quantity and quality

When explaining why most HSs do not achieve native-like proficiency in their heritage language, most researchers consider reduced input as one of the major factors (e.g., Kupisch & Rothman, 2018; Montrul 2008; 2016; Polinsky, 2018b; Polinsky & Scontras, 2020). Unlike age effects, which predict an HS advantage, input conditions do not necessarily predict an HS advantage over L2ers. If only input (and not AoA) is relevant for ultimate attainment in language acquisition, then HSs and L2ers would be expected to perform very similarly under comparable conditions of reduced input. Of course, AoA and input effects are not mutually exclusive: both groups might be subject to reduced input and thus differ from monolinguals, though HSs could still have selective advantages due to AoA.

The quality of input also matters. Polinsky and Scontras (2020, p. 14) emphasized that "increased exposure to the heritage language will only get heritage speakers so far; they also need exposure from a variety of sources." Unlike predominantly monolingually-raised children who receive mostly NS input, HSs and L2ers likely have exposure to non-native or non-monolingual varieties, e.g., HSs from their parents who undergo L1 attrition (more below), L2ers from non-native teachers. While monolingual children and HSs receive naturalistic input since birth, L2ers mainly receive classroom input, at least initially.

Unfortunately, measuring the quantity and quality of the input is problematic. While L2ers' learning experiences might be relatively easy to track (e.g., how many courses/years of classroom learning were taken and how much time was spent living in countries where the target language is spoken), tracking HSs' language experiences from childhood is difficult. Most adult HS studies rely on retrospective self-reporting to estimate the input HSs have had (for child HSs, see Unsworth, 2019 for an overview on the parental questionnaire). Using bilingual child corpora (instead of adult monolingual ones) might be one way to approximate HS childhood input. Taken together with findings from experimental tasks, such corpora help to determine whether first-generation immigrants, whose speech constitutes the input for HSs, speak a variety different from the baseline spoken in the home country (due to L1 attrition, L2 transfer, or other factors; see Polinsky & Scontras, 2020, for a discussion on baselines in HS studies). If these bilingual children or child HSs are exposed to a qualitatively different variety compared to monolingually-raised children, it may explain why their ultimate attainment differs from that of monolingual adults.

An indirect way to examine input among HSs is to compare sequential bilinguals to simultaneous bilinguals. HSs can be classified as either sequential bilinguals or simultaneous bilinguals based on age of reduced input in the heritage language, which is also operationalized as age of onset of bilingualism (i.e., age of acquiring the majority language, i.e., English in the United States). Given that sequential bilinguals have a longer period of sustained exposure to the heritage language before being exposed to the majority language, the extent of "incomplete acquisition" is greater in simultaneous bilinguals than sequential bilinguals (e.g., Spanish HSs: Montrul, 2002).

A related concept under the broad term of input is frequency, which has become increasingly important in language acquisition discussions (e.g., variational learning model by C. Yang, 2004; see Slabakova, 2015a, who finds both L1 transfer and frequency to be important in L2

acquisition).[4] However, frequency as a determining factor of ultimate attainment has been undermined by the poor acquisition of agreement and case markers. Though they are highly frequent, they are two of the most vulnerable areas in heritage and L2 grammar (e.g., Polinsky, 2018b, Chapter 5; the Bottleneck Hypothesis, Slabakova, 2008, 2014, 2019). As warned by O'Grady, Lee, and Kwak (2009, p. 72, cited in Slabakova, 2015a), "[i]n considering the role of input frequency in language acquisition (first or second), it is vital to bear in mind a key point: what counts is not how many times learners hear a particular form—it is how many times they encounter mappings between a form and its meaning." Thus, while no one denies that input and frequency play an important role, exactly how much is sufficient for successful language acquisition remains to be explored.

1.2.3. HS/L2 comparisons: Dominant language transfer

Studies have shown that both HSs and L2ers are subject to dominant language transfer (e.g., Montrul, 2010a, 2014; Montrul & Ionin, 2010, 2012) or L1 transfer (e.g., Schwartz & Sprouse, 1996; White, 2003), resulting in non-native-like outcomes. Summarizing previous studies, Montrul (2016) concludes that dominant language transfer affects both HSs and L2ers and is more prominent in low-proficiency groups (p. 272), though complexity (such as interface phenomena, discussed below) and frequency are also important. However, with more language dyads studied, Polinsky (2018b) points out that HSs' dominant language transfer is not as pronounced as L2ers.[5] Importantly, dominant language transfer is often confounded with other factors such as

[4] Frequency here is used broadly to mean how often a certain linguistic structure occurs. Under different accounts, it can be frequency of the target form or frequency with which the relevant exemplars occur in the input.
[5] In addition to transfer, Polinsky and Scontras (2020) identify two other factors contributing to non-target-like patterns by HSs: attrition and divergent attainment, which require comparisons between adult HSs, child HSs, and the so-called "adult language forgetter" (e.g., immigrants undergoing L1 attrition after having acquired it completely before migrating); see Polinsky (2018a) for a range of comparisons.

simplification. The dominant language in most heritage language studies is English (see Scontras & Putnam, 2020, which counted the dominant languages represented in research articles published in the *Heritage Language Journal*) and these studies often find that HSs fail to acquire certain phenomena in their heritage language. If these phenomena are also absent in English (e.g., grammatical gender, case marking, long-distance reflexives), this failure could be due to transfer from English or overall simplification. The only way to disentangle these two explanations is to compare two HS groups with different dominant languages that differ in the relevant way. For example, even though both Korean and Mandarin have long-distance reflexives while English does not, J.-H. Kim (2007) found that Korean HSs in both the United States (English-dominant) and China (Mandarin-dominant) prefer local readings of reflexives compared to Korean NSs.

Another example is found in the acquisition of overt and null pronouns. Previous studies have assumed that the increased use of overt pronouns in Spanish by first-generation immigrants and especially second-generation immigrants (i.e., HSs) is primarily due to contact with English (e.g., Otheguy, Zentella, & Livert, 2007). However, other studies have shown that the increased use of overt pronouns appears even when both the societal and heritage languages allow null pronouns (e.g., Russian vs. Hebrew: Dubinina & Polinsky, 2013, cited in Polinsky, 2018b). Thus, studies with different languages in contact other than English are needed. Comparing two contact languages, van Osch (2019) found evidence for dominant language transfer as Dutch-dominant Spanish HSs (in the Netherlands) outperformed proficiency-matched English-dominant Spanish HSs (in the United States) on Spanish definiteness, which is more like Dutch.

1.2.4. HS/ L2 comparisons: processing consideration

Focusing on HSs, Polinsky and Scontras (2020) consider reduced input (both in quantity and quality) and less-efficient processing in the non-dominant heritage language as the two main potential triggers for non-native-like HS outcomes. Building on the psycholinguistic findings from monolingual adults and children, psycholinguistic studies on HSs can examine the role of processing more directly; see Felser (2020) and Gürel (2020) for some comments. The processing consideration has long been considered in all types of bilinguals (e.g., the Interface Hypothesis, Sorace, 2011). Studies with HSs using online tasks have started to emerge, mostly in Spanish (e.g., Jegerski, 2015, 2018a, 2018b; Jegerski, Keating, & VanPatten, 2016; Jegerski & Sekerina, 2020; Keating, Jegerski, & VanPatten, 2016) and Russian (e.g., Parshina, Laurinavichyute, & Sekerina, 2020; Sekerina & Sauermann, 2015; Sekerina & Trueswell, 2011; Meir, Parshina, & Sekerina, 2020) (see also Turkish: Jacob et al., 2019; Gračanin-Yuksek et al., 2020). However, comparisons between HSs and L2ers remain to be conducted. Given that the present dissertation uses only offline, untimed tasks, the role of processing will only be mentioned briefly.

1.2.5. HS/ L2 comparisons: task effects

Furthermore, HSs and L2ers may perform differently depending on the type of task. As HSs acquire the heritage language naturalistically from aural input, they are better at processing the language aurally and may have little metalinguistic awareness, especially if they have not been formally instructed in the heritage language. Thus, the advantage HSs have over L2ers may (only) show up in oral/aural tasks or tasks focusing on meaning (such as a picture-sentence matching

task), whereas L2ers may outperform HSs in written tests or tasks focusing on form (such as an acceptability judgment task) that requires more metalinguistic skills (Montrul, 2008).[6]

For example, in a study on HS relative clauses, Orfitelli and Polinsky (2017) found Russian HSs were native-like on a comprehension task (i.e., an auditory sentence-picture matching task), but not on a written grammaticality judgement test. Conversely, in a study on L2ers and grammatical gender, Grüter, Lew-Williams, and Fernald (2012) found that L2-Spanish learners were native-like in offline comprehension (i.e., a sentence-picture matching task), but not on elicited production and online processing.

By comparing HSs and L2ers on oral and written tasks, Montrul, Foote, and Perpiñán (2008) found that HSs outperformed L2ers of Spanish in an oral picture naming task while L2ers outperformed HSs in written tasks; all tasks examined Spanish gender agreement. Comparing form-focused and meaning-focused tasks (testing Spanish definite articles), Montrul and Ionin (2012) found that L2ers of Spanish performed better in a form-focused task (i.e., a sentence-picture acceptability judgment task) than in a meaning-focused task (i.e., a picture-sentence matching task), while Spanish HSs were native-like in both tasks. Note that explicitness plays a role when modality is held constant (written: Montrul & Ionin, 2012; spoken: Montrul et al., 2014). For example, Montrul et al. (2014) tested Spanish HSs and L2ers using three spoken word recognition tasks that varied in explicitness, and found that HSs were native-like on the most implicit task but patterned with L2ers in the other two tasks.

[6] However, Van Osch and Sleeman (2018) found that Spanish HSs in the Netherlands performed better on an acceptability judgement task than on an oral production task, which may be a consequence of different societal circumstances.

1.2.6. Vulnerable domains: Phonology vs. morphosyntax

By now, the emerging consensus is that the relative importance of AoA differs across linguistic domains (under a modular approach to language).[7] While HSs typically outperform L2ers on phonology (e.g., Korean: Oh, Jun, Knightly, & Au, 2003; Mandarin: C. Chang, Yao, Haynes, & Rhodes, 2011; C. Chang & Yao, 2016; Spanish: Au et al., 2002; J. Y. Kim, 2016, 2020), they do not necessarily have an advantage on morphosyntax.

To be specific, Oh et al. (2003) found that "childhood speakers" of Korean have an advantage over L2ers in both perception and production, while "childhood hearers" only have a perception advantage (but childhood hearers of Spanish showed a production advantage in Au et al., 2002). Similarly, J. Y. Kim (2016, 2020) found that Spanish HSs have an advantage over L2ers in perception, but not production, possibly because they hear Spanish more than they actually speak it.

Early exposure to a language gives learners an advantage in sounds even when they have no conscious recollection of it. Using functional MRI, Pierce et al. (2014) found that, unlike French monolinguals, international children adopted by French-speaking Canadian parents and who had stopped hearing Chinese by age one or two still showed brain responses to tones which were similar to those produced by Chinese-French bilinguals. However, Pallier et al. (2003) and Ventureyra et al. (2004) found no such language retention with Korean phonemes by internationally adopted children in France, which Oh et al. (2019) ascribed to later acquisition of

[7] Generative linguistics assume a modular approach to language: language faculty consists of different domains, or modules, such as phonology, semantics, syntax while the connections between modules are the "interfaces" (see Fodor, 1983; Jackendoff, 2002, also see Montrul, 2012, for discussing linguistic modularity and maturational effects across domains by HSs vs. L2ers). For a non-modular approach to language acquisition, see O'Grady (2005) for a non-generative, non-modular emergentist model.

phonemes in Korean than tones in Mandarin by monolingual children. For a review on language retention or loss in internationally adopted children, see Pierce, Genesee, and Klein (2019).

Within morphosyntax, most studies have found a selective HS advantage on core aspects of syntax, which develop before age three, but not in the domains of semantics, syntax-discourse interface, and inflectional morphology (for reviews, see Montrul, 2012, 2016; Polinsky, 2018b). However, these conclusions are mostly based on comparisons across multiple studies, without proficiency-matched participants. Moreover, many studies did not control for proficiency in the target language, which is problematic. Some studies did not independently test language proficiency (e.g., C. Chang et al., 2011), though some recruited students from classes of comparable levels (e.g., O'Grady et al. 2001; Lee 2016). Some used an independent proficiency test but found that participating HSs tended to be more proficient than L2ers (e.g., Spanish: J. Y. Kim, 2020), especially in languages where highly proficient L2ers are harder to find (e.g., Mandarin: C. Chen, 2019; Russian: Ionin et al. 2020). It is also possible that some mixed findings could be due to task effects; certain types of tasks might be biased in favor of different types of speakers/learners, as discussed above.

While not many studies have directly compared HSs and L2ers **across domains**, among those studies that do, there is an HS advantage with phonology but not morphosyntax (Spanish: Au et al., 2002, 2008; Knightly et al., 2003; Korean: Lee-Ellis, 2012). This finding points to the importance of AoA on phonology over morphosyntax, which is consistent with the general finding from L2 acquisition that it is more difficult for adult L2ers to be native-like in phonology than in morphosyntax (e.g., Abrahamsson & Hyltenstam, 2009; Granena & Long, 2013).

1.2.7. Vulnerable domains: Morphology vs. syntax vs. interface

Going beyond the split between phonology and morphosyntax, I discuss the relationship between morphology and syntax under the Bottleneck Hypothesis (Slabakova, 2008, 2014, 2019). The Interface Hypothesis (Sorace, 2011; Tsimpli & Sorace, 2006) is briefly discussed as the interfaces are considered one of the most challenging phenomena to acquire in all bilingual populations. While both hypotheses were initially proposed for L2 acquisition, they have been extended to HSs (Montrul, 2018; Montrul & Polinsky, 2011). The Interface Hypothesis states that external interfaces (between syntax and other cognitive domains, such as syntax-pragmatics or syntax-discourse) are more challenging to acquire than internal interfaces (such as syntax-semantics). A former version (Tsimpli & Sorace, 2006) states that interfaces are more difficult to acquire than, say, narrow syntax, without making the external vs. internal interface distinction. While the Interface Hypothesis is influential and supported in many earlier studies examining overt vs. null subjects (e.g., Sorace & Filiaci, 2006; Sorace & Serratrice, 2009; Tsimpli & Sorace, 2006), many recent studies with different interface phenomena or different languages do not find clear support for the updated hypothesis in L2ers (e.g., English and Spanish: Slabakova, 2015a; Basque: Rodríguez-Ordóñez & Sainzmaza-Lecanda , 2018; Japanese: Okuma, 2015; Spanish: Gómez Soler, 2017; Spanish and Greek: Margaza & Gavarró, 2020). The hypothesis is also not well-supported among HSs (e.g., Greek child HSs: Daskalaki et al. 2019; Spanish HSs: Leal, Rothman, & Slabakova, 2014; Leal Méndez, Rothman, & Slabakova, 2015; Hoot, 2017; Japanese/Korean HSs vs. L2ers: Laleko & Polinsky, 2016), but see Mai (2012) on Mandarin HSs and L2ers.

The Bottleneck Hypothesis states that morphology is more challenging for L2ers than semantics and syntax. Jensen, Slabakova, Westergaard, and Lundquist (2020) tested L1-Norwegian L2-English learners on both syntax (Norwegian having verb-second word order) and

functional morphology (Norwegian having subject-verb agreement) and found that L2ers perform better on syntax than functional morphology, supporting the Bottleneck Hypothesis. One strength of Jensen et al.'s (2020) study is that they used an acceptability judgment task to test both syntax and morphology to avoid possible task effects. Mikhaylova (2018) tested English-dominant Russian HSs (and in Mikhaylova, 2012, also L1-English L2-Russian learners) on Russian aspect and found that aspectual morphology is the bottleneck for acquiring Russian aspect, supporting the Bottleneck hypothesis. Similarly, Polinsky (2011) also uses the Bottleneck hypothesis to explain why adult Russian HSs have difficulty interpreting Russian relative clauses.[8]

While not positing a clear division between syntax, semantics, and morphology, Mai and Deng (2019) examines Mandarin HSs in different domains within morphosyntax using the *shì…de* cleft construction and compared them to L2ers from a previous study. Compared to L2ers, HSs were less influenced by English, but still show selective vulnerabilities in the heritage grammar in performing better on word order and the temporal feature than telicity and discourse features.

One problem with the Bottleneck Hypothesis is the difficulty of classifying some linguistic phenomena into certain domains. Under the framework of distributed morphology (Halle & Marantz, 1993), there is no division between syntax and morphology. Similar classification issues were raised with the Interface Hypothesis (e.g., see Slabakova, 2011, on how to classify Topic and Focus). I will discuss the relevance of these two hypotheses when classifying the four phenomena in this dissertation into different domains.

[8] Another study that potentially supports the Bottleneck Hypothesis is Håkansson (1995). Data collected from written material and spoken language from five Swedish HSs revealed attrition of noun phrase morphology, but not word order. However, this study has a small number of participants and the dominant languages of the HSs are different.

1.2.8. Literature gaps

To explore which domains are vulnerable to heritage and L2 acquisition, more studies need to test the same group of participants across domains (as called for in, e.g., Montrul, 2018). Additionally, in the literature of selective advantages by HSs over L2ers, extensive work has been done on Spanish (e.g., Montrul et al., 2008; Montrul & Ionin, 2012; J. Y. Kim, 2016, 2020) as well as Korean (e.g., O'Grady et al., 2001; Lee-Ellis, 2012; J.-H. Kim et al., 2010; S. Lee, 2012), but relatively little is known whether such selective advantages generalize to heritage Mandarin. Compared to Spanish, Mandarin HSs might have greater difficulty acquiring or maintaining Mandarin due to greater typological distance and the non-alphabetic writing system, among others. However, because Mandarin is hard for English speakers to acquire, the heritage advantage might be even more pronounced. While both HS and L2 groups might have difficulty achieving native-like proficiency, HSs might still outperform L2ers.

This dissertation aims to expand our knowledge of HS/L2 acquisition in two ways: (i) by examining whether the selective advantages found for HSs in other languages hold for Mandarin, which is typologically distant from languages used in previous HS/L2 comparisons and is among one of the hardest languages for English speakers to acquire; and (ii) by going beyond a phonology/morphosyntax comparison, and examining whether there are selective advantages within different subdomains of morphosyntax/semantics.

1.3. Organization

The rest of the dissertation is organized as follows. In Chapter 2, the properties of the four linguistic phenomena (**third tone sandhi, aspect marking, relative clauses,** and **long-distance reflexives**) are discussed as well as L1, L2, and heritage language acquisition of these four

linguistic phenomena, in that order. Two broad research questions are asked at the end of Chapter 2. Chapter 3 through Chapter 7 discuss the three experimental tasks (the Tone Identification Task, the Acceptability Judgement Task, and the Truth Value Judgement Task) testing the four linguistic phenomena and report on the results of those tasks. Chapter 3 presents the methods, participants, and results for the Tone Identification Task.[9] Chapter 4 presents the methodology and participants of the two judgment tasks on phenomena in morphology, syntax, and semantics, which are reported in Chapters 5-7 (Chapter 5 on aspect, Chapter 6 on relative clauses, and Chapter 7 on anaphors, focusing on long-distance reflexives). Chapter 8 provides a general discussion and conclusion.

[9] Given that the number of participants in the Tone Identification Task is considerably smaller than in the other two judgment tasks, I describe the participants and the results of the Tone Identification Task separately in Chapter 3.

Chapter 2. Four Mandarin phenomena under investigation

This dissertation investigates four linguistic phenomena in Mandarin. For each one, properties were chosen in consideration with English/Mandarin differences and prior literature. L1, L2, and heritage language acquisition of these phenomena are reviewed. The literature on Mandarin tone sandhi, aspect, relative clauses (RCs), and anaphors is extensive; I focus on the disyllabic sequence of tone sandhi, the interaction between grammatical aspect and lexical aspect, the subject-object asymmetry and head direction of RCs, and long-distance (LD) readings of reflexives. In discussing these phenomena, I also consider their frequency in the input, and ease or difficulty of processing since both frequency (e.g., Ellis, 2002, 2006) and processing may play a role in acquisition (HSs: Polinsky & Scontras, 2020; learners: Phillips & Ehrenhofer, 2015). However, I will not go into detail with regard to processing considerations, because the tasks in this dissertation are all offline, untimed tasks that cannot examine the role of processing. A pedagogical note on how these features are typically taught in the L2 classroom appears before the research questions and the hypotheses.

2.1. Mandarin tone sandhi

2.1.1. Properties of Mandarin tone sandhi

Mandarin has four lexically contrastive tones: high-level (T1), high-rising (T2), low-falling-rising (or low-falling; T3), and high-falling (T4). (I will leave out neutral tones in this dissertation.) Using the five-point pitch value developed by Chao (1930), T1 is transcribed as [55], T2 as [35], T3 as [214] (two other variants – [21] and [35] – are discussed below), and T4 as [51]. As a common notation, the pitch values are in brackets. In Pinyin (Romanized script), the tone number (1, 2, 3, and 4) is indicated after the syllable, e.g., “ma1” stands for “ma” pronounced with T1. The

syllable *ma* with T1 though T4 are thus ma1, ma2, ma3 and ma4, which mean *mother*, *hemp*, *horse* and *scold* respectively in Mandarin. An alternate way is to indicate the tone on the vowel, as in mā, má, mǎ, and mà. Table 2.1 summarizes the lexical tones in Mandarin.

Table 2.1. Mandarin lexical tones

	Height/contour	Pitch value	Examples
T1	high-level	[55]	ma1 'mother'
T2	high-rising	[35]	ma2 'hemp'
T3	low-falling-rising or low-falling	[214] or [21]; [35] under T3 sandhi	ma3 'horse'
T4	high-falling	[51]	ma4 'scold'

Tone sandhi is the tonal alternation in natural speech, and T3 sandhi is among the most studied. The T3 sandhi rule is that, when two underlying T3 syllables occur consecutively, the first syllable becomes T2 [35] (see e.g., Shih, 1986 for more analyses). For example, the Mandarin greeting 'ni3 hao3' ('you good') is pronounced as 'ni2 hao3' (but ni3 and hao3 if pronounced separately). In addition to disyllabic sequences, tone sandhi also applies in multi-syllabic sequences, which introduce more complexity (e.g., Speer, Shih, & Slowiaczek, 1989). When an underlying T3 syllable precedes a non-T3 syllable, namely T1, T2, or T4, the T3 syllable is pronounced as a half-T3 ([21]; a low-falling tone). While T3 sandhi is phonological in nature and has no straightforward phonetic explanation, the half-T3 rule is phonetically motivated in that T3 is changed from [214] to [21], without the final rise to reduce the articulatory effort (see e.g., J. Zhang & Lai, 2010). Unlike the T3 sandhi rule where T3 is changed to a categorically different tone (i.e., T2), the half-T3 is not categorically different from T3. Thus, some scholars consider half-T3 to be tonal coarticulation and not tone sandhi (e.g., Shih & Sproat, 1992, cited in W. Jin, 2019), while others consider half-T3 as another instance of the tone sandhi phenomena and name it "Half-T3 sandhi" (e.g., C. Yang, 2016; H. Zhang, 2013). The Mandarin T3 (sandhi) rules are captured in (1). In this

dissertation, I use full-T3 [214], sandhied T3 (or raised-T3 or T3S by some scholars) [35] (the T3 that undergoes tone sandhi and is realized as T2) and half-T3 [21] to describe the three phonetic variants/realizations of T3, while using T3 to mean the underlying form represented in speakers' minds.

(1) Mandarin T3 (sandhi) rules
Full T3 sandhi rule: T3 [214] → T2 [25] / __ + T3
Half-T3 (sandhi) rule: T3 [214] → low-falling tone [21]/ __ + T1/2/4

While many NSs and instructed learners are aware of the T3 sandhi rule, few are aware of the half-T3 rule. However, half-T3 actually has the widest distribution. While full-T3 [214] is traditionally described as occurring in isolation and utterance-final positions, Beijing Mandarin uses half-T3 in utterance-final positions (Duanmu, 2000; cited in H. Zhang, 2018b) and Taiwanese Mandarin even allows half-T3 in isolation (Tai, 1978, p. 117, cited in H. Zhang, 2018b). Currently, there is a debate as to whether the underlying form of T3 is indeed the full-T3 or if it should be changed to the more-widely distributed half-T3. The debate not only has implications for theoretical phonology but also in pedagogy (e.g., how T3 should be taught, which variant should be taught first, etc.) (see e.g., H. Zhang, 2017). It is not the goal of this present dissertation to solve this debate and I will still refer to T3 as the underlying form represented in speakers' minds.

Most perceptual studies have shown that NSs cannot differentiate between a T2 and a sandhied T3, though some studies have found the opposite (e.g., Y.-J. Lin & Y.-Y. Hsu, 2018). In production, some acoustic studies have found differences between them (for corpus studies, see, e.g., J. Yuan & Y. Chen, 2014). For example, C. Zhang and Peng (2013) tested adult NSs on both perception and production using pseudo-words and real words that are minimal pairs of similar frequency (er2yu3 兒語 'baby talk' vs. er3yu3 耳語 'whisper'). No perceptual differences were

found; acoustical differences were found from pseudo-words, but not from real words. Comparing sandhied T3 and half-T3, J. Zhang and Lai (2010) found that NSs made fewer mistakes in half-T3 than sandhied T3, since the former is more phonetically motivated.

Using event-related potentials (ERPs) in their psycholinguistics research on tone sandhi, C. Zhang, Xia, and Peng (2015) found that T3 sandhi (T3T3 sequences) is more difficult to "encode" in speech production compared to T2T3 sequences. This processing difficulty, however, has not been used to explain the findings from T3 sandhi acquisition studies.

In terms of frequency of the four Mandarin tones, while Tone 3 is the least frequent (e.g., at 16% based on characters, *Junda Chinese Text Computing*, cited in Xiaoqing Li & Y. Chen, 2015, p. 17, and also at 16% based on speech corpus, Y. Wu et al, 2020), learners presumably still receive much input of T3, given that there are only four tones. Some frequent T3 words include *wo* 'you', *ni* 'you', *hao* 'good; very', *hen* 'very/be', *xiao* 'small', *lao* 'old' or just an affix (as in *lao3shi1* 'teacher)'. For all disyllabic sequences, there are only 15 possibilities (due to T3 sandhi) and learners never hear the incorrect T3T3 pronunciations. Thus, T3 sandhi is presumably still more frequent than other phenomena tested in this dissertation. Some of the T3 words mentioned above are very commonly followed by other T3 words in disyllabic or multi-syllabic sequences, such as *xiao3gou3* '(small) dog', *hen3hao3* 'very good', and *wo3 hen3 hao3* 'I am (very) good'. Thus, in terms of frequency, both HSs and L2ers should have plenty of experiences with T3 sandhi, especially HSs who heard Mandarin at a younger age.

2.1.2. Acquisition of Mandarin tone sandhi

In L1 acquisition, lexical tone perception is developed before age one (see Tsao, 2016, for a review). Children acquire T1, T4, T2, and finally T3. Generally, tone errors are rare beyond age

two, though some have reported children who did not master all four tones by 2;6 (for reviews, see H. Zhu, 2016; Tsay 2016). However, a recent study by Wong and Strange (2017) found that the tone production by children as old as six was still rated significantly lower in accuracy compared to adults (see also Wong, 2012, 2013). Wong and Strange (2017) also found that children made more tone errors in first syllables than in second syllables.

The acquisition of tone sandhi has not been well-documented until recently. H. Zhu (2002) speculated that children might acquire tone sandhi very rapidly, making it difficult to pinpoint the exact timing in a cross-sectional study. Additionally, the acquisition of tone sandhi is difficult to study because highly frequent compounds such as 'ni2hao3' cannot be taken as evidence for knowing tone sandhi (such as English 'went' cannot be taken as evidence for knowing past tense); learners may just remember this sequence as a lexicalized chunk, and do not know that 'ni' is T3 in isolation. To establish whether participants know T3 sandhi, it is necessary to independently show that they know the two syllables are T3 in isolation, and it is only when the two syllables occur together that the first T3 becomes T2. While the earlier studies found that T3 sandhi is in place by age three when the data is transcribed by NSs based on perceptual transcriptions (e.g., H. Zhu, 2002; Y.-H. Huang, 2006; Wang, 2011, cited in P. Tang et al. 2019), recent acoustic studies have found that this is not the case (P. Tang et al., 2019; Xu Rattanasone et al., 2018). For example, P. Tang et al. (2019) found that while 3-year-olds were able to productively apply the T3 sandhi rule to novel disyllabic words, even 5-year-olds still differed from adults when applying this rule to trisyllabic words. Comparing the two variants, children acquired half-T3 earlier than sandhied T3.

For L2ers, tones are notoriously difficult (for reviews, see Y. Wang, Sereno, & Jongman, 2006; Y. Wang, Jongman, & Sereno, 2016; H. Zhang, 2018a). Generally, T1 and T4 are acquired

earlier than T2 and T3 (e.g., by L1-English/Japanese/Korean L2-Mandarin learners, H. Zhang, 2013). While some studies have found that learners who speak a tonal L1 outperform those who speak a non-tonal L1 such as English (e.g., Wayland & Guion, 2004, which found that Chinese speakers outperformed English speakers in discriminating Thai tones when both groups had no prior experience with Thai), Hao (2012) found that both NSs of Cantonese and English have difficulty distinguishing T2 and T3, with NSs of Cantonese having additional difficulty distinguishing T1 and T4. X. Wang (2006) also found that NSs of Japanese and English outperformed NSs of Hmong (a tone language) trying to learn Mandarin as a L2. Additionally, perception difficulty and production difficulty of tones do not always correlate. I focus on tone perception below.[10]

For L1-English L2ers, monosyllabic T4 is the easiest to identify in isolation (or in the final position) because its falling pitch is acoustically similar to the end of English declaratives (e.g., Y. Wang et al., 2006). Due to acoustic similarity (similar F0 contours), T2 and T3 are difficult to discriminate for both tonal and non-tonal L2ers, and under some circumstances, even for NSs (e.g., Y. H. S. Chang, 2011). For example, Hao (2012) found that both Cantonese and English NSs have problems differentiating T2 and T3 in Mandarin. Pelzl et al. (2019) further found that while T3 is difficult for both NSs and L2ers of Mandarin to identify, T2 is only difficult for L2ers.

Recent studies have found that L1-English L2ers who performed well on monosyllabic tones did not necessarily perform well on disyllabic tones (e.g., C. Chang & Bowles, 2015; Pelzl et al., 2019). For L1-English L2ers on disyllabic tones (e.g., Hao, 2018; Pelzl et al., 2019), T4

[10] Only studies on listeners who actually know Mandarin are reviewed here, since this dissertation is concerned with HSs and L2ers of Mandarin. For studies on listeners without Mandarin learning experiences, see e.g., Gandour (1983); Y. S. Lee et al. (1996); So and Best (2010), A. Chen et al. (2016).

becomes the hardest to identify when presented in non-final position due to interference from English intonation. The mutual confusion between T2 and T3 persists for disyllabic sequences in both first and second syllables (e.g., Q. Chen, 1997, cited in Hao, 2018). First syllables are also more difficult to identify than second syllables (e.g., Hao, 2012).

While some prior studies on disyllabic sequences exclude the T3T3 combination from the analysis due to T3 sandhi (e.g., Hao, 2018), several studies have begun to examine T3 sandhi in L2ers, mostly in production: Table 2.2 summarizes the L2 studies on T3 sandhi, including both production and perception studies. While correct production was possible when evaluated by NSs (e.g., C. Yang, 2016; H. Zhang, 2013), S. Chen et al. (2019), the first comprehensive acoustic study on T3 sandhi by L2ers, found that L2ers' production is not native-like. To my knowledge, the only perception study on T3 sandhi is H. Zhang (2017, 2018a, 2018b). With L1-English L2ers, both C. Yang (2016) and W. Jin (2019) found higher accuracy for half-T3 than full-T3 (significant differences in C. Yang's study, but not in W. Jin's) while H. Zhang (2017; see also H. Zhang, 2013, 2018a, 2018b) found the opposite. Thus, this issue is not settled yet. While the T3 sandhi rule is more commonly taught than the half-T3 rule (survey results from teachers: C. Yang & W. Jin, 2018; survey results from students/L2 participants: W. Jin, 2019), W. Jin (2019) found that L2ers performed better on half-T3 than sandhied T3 and attributed this finding to the stronger phonetic motivation of the half-T3 than sandhied T3. Specifically, H. Zhang (2018b) found that half-T3 (low-falling) [21] is mostly misperceived as T4 (high-falling) [51] because both have falling contours; furthermore, half-T3 is also sometimes misperceived as the neutral tone because both half-T3 and the neutral tone have a short duration. Similarly, in production, half-T3 (low-falling) [21] is mispronounced as T4 (high-falling) [51] because both have falling contours.

Table 2.2. Selected studies on Mandarin T3 sandhi by L2ers

	L1s	Task format	Pattern summary or key finding
C. Yang (2016; Ch 6)	English	Production (familiar words and non-words)	Significantly higher accuracy of half-T3 than T3 sandhi
H. Zhang (2018b; Ch 6)	L1-English/Japanese/Korean L2ers on disyllabic production; L1-English L2ers on trisyllabic production and perception	Production (real words embedded in sentences) and perception (transcribing the tones of the pseudo-words they heard)	Higher accuracy for T3 sandhi than half-T3 (for beginner and intermediate L2ers, but not for advanced L2ers) in production; L2ers were unable to transcribe half-T3 as T3 in perception
S. Chen et al. (2019)	Cantonese & English	Production, including both real words and wug (nonsense) words	L2ers' production is not native-like, with the Cantonese group performing better than the English group
W. Jin (2019)	English	Production (passage readings, thus real words)	Higher accuracy for half-T3 than T3 sandhi, but no significant differences; similar accuracy for half-T3 in T3T4, T3T2, and T3T1

While heritage phonology is a burgeoning field (see C. Chang, to appear, for an overview; e.g., Tse, 2016, on Toronto heritage Cantonese), studies on heritage Mandarin, besides C. Chang and colleagues and B. Yang (2015; Ch 6), remain scarce,. Chang and colleagues found that HSs have an advantage over L2ers in producing Mandarin vowels, plosives, retroflex (C. Chang et al., 2011) and tones (C. Chang & Yao, 2016), but not with neutral tones (Chang & Yao, 2019).[11] Examining trisyllabic sequences, B. Yang (2015) also found that HSs outperformed L2ers in recognizing the starting point of tones in their perceptual space. A related study by Tsukada et al. (2015) found that HSs of Cantonese did not outperform L1-English L2ers in discriminating

[11] C. Chang and Yao (2019) found that L2ers patterned more like NSs than HSs in some (but not all) of the measures on neutral tones. However, as acknowledged by the authors, since neutral tones in non-obligatory contexts vary across dialects (more consistent in northern Mandarin but not in southern Mandarin), L2ers' advantage over HSs in neutral tones may be due to (1) L2ers' familiarity with the standard northern Mandarin commonly taught in L2 classrooms, and (2) some HSs' exposure to southern Mandarin.

monosyllabic Mandarin tones and even had additional problems differentiating between Mandarin T1 and T4, which they map to Cantonese T1, similar to finding from L1-Cantonese L2-Mandarin learners (Hao, 2012). To my knowledge, no study has examined the perception of T3 sandhi by HSs or compared HSs and L2ers on this phenomenon.

T3 sandhi is chosen for this dissertation to represent the subdomain of phonology because (a) it has no equivalent in English, and should be equally difficult for HSs and L2ers under dominant language transfer; (b) it is an early-acquired phenomenon, so HSs may have a selective advantage due to age of acquisition (AoA); and (c) it is a complex phenomenon which is typically not emphasized in the classroom, unlike basic tones.

2.2. Mandarin aspect

2.2.1. Properties of Mandarin aspect

Unlike English, Mandarin lacks (overt) tense morphology (like past tense *-ed*) (e.g., Lin, 2003) but has a rich aspectual system (like progressive *-ing*). Mandarin has two perfective markers *-le* and *-guo*, and two imperfective markers *zai* and *-zhe*.[12] While *-le*, *-guo,* and *-zhe* appear after the verb as suffixes, *zai* appears pre-verbally as a separate word. All four can be optional depending on contexts. The perfective markers tend to be interpreted as past and the imperfective markers as present, but this is not always the case. In (2), the bounded event marked by *-le* can take place in the future, suggesting that *-le* does not indicate past tense.[13] The perfective marker *-le* signals the initiation or termination, and not necessarily the completion of an event. In (3) (Smith, 1997, p.

[12] This verb-final *-le* has the same form as the sentence-final *LE*, which marks a "currently relevant state" not discussed in this dissertation. I gloss the verb-final aspect marker as *-le* and the sentence-final particle as *LE*.
[13] The abbreviation used in this dissertation: CL = 'classifier'. The four aspect markers *-le, -guo, zai* and *-zhe* are glossed as perf(ective), exp(eriential), prog(ressive), and dur(ative) respectively.

68), the completion of *xie* 'write' is marked by a separate morpheme *wan* 'finish'. The sequence 'write-finish' is an example of Resultative Verb Compounds commonly used in Mandarin.

(2) Wǒ míngtiān xià-le bān qù kàn diànyǐng
I tomorrow off-perf work go see movies
'Tomorrow I will go to the movies after work'.

(3) Wo zuotian xie-le yi-feng xin, keshi mei xie-wan
I yesterday write-perf one-CL letter, but not write-finish
'I wrote a letter yesterday, but did not finish it.' (infelicitous in English)

As an experiential marker, *-guo* is incompatible with future adverbials. A notable difference between *-guo* and *-le* is that *-guo* signals a discontinuity with the present, but *-le* does not (e.g., Smith, 1991). In (4), with *-le* in (a), the interpretation is that they may or may not still be in Hong Kong; with *-guo* in (b), the interpretation is that they are no longer in Hong Kong (examples from Smith, 1994, p. 117).

(4) a. Tamen shang ge yue qu-le Xiang Gang
they last CL month go-perf Hong Kong
'Last month they went to Hong Kong (they may still be there)'

b. Tamen shang ge yue qu-guo Xiang Gang
they last CL month go-exp Hong Kong
'Last month they went to Hong Kong (and they are no longer there).'

The progressive marker *zai* focuses on the progressive phase of an event and is semantically close to the English progressive marker (e.g., Smith, 1997, p. 272). It marks unbounded events in progress, and tends to be interpreted as present. However, *zai* can be used with past adverbials, as in (5).

(5) Wǒ zuótiān xiàwǔ yīzhí zài xiě zuòyè
I yesterday afternoon always prog write homework
'I was doing homework yesterday (all) afternoon'.

The durative marker *-zhe* emphasizes the duration of an event and the resultant state. The major function of *-zhe* (51% of the time according to R. Xiao & McEnery, 2004b, p. 184) is to "express overlapping actions in the background" ; see (6). Verbs denoting posture or location can only occur with *-zhe*, but not *zai*; see (7) (see e.g., Woo, 2015). Mandarin *-zhe* marks unbounded situations, and tends to be interpreted as present. It is sometimes labelled as a "stative marker" (e.g., M. Liu, 2015) or "continuous marker" (e.g., S.-W. Tang, 2016). Some researchers argue that there are two kinds of *-zhe* (progressive and resultant-stative), depending on the verb, e.g., Tsai (2008).

(6) Na haizi ku-zhe yao baba
that child cry-due want dad
"While crying, that child called out for her father"
(7) Lisi zuo-zhe
Lisi sit-dur
'Lisi is sitting (somewhere).'

Importantly, these grammatical aspect markers cannot freely combine with all lexical predicates. Below I introduce Vendler's four-way distinction of lexical verbs (1967) before the interaction between grammatical aspect and lexical aspect. The four-way distinction is based on three binary semantic features: dynamicity, telicity, and punctuality (see Table 2.3). State verbs are non-dynamic while the other three are dynamic events. Both state and activity verbs are non-telic, which means that they do not have an endpoint, while both accomplishment and achievement verbs are telic, i.e., have an endpoint. Punctuality indicates that an event is instantaneous and non-durative; only achievement verbs are punctual. Note that the complements or arguments of the verbs matter: while 'run' is an activity verb, 'run a mile', which is technically a verb phrase (VP), is an accomplishment verb. This dissertation uses the term 'states', 'activities', 'accomplishments', and 'achievements' at the VP level hereafter.

Table 2.3. Four verb classes (Vendler, 1967)

	[± Dynamic]	[± Telic]	[± Punctual]	Examples
States	-	-	-	Know, love
Activities	+	-	-	Run, walk
Accomplishments	+	+	-	Run a mile, paint a picture
Achievements	+	+	+	Recognize, find

Vendler's framework of 'verb class' has been criticized and replaced by 'situation type' or 'situation aspect' in Smith (1991, 1994); additional verb classes have been proposed, such as semelfactive (Smith, 1991, 1994) and mixed telic-stative in Mandarin (P. Li & Bowerman, 1998), which are not discussed here. Indeed, disagreement exists on verb classification, perhaps to a greater degree in Mandarin. According to Tai (1984), Mandarin does not have accomplishments and many English achievements are realized in Mandarin as Resultative Verb Compounds (e.g., pick-descend, fall-down, write-finish). Whether Resultative Verb Compounds are accomplishments (Smith, 1991; Y. Li, 2016) or achievements (Tai, 1984; P. Li & Shirai, 2000) is controversial. For states, researchers such as Smith (1991) and R. Xiao and McEnery (2004b; see also Z. Xiao and McEnery, 2004a) classify stage-level states as temporary and changeable (e.g., *bing* 'be ill' and *mang* 'be busy') while individual-level states are permanent and intrinsic (e.g., *xiang* 'resemble,' *xìng* 'have the last name of' and *piaoliang* 'beautiful'). Mandarin states such as 'be ill', 'be busy' and 'beautiful' might be classified as adjectives rather than verbs in some analyses. Given that a Mandarin adjective can be the predicate in a sentence without a copula, it is debated whether Mandarin has adjectives as a separate syntactic category. In this dissertation, Mandarin adjective-like verbs (or simply adjectives) are included as state verbs following previous studies on aspect (e.g., R. Xiao & McEnery, 2004b; J. Chen & Shirai, 2010).

I keep Vendler's four-way classification because it is commonly used in acquisition research and considered valid in Mandarin under some accounts (e.g., Soh & J. Kuo, 2005; Y. Li, 2016). Soh and J. Kuo (2005) argue that Mandarin accomplishments pattern like English accomplishments when the direct objects are quantified (by numerals since Mandarin lacks articles).

Below I discuss the interaction between grammatical and lexical aspect. While exceptions exist, the general pattern is summarized in Table 2.4 (see Smith, 1997; R. Xiao & McEnery, 2004b). Perfective *-le* is largely compatible with all four lexical predicates. However, *-le* is compatible only with stage-level states, but not individual-level states.[14] For individual-level states and activities to be compatible with *-le*, extra delimiting mechanisms are required to provide an endpoint (R. Xiao & McEnery, 2004b, p. 107-111). Without extra delimiting mechanisms, states and activities with *-le* result in incomplete sentences; see examples (8) and (9) (S. Yang, 1995, cited in R. Xiao & McEnery, 2004b, p. 102-3). In (10) (example from F.-H. Liu, 2017, p. 218), activities with *-le* are made complete be adding another clause or sentence-final *LE*. In seemingly acceptable cases without objects, *-le* is ambiguous between the verb-final *-le* and the homophonous sentence-final *LE*.

(8) a. *Liming ai-le Xiaojuan
*Liming love-perf Xiaojuan
'Liming loved Xiaojuan'

b. Liming ai-le Xiaojuan san-nian
Liming love-perf Xiaojuan three-year
"Liming loved Xiaojuan for 3 years"
(States bounded by a for-adverbial)

[14] A. Li (2016, p. 82) describes a rare case where a state verb "has a starting point" and followed by *-le* (gloss modified; DE is a possessive marker).
Bàba bǎ wǒ guòjì gěi gūgū wǒ jiù xìng-le gūfu de xìng
father BA I adopt (continue family line) to aunt I then surname-perf uncle-in-law DE surname
'Father gave me to his sister for adoption, and I took her husband's last name.'

(9) a. *Lisi tui-le che
*Lisi push-perf cart
"Lisi pushed the cart"

b. Lisi tui-le tui che
Lisi push-perf push cart
"Lisi pushed the cart a bit"
(Activity bounded by verb reduplication)

(10) Xiaowang chi-le pingguo, Xiaozhang chi-le xiangjiao.
Xiaowang eat-perf apple Xiaozhang eat-perf banana
'Xiaowang ate (some) apple; Xiaozhang ate (some) banana.'

Table 2.4. Interaction between lexical and grammatical aspect in Mandarin (✓= grammatical; ✗= ungrammatical; ? = marginal)

	Perfective-*le*	Experiential -*guo*	Progressive *zai*	Durative -*zhe*
States[a]	✗ (incomplete)	✓	✗	✗
Activities	✗/? (incomplete)	✓	✓	✓/? (incomplete)
Accomplishments	✓	✓	✓	✗/?
Achievements	✓	✓	✗	✗

Note: [a]States here include only individual-level states

Experiential -*guo* is largely compatible with all four lexical predicates (e.g., R. Xiao & McEnery, 2004b, p. 143). Progressive *zai* is compatible with activities and accomplishments, but not achievements and individual-level states (as *zai* marks unbounded events in progress). Thus, unlike English, *wo (xianzai) zai dao* I now *zai* arrive 'I am arriving (now)' is ungrammatical in Mandarin. (Not all achievements in English are compatible with -*ing,* as examples such as 'I am noticing a problem' or 'I am losing the keys' are ungrammatical according to e.g., Kearns (2011, p. 165); however, they may be acceptable by some NSs.) Similarly, Resultative Verb Compounds pattern with achievements in being incompatible with *zai*.[15] The ungrammatical combination of *zai* with

[15] To be specific, there are different kinds of Resultative Verb Compounds. Those that are [+durative] (similar to 'degree achievement' in the "scale structure" framework) are compatible with *zai* (e.g., H. Xu, 2015).

achievements and Resultative Verb Compounds is attested in both L2 and heritage language acquisition (discussed later).

Like *-le* and *zai*, durative *-zhe* is compatible with stage-level states, but not individual-level states (Smith, 1997, p. 273; but see R. Xiao & McEnery, 2004b, p. 189, for some exceptions). Like *zai*, *-zhe* is incompatible with achievements. While *-zhe* is considered to be compatible with activities by some scholars (e.g., C.-T. J. Huang et al., 2009, p. 101; R. Jia, 2016, p. 114, R. Xiao & McEnery, 2004b), others argue that this combination results in incomplete sentences (e.g., J.-I. Li & Hsieh, 2015; Y. Guo, 2020); see (11) (example from J.-I. Li & Hsieh, 2015, p. 30; gloss modified). With some exceptions (R. Xiao & McEnery, 2004b, p. 193), *-zhe* rarely occurs with accomplishments. Y. Guo (2020) classified *-zhe* with accomplishments as ungrammatical while R. Jia (2016) classified it as grammatical (but the example she gave was "draw picture" without a quantified direct object, which is required in Mandarin accomplishments, according to Soh & J. Kuo, 2005).

(11) a. Ta zai fangjian li ting-zhe yinyue.
He at room in listen-dur music
'He is listening to the music in the room.'
b. ?Ta ting-zhe yinyue.
he listen-due music

In psycholinguistics, many studies have examined tense/aspect across languages, but few have examined Mandarin aspect. For example, to investigate English tense/aspect, one line of research is to use ERP methodologies to examine morpho-syntactic violations (disagreement/incongruity) by creating incongruous temporal contexts (e.g., by adverbials) and morpho-syntactic markers (e.g., Flecken, Walbert, & Dijkstra, 2015). In Mandarin, without (overt) tense markers, such

incongruity is created by incongruous temporal context (adverbials) and aspect markers or aspectual expressions (e.g., Qiu & Zhou, 2012; Collart & Chan, 2019) or by putting progressive *zhengzai* and perfective *-le* together (e.g., Y. Zhang & J. Zhang, 2008).[16] Note that while *-le* can be used with future adverbials in sentences like (2), repeated below as (12), future adverbials with *-le* is ungrammatical in simple declarative sentences, as in (13) (taken from Collart & Chan, 2019; gloss modified). Processing studies on aspectual violations are less informative to the current dissertation as I test incompatibility of lexical aspect and aspect markers (without giving temporal contexts).

(12) Wǒ míngtiān xià-le bān qù kàn diànyǐng
I tomorrow off-perf work go see movies
'Tomorrow I will go to the movies after work'.

(13) Mama zuotian / #mingtian xi-le yifu.
Mother yesterday / #tomorrow wash-perf clothes
'Yesterday/#Tomorrow, mom washed the clothes.'

Another line of research examines how speakers process perfective vs. imperfective for (in)completed events. While perfectives are often processed faster than imperfectives with accomplishments (e.g., Cantonese: Yap et al., 2009; Mandarin: Yap et al., 2004, cited in Yap et al., 2009; English: Madden & Zwaan, 2003), imperfectives are processed faster with activities (e.g., Cantonese: Yap et al., 2009). Such associations are compatible with the prototype account (e.g., Li & Shirai, 2000; Shirai & Andersen, 1995), which, in the L1 context, means that "children tend to create more dramatic association based on skewed distribution in the adults' input." (J. Chen & Shirai, 2010, p. 19); see also the Distributional Bias Hypothesis (e.g., Andersen & Shirai, 1996). Potentially relevant to the present dissertation is the fact that processing *-le* with accomplishments

[16] For the differences between *zai* and *zhengzai*, see J.-H. Wang (2015).

and *zai* with activities is easier for NSs, and by extension, to HSs and L2ers. Given that the present dissertation adopts an offline task to test aspect, I do not focus on the processing considerations, though it might indeed affect the acceptability ratings participants give, with higher acceptance to only the more prototypical/ frequent ones, discussed next.

In terms of frequency on aspectual marking, the range varies widely. In elicited production, L. Jin and Hendriks (2005) found that less than 40% of the predicates were marked with aspect markers by adult NSs (as well as L1 children and L2ers), and more than 60% of them with achievements. In audio- and video-recorded conversations, C.-C. Huang (2003) found only 21%-28% of the verbs in child-directed speech and only 2%-4% in adult-to-adult speech were marked with aspect markers. Using three corpora of the same size, S. Yang and Y. Huang (2013, cited in F. Wang & F. Wu, 2020) found that the frequency of occurrence of aspect markers differs across genres: 28% in the spoken corpus, 53% in the novel corpus, and 18% in the news corpus. Aspect markers are probably less necessary in spoken Mandarin due to the "present-time orientation" but more necessary in novels to frame events (F. Wang & F. Wu, 2020).

Among the four aspect markers, *-le* is by far the most frequent, followed by *-zhe*, while *zai* and *-guo* are of similar frequency (the written corpus in R. Xiao & McEnery, 2004b; the speech corpora in J. Chen & Shirai, 2010). The distribution of aspectual markers heavily depends on the lexical predicates. To get a sense of how strong the four aspect markers are associated with different lexical predicates, I summarize the corpus findings from R. Xiao and McEnery (2004b), which used the Weekly corpus based on newspaper texts. (The figures below do not necessarily add up to 100% as they classified predicates into six categories instead of four and the numbers reported here are rounded.) For *-le*, 50% of the predicates marked with *-le* are achievements, 30% are accomplishments, 13% are activities, 3% are individual-level states, and 2% are stage-level

states (R. Xiao & McEnery, 2004b, p. 104). For *-guo*, 43% of the predicates are accomplishments, 30% are activities, 17% are achievements, and 8% are individual-level states (R. Xiao & McEnery, 2004b, p. 143). For *zai*, 83% of the predicates are activities, 9% are accomplishments, 3% are achievements, and 2% are stage-level states (R. Xiao & McEnery, 2004b, p. 209). For *-zhe*, glossing over the usages, 55% of the predicates are activities, 27% are stage-level states, 15% are individual-level states, and less than 1% are accomplishments (R. Xiao & McEnery, 2004b, p. 188). Similar to the processing consideration above, the frequency might explain the performance of HSs and L2ers, and even NSs, in an acceptability task used in the present dissertation.

2.2.2. Acquisition of Mandarin aspect

Longitudinal data from Mandarin-speaking children indicates that the acquisition order is *-le*, followed by *zai* and *-zhe*, and finally *-guo* (e.g., Erbaugh, 1992). Studies have shown that, by age three, Mandarin-speaking children use *-le* with different lexical aspect classes, although mostly with achievements, followed by accomplishments (P. Li & Bowerman, 1998). (In L. Jin and Hendriks (2005), adults and 10-year-olds predominantly only use *-le* with achievements, followed by accomplishments, without using *-le* with neither activities nor states.) Children use *zai* exclusively with activities; only at age five (or later) do they began to use *zai* with accomplishments (P. Li & Bowerman, 1998). While early natural production studies found children to never make mistakes using *zai* with states (Erbaugh, 1978, 1992; J. Chen & Shirai, 2010), such mistakes were found in elicited production (P. Li & Bowerman, 1998) and elicited story telling (20% in 5-year-olds in L. Jin & Hendriks, 2005). While P. Li and Bowerman (1998) found that 5-years-old incorrectly produced *zai* with states in elicited production (though only rarely), children did not incorrectly produce *zai* with achievements (Resultative Verb Compounds

in their study). For *-zhe*, L. Jin and Hendriks (2005) found that 5-year-olds misused *-zhe* with achievements at 14% (higher with statives and activities – recall that *-zhe* with statives and activities are not grammatical in simple declarative sentences). Both *zai* and *-zhe* first emerge with activities before extending to accomplishments (J. Chen & Shirai, 2010). For *-guo*, this emerges with accomplishments and achievements before extending to atelic verbs (J. Chen & Shirai, 2010).

L2-Mandarin studies have generally found that learners are sensitive to the interaction between lexical and grammatical aspect; see F.-H. Liu (2017) for an overview and Table 2.5 for a selective summary. A common pattern is the misuse and mis-acceptance of *zai* with achievements and Resultative Verb Compounds, e.g., L. Jin and Hendriks (2005), F.-H. Liu (2012), and Y. Guo (2020). I discuss F.-H. Liu (2012) and Y. Guo (2020) in detail as they included judgement tasks (among others), which this dissertation uses to test aspect. Both studies examined imperfective markers by L1-English L2-Mandarin learners at three proficiency levels. F.-H. Liu (2012) found that L2ers patterned with NSs in allowing *zai* with accomplishments, but only the more advanced learners were native-like with activities and only the most advanced learners were native-like with states and achievements. Y. Guo (2020) examined both *zai* and *-zhe*: L2ers incorrectly accepted *zai* and *-zhe* with achievements (though the advanced group correctly rejected *-zhe* with achievements), but correctly accepted *zai* with accomplishments (though only the advanced group correctly rejected *-zhe* with accomplishments). Y. Guo (2020) additionally tested the incompleteness effect of *-zhe* using a sentence completeness judgment task (to compare *zai* and *-zhe*) and found that L2ers could not detect the incompleteness effect of *-zhe*.

While aspect has been much studied with HSs in morphologically complex languages (e.g., Spanish: Montrul, 2002; Russian: Mikhaylova, 2018, 2019), only a few studies have examined Mandarin aspect by HSs; see Table 2.6.

Table 2.5. Selected studies on Mandarin aspect by L2ers, focusing on the interaction between grammatical and lexical aspect (Resultative Verb Compounds = RVCs)

	L1s	Task format	Pattern summary or key finding
L. Jin & Hendriks (2005) on *-le, zai,* and *-zhe*	English	Two elicited production tasks	L2ers incorrectly used *zai* with achievements and RVCs, possibly due to L1-English transfer
F.-H. Liu (2012) on *zai*	English	A judgement task	L2ers incorrectly accepted *zai* with achievements, possibly due to L1-English transfer
Y. Shi (2013) on *-le*	English (also the dominant language of HSs in the United States)	A fill-in-the-blank task (add *-le* if needed)	Undersupply of *-le* in accomplishments, achievements, and RVCs, possibly due to being overly cautious
Y. Guo (2020) on *zai* and *-zhe*	English	A grammaticality judgment task	L2ers incorrectly accepted *zai* with achievements, possibly due to L1-English transfer

Table 2.6. Selected studies on Mandarin aspect by HSs, focusing on the interaction between grammatical and lexical aspect (Resultative Verb Compounds = RVCs)

	Dominant languages of HSs	Task format	Pattern summary or key finding
L. Jia & Bayley (2008) on *-le*	English (for child HSs in the United States)	An elicitation task, a multiple-choice test (choose among *-le*, other aspect markers, or leave it blank), and a sentence-completion task (fill in *-le* or leave it blank)	No influence of lexical aspect on the use of *-le*
M. Shi (2011) (master's thesis)	Mostly Dutch (for adult HSs in the Netherlands)	Two elicited production tasks	HSs incorrectly used *zai* with achievements and RVCs, possibly due to the imperfective being acquired later than the perfective and the infrequent input of *zai* (per J. Chen & Shirai, 2010)
Y. Shi (2013) on *-le*	English (also the L1 of L2ers)	A fill-in-the-blank task (add *-le* if needed)	Undersupply of *–le* in accomplishments, achievements, and RVCs, possibly due to being overly cautious; HSs outperformed L2ers
R. Jia (2016) on *le-, zai,* and *-zhe*	English (for child HSs in Canada)	An elicited production task and a grammaticality judgment task	HSs incorrectly produced and accepted *zai* with achievements and RVCs due to English transfer of *-ing*

Unlike other studies, L. Jia and Bayley (2008) did not find influence of lexical aspect on the use of *-le*. R. Jia (2016, p. 134) speculates that L. Jia and Bayley's frog story elicitation task might not be able to elicit a wide variety of lexical verbs. Like L1-English L2ers (e.g., L. Jin & Hendriks, 2005), M. Shi (2011) found that adult HSs in Netherlands misused *zai* with achievements and Resultative Verb Compounds (however, this study had very few participants). M. Shi interpreted such mistakes in two ways: first, the imperfective was acquired after the perfective, and second, (parental) input of *zai* was rare as reported in J. Chen and Shirai (2010). R. Jia (2016) also found that child HSs incorrectly produced and accepted *zai* with achievements and Resultative Verb Compounds, and interpreted them as progressive *-ing* due to English transfer. Y. Shi (2013), the only study to compare HSs and L2ers in this domain, found an advantage for HSs in *-le*. However, this study had unequal, non-proficiency-matched groups. Thus, it is unknown whether the differences between HSs and L2ers were due to language proficiency or learner types.

The Aspect Hypothesis (Shirai & Andersen 1995; Andersen & Shirai, 1996) is briefly discussed next. It is not the focus of the present dissertation because this hypothesis predicts emergence and developmental trajectory, which is better examined by longitudinal data or at least different proficiency groups.[17] Based on the lexical aspects proposed by Vendler (1967), the Aspect Hypothesis is stated in (14) (taken from J. Chen & Shirai, 2010, p. 2):

(14) A. Children first use past or perfective marking on achievement and accomplishment verbs, eventually extending its use to activity and stative verbs.

B. In languages that encode the perfective/imperfective distinction, imperfective past develops later than perfective past.

C. In languages that have progressive aspect, children first use progressive aspect marking mostly with activity verbs, then extend [it] to accomplishment and achievement verbs.

D. Children do not incorrectly overextend progressive aspect markings to stative verbs.

[17] The differences between the Aspect Hypothesis and the prototype hypothesis mentioned earlier is that the latter does not have a developmental component (F. H. Liu, 2017, p. 220).

Applying the Aspect Hypothesis to Mandarin, Prediction A predicts that *-le* and *-guo* are first used with achievements and accomplishments before activities and states. Prediction B predicts that *-le* and *-guo* are acquired earlier than *zai* and *-zhe*. Prediction C predicts that children will use *zai* first with activities, before extending to accomplishments. Note that following Prediction C, learners will then extend *zai* to achievements, but this combination is ungrammatical (see F.-H. Liu, 2017, for some critique on the Aspect Hypothesis). Closely related to Prediction C, Prediction D predicts that children will not, by overextension, incorrectly use *zai* with states.

Most studies with Mandarin-speaking children have generally found support for the Aspect Hypothesis, including Li (1990), Li & Bowerman (1998), and J. Chen & Shirai (2010). However, J. Chen & Shirai (2010) has also provided evidence against the hypothesis. They examined a longitudinal corpus and found early emergence of *-le* with all four lexical predicates, including states, due to high frequency in the parent input, which was also examined by the authors. (The examples given by J. Chen & Shirai (2010) seem to be sentence-final *LE* rather than verb-final aspectual *-le*. In the present dissertation, I focus on individual-level states that are incompatible with *-le* to examine possible English transfer of past tense.)

In L2-Mandarin, while L. Jin and Hendriks (2005) supports the Aspect Hypothesis, Tong and Shirai (2016) does not, especially not its developmental prediction (contrary to the prediction, they actually found stronger associations between aspect markers and lexical aspect at later stages of development). For an updated discussion on the Aspect Hypothesis in L2-Mandarin, see F.-H. Liu (2017). As mentioned above, F.-H. Liu (2017) pointed out the fact that Mandarin progressive *zai* does not occur with achievements is contrary to Prediction C. In addition, the Aspect Hypothesis does not address L1 transfer found in many L2 studies, though F.-H. Liu (2017) acknowledges that

L1 transfer has been addressed in subsequent work (see Shirai, 2007). In heritage Mandarin, only M. Shi (2011) discusses the Aspect Hypothesis and seems to find support for it.

This dissertation adds to the literature by comparing proficiency-matched HSs and L2ers on all four aspect markers. While aspect is at the interface of syntax, semantics, and morphology, I classify the property examined in this dissertation as falling at **the morpho-semantics interface** because the task tested the acceptability of aspectual marking on different kinds of lexical predicates. **This phenomenon is chosen because** (a) there are clear differences between Mandarin and English aspect, allowing for an investigation of cross-linguistic influence; (b) aspect markers are an early-acquired phenomenon, which may confer an advantage onto HSs; and (c) morphology has been termed the 'bottleneck' of L2 acquisition (Slabakova 2008, 2014), so this domain is expected to be particularly challenging for L2ers (though see Montrul, 2018, which extends the Bottleneck Hypothesis to HSs).[18] Note that it is presently unclear whether the Bottleneck Hypothesis applies equally to Mandarin, a language with impoverished inflectional morphology, and to morphologically rich languages.

2.3. Mandarin relative clauses (RCs)

2.3.1. Properties of Mandarin relative clauses

In the English complex noun phrase (NP) *the woman [who sees the man]*, *woman* is the 'head noun' of a RC *who sees the man.* (A 'head noun' is the 'head' of a noun phrase, modified by an adjective or a RC.) Unlike the English 'head noun' which precedes the RC, the head noun

[18] In Slabakova (2015b), L1-English L2-Mandarin learners successfully acquire the temporal meaning in Mandarin despite lack of tense in Mandarin. Unlike the studies reviewed above that focused on forms (grammatical or ungrammatical combinations), Slabakova (2015b) focuses on meaning informed by a universal deictic pattern of temporality: bounded/telic events tend to be interpreted as past and unbounded/atelic events as present (Smith & Erbaugh, 2005).

of Mandarin follows the RC. Thus, English RCs are head-initial while Mandarin RCs are head-final. There are different types of extraction, most commonly subject-extracted RCs (SRCs) and object-extracted RCs (ORCs); see Table 2.7.[19] Note that in English, SRCs resemble the Subject-Verb-Object (SVO) word order, but in Mandarin, ORCs resemble the SVO word order.

Table 2.7. RCs in English and Mandarin (*de* is the Mandarin RC marker)

	English	Mandarin
SRC	**The man** [RC who __ sees the woman] Subject Verb Object	[RC __ kànjiàn nǚrén *de*] **nánrén** [RC __ see woman *de*] man Verb Object Subject
ORC	**The man** [RC who the woman sees __] Object Subject Verb	[RC nǚrén kànjiàn __ *de*] **nánrén** [RC woman see __ *de*] man Subject Verb Object

Typologically rare in having head-final RCs and a canonical SVO word order (e.g., Dryer, 2005; Comrie, 2008), Mandarin RCs receive much attention in psycholinguistics and acquisition research to help tease apart different theories. I briefly discuss the psycholinguistics studies before focusing on the acquisition studies in section 2.3.2.

In the psycholinguistics and acquisition research on RCs, the subject-object asymmetry, often called the SRC advantage or SRC preference, means that SRCs are acquired earlier, read or processed faster, and with higher accuracy. In recent psycholinguistics literature on this topic, the two main competing theories are memory-based theories and expectation-based theories. If a structure is more frequent than another structure, the expectation-based theories predict that the more frequent structure will be easier to acquire and process (e.g., Reali & Christiansen, 2007).[20]

[19] The grammatical status of the word *de* is controversial. Several terms have been used in the Mandarin RC literature, including genitive, relativizer, complementizer, and simply, *de* (Sun, 2015). In this dissertation, *de* is simply considered as a RC marker and glossed as such. The underscore denotes a gap, but whether the gap exists is not critical here. For clarity, RCs are bracketed.

[20] More recent proposals under the expectation-based theories are surprisal (Hale, 2001; Levy, 2008) and entropy reduction (Hale, 2003, 2006), both informed by information theory.

If the linear distance (or intervening elements) between the head noun and the gap (or the embedded verb) is shorter in one structure than another, the memory-based theories (e.g., Gibson, 1998, 2000) predict that the former is easier to acquire and process because it requires fewer processing resources to be kept in memory. In head-initial RCs like English, expectation-based theories predict an SRC advantage because SRCs are more frequent than ORCs. Similarly, memory-based theories also predict an SRC advantage in English because the linear distance between the head noun and the gap (or the embedded verb) in SRCs is shorter than that in ORCs. Given that both theories predict an SRC advantage in English, it is difficult to evaluate which theory has more explanatory power. In head-final RCs such as Mandarin, expectation-based theories predict an SRC advantage while memory-based theories predict an ORC advantage (note the linear distance between the head noun and the gap is shorter in ORCs than in SRCs); see Table 2.8. Thus, head-final RCs make an ideal testing case for RC processing.

Table 2.8. Theories and predictions on the SRC/ORC asymmetry in English and Mandarin

	expectation-based theories (e.g., Reali & Christiansen, 2007; Hale 2001; Levy 2008)	memory-based theories (e.g., Gibson 1998; 2000)
English	an SRC advantage	an SRC advantage
Mandarin	an SRC advantage	an ORC advantage

In Mandarin, inconsistent results have been reported with adult NSs. While both an SRC advantage (e.g., C. Lin & Bever, 2011; Vasishth, Z. Chen, Q. Li, & G. Guo, 2013; Jäger et al., 2015; Y.-T. Sung, Cha, Tu, M.-D. Wu, & W.-C. Lin, 2016a) and an ORC advantage (e.g., Hsiao & Gibson, 2003; Y. Lin & Garnsey, 2011; Packard et al., 2011; Gibson & H.-H. I. Wu, 2013; Xu, Duann, Hung, & D. H. Wu, 2019) have been reported, Jäger et al. (2015) argues that there is a universal SRC advantage and the apparent ORC advantage reported in Mandarin was due to

temporary/local ambiguities as a confounding factor.[21] After removing temporary ambiguities by adding classifiers, adverbials, and frequency phrases in their stimuli, Jäger et al. found an SRC advantage. Corroborating Jäger et al.'s finding, Mansbridge, Tamaoka, Xiong, & Verdonschot (2017) found an ORC advantage with temporarily ambiguous RCs, but an SRC advantage with unambiguous RCs. The points of difficulty that occurred in their eye-tracking data were predicted by both the expectation-based and the working-memory based theories. Although the present dissertation adopts offline tasks, the observed processing asymmetry might still affect the offline performance of the participants, with better performance on SRCs over ORCs due to the expectation-based theories.

Another note on frequency is that while RCs are common, RCs with two animate referents/nouns are infrequent cross-linguistically (see F. Wu et al., 2012 for a summary, citing English and German corpus studies) though many experimental studies focusing on SRC/ORC asymmetry have used such structures to control for animacy effects. According to Hsiao (2003, p. 105, cited in F. Wu et al., 2012), out of 882 Mandarin RCs in a corpus, only six had two animate nouns. K. Kuo and Vasishth (2006, cited in F. Wu et al., 2012) found that out of 164 RCs in another corpus, only 13 (out of 119) SRCs and 3 (out of 45) ORCs had two animate nouns. In two corpus studies that examine RC types and head nouns, more than 85% of the ORCs had inanimate head nouns (extracted objects) while more than 65% of SRCs had animate head nouns (extracted subjects) (Pu, 2007; F. Wu, 2009, both cited in F. Wu et al. 2012). Thus, HSs and L2ers may have difficulty with RCs that have two animate nouns because these are rare in the input.

[21] Temporary ambiguities arise because of the surface similarities between ORCs (e.g., [woman see *de*] man) and regular SVO sentences (e.g., woman see man). Upon processing the first two words in ORCs, i.e., 'woman' and 'see', readers may misanalyse them as the subject and the verb (in an SVO sentence) until the disambiguating *de*. In head-initial RCs, such temporary ambiguities do not occur.

2.3.2. Acquisition of Mandarin relative clauses

Previous studies have shown that Mandarin-speaking children acquire Mandarin RCs starting at age three or four and the acquisition stabilizes at age five (C. Hsu, 2014). In L2-Mandarin, Hu and C. Liu (2007) found that L1-English L2ers outperformed L1-Korean L2ers in judging the well-formedness of Mandarin RCs and suggested the surface similarity between Korean and Mandarin RCs (both head-final) hinders the acquisition of Mandarin RCs by L1-Korean L2ers.

Cross-linguistically, SRCs have often been found to be easier than ORCs in acquisition. This SRC advantage can be explained by the Noun Phrase Accessibility Hierarchy (NPAH), originally proposed by Keenan and Comrie (1977) as a typological universal or generalization, and might reflect "psychological ease of comprehension" (p. 88). Simply put, the NPAH predicts that a subject is easier to relativize cross-linguistically, hence the SRC advantage. Studies on head-initial RCs often find an SRC advantage, supporting the NPAH (e.g., L1-English: C. Kim & O'Grady, 2016; L2-English: Doughty, 1991; Gass, 1980). However, the results from head-final RCs are mixed, especially in Mandarin. In L1 acquisition, the SRC advantage (e.g., C. Hsu, Hermon & Zukowski, 2009; C. Hsu, 2014; Hu, Gavarró, Vernice, & Guasti, 2016; Hu, Gavarró, & Guasti, 2016) seems to have received more support than the ORC advantage (e.g., J. Chen & Shirai, 2015; W. He, N. Xu, & Ji, 2017).

These mixed findings are reflected in L2-Mandarin studies as well; both the SRC advantage (e.g., Y. Xu, 2013, 2014a; Cherici, Y. Chang, & Tanaka, 2019) and the ORC advantage (e.g., Yaqiong Wang & Feng, 2014; Cui, 2013; L. Chang, 2017) have been reported; see Table 2.9 for a summary. Testing L1-English L2ers, Y. Xu (2013; 2014a) and Cherici et al. (2019) found an SRC advantage, while Yaqiong Wang and Feng (2014) found an ORC advantage. Cui (2013) found an

ORC advantage from L2 learners with various L1 backgrounds using a multiple-choice questionnaire (asking "who helped whom?"). Comparing two L1 groups, L. Chang (2017) found both L1-English and L1-Japanese L2-Mandarin learners produced more ORCs than SRCs, but the trend was reversed for L1-English L2ers with higher proficiency. In an offline forced-choice task, C. Chen (2017) found that L1-English L2ers marginally outperformed L1-Korean L2ers in subject-modifying ORCs. However, it is likely that some L1-English L2ers were able to answer Mandarin ORCs (i.e., [Subject-Verb *de*] Object) correctly by utilizing SVO order, without fully acquiring Mandarin RCs. This SVO strategy might also explain Cui's (2013) findings. The picture-based TVJT in C. Chen (2019c), and in this dissertation, did not allow participants to answer all conditions correctly by simply utilizing an SVO word order cue. After removing the SVO cues, C. Chen (2019c) found that L1-English L2ers showed an SRC advantage and outperformed L1-Korean L2ers in SRCs, but not ORCs.

Table 2.9. Selected studies on Mandarin RCs by L2ers, focusing on subject-object asymmetry

	L1s	Test format	SRC or ORC advantage
Xu (2013)	English	A written sentence completion task	SRC
Yaqiong Wang & Feng (2014)	English	A listening-oral translation task (both L2-to-L1 and L1-to-L2 translation)	ORC
Xu (2014a)	English	A written sentence combination task	SRC
Cui (2013)	Various	A multiple-choice questionnaire (asking "who helped whom?")	ORC advantage, but data from self-paced reading are mixed
L. Chang (2017)	English & Japanese	Naturally produced compositions (learner corpus)	ORC (but not throughout all levels)
C. Chen (2017)	English & Korean	A forced-choice task (asking, e.g., who got invited?)	No advantage
Cherici, Y. Chang, & Tanaka (2019)	English	An elicited production task	SRC
C. Chen (2019c)	English & Korean	A picture-based TVJT	SRC advantage from the English group

Like L1 processing, L2 processing of Mandarin RCs yields divergent results (e.g., SRC advantage: Xu, 2014c, Q. Li, X. Guo, Yiru Yao, & Müller, 2016; Q. Yao & Renaud, 2016; ORC advantage: Packard, 2008; Y.-T. Sung, Tu, Cha & M.-D. Wu, 2016b; Yun Yao, 2018). Given that online studies have temporary ambiguity/garden-path problems that are not addressed in offline studies, the L2 processing studies are not reviewed.

In heritage language acquisition, to my knowledge, there is no study on Mandarin RCs by English-dominant adult HSs. There are, however, a few studies on Russian (both child and adult HSs: Polinsky, 2011) and Korean RCs (adult HSs: Lee-Ellis 2011; T. Lee, 2016; O'Grady et al., 2001). I briefly discuss Korean RCs since they are head-final. Both head errors (assuming head-final as head-initial) and reversal errors (assuming ORCs as SRCs, and vice versa) are commonly reported in the acquisition of Korean RCs (Lee-Ellis 2011; T. Lee, 2016; O'Grady et al., 2001. With reversal errors, participants presumably know that Korean RCs are head-final, but are unable to use case markers that are necessary to distinguish between SRCs vs. ORCs in Korean.[22] While head errors are sometimes discussed in Mandarin RCs, reversal errors are typically not.

While there are no studies on Mandarin RCs by adult HSs, there are a few studies with Mandarin-English or Cantonese-English bilingual or even trilingual children. While Chan et al. (2017) and R. Jia and Paradis (2020) consider these children to be HSs, Kidd et al., (2015) and

[22] In Korean (and Japanese), both SRCs and ORCs have NV sequence in the RC region but differ in the case marking; see examples below (O'Grady et al., 2003; NOM, nominative case; ACC, accusative case; RES, present tense; PAST, past tense).

(i) Korean SRC

[$_{RC}$	_	namca-lul	po	nun]yeca	
[$_{RC}$	_	man-ACC	see	RC.PRS]	woman

'The woman [$_{RC}$ who sees the man]'

(ii) Korean ORC

[$_{RC}$	namca-ka	_	po	nun]yeca	
[$_{RC}$	man-NOM	_	see	RC.PRS]	woman

'The woman [$_{RC}$ who the man sees]'

Tsoi et al. (2019) do not. With regard to RC asymmetry, Tsoi et al. (2019) found an SRC advantage in Mandarin while Yip and Matthews (2007) found an ORC advantage in Cantonese.

Most studies with bilingual children have found cross-linguistic transfer of RC head direction: English-to-Cantonese transfer in English-dominant simultaneous bilingual children in Australia (Kidd et al., 2015), English-to-Mandarin transfer in English-dominant children in Canada (R. Jia & Paradis, 2020, see also Jia, 2016) and English-to-Mandarin transfer in English-dominant children in Australia (Tsoi et al., 2019).[23]

R. Jia and Paradis (2020) found that monolingual Mandarin-speaking children outperformed heritage bilingual children in Canada: some (10 out of 29 children, but only 8% out of all utterances) bilingual children incorrectly produce head-initial Mandarin RCs, such as head-RC-*de*, showing dominant language transfer from English. However, these children were comparable with monolingual peers in comprehension using an audio picture-selection task. Interestingly, when asked to comprehend ungrammatical head-initial Mandarin RCs, both bilingual children and monolingual peers have an easier time comprehending ungrammatical head-initial Mandarin SRCs than ungrammatical head-initial Mandarin ORCs because head-initial SRCs (i.e., head $_{RC}$[verb-object-*de*] have SVO word order and thus children could still comprehend "who did what to whom" (p. 171). In comprehension, English-dominant English-Mandarin bilingual children misidentified the RC subject as the head in ORCs (Mandarin: Tsoi et al., 2019; Cantonese: Kidd et al., 2015). (Note that in Mandarin, only ORCs begin with nouns, so head errors

[23] Three related studies on cross-linguistic transfer on RCs were conducted in Hong Kong and Singapore. Yip and Matthews (2007) found Cantonese-to-English transfer in simultaneous Cantonese-dominant bilingual children using diary studies (e.g., children produced head-initial RC in Cantonese). Chan et al. (2017) found English-to-Cantonese transfer in trilingual children in comprehension (L1-Cantonese L2-English L3-Mandarin; children are Cantonese-dominant, and Mandarin is their weakest language). In a production task, Yan and Matthews (2017) found English-to-Mandarin transfer with bilingual children in Singapore (language dominance unclear, presumed to be English-dominant by the authors). Due to the sociolinguistic contexts in Hong Kong and Singapore, bilingual speakers there, even if they are English-dominant, are not necessarily considered as HSs of Mandarin or Cantonese. Chan et al. (2017), however, did consider the children in their study to be HSs of Cantonese.

are more likely with ORCs than with verb-initial SRCs.) These head errors were also found in O'Grady et al. (2001), where adult Korean HSs performed similarly to L2ers in misanalysing the first noun (in both SRCs and ORCs) as the head due to English transfer.

This dissertation contributes to the debate concerning SRC/ORC advantage with new data from both HS and L2 Mandarin. Among the phenomena tested in this dissertation, RCs represent syntax. While RCs, like grammatical aspect, fall into the general domain of (morpho-)syntax, an HS advantage is less likely here because (a) RCs are a later-acquired phenomenon; and (b) L2ers are generally found to be more successful with purely syntactic phenomena than with morphology (Slabakova, 2008, 2014).

2.4. Mandarin anaphors, with a focus on long-distance reflexives

2.4.1. Properties of Mandarin anaphors, with a focus on long-distance reflexives

According to Binding Principle A (Chomsky, 1981), a reflexive must be bound in its binding domain, roughly a clause. In (15), an English reflexive must refer to a local antecedent, and not a long-distance (LD) antecedent. Unlike English, Mandarin has two types of reflexives, the complex reflexive *taziji* (himself/herself) and the simplex reflexive *ziji* (self). Like the English 'himself/herself', *taziji* requires a local antecedent, as in (16), which is the Mandarin equivalent of (15) . In contrast, *ziji* (self) can take either a LD or a local antecedent, as in (17). However, blocking effects apply when a local antecedent is a first or second person singular pronoun and "block" *ziji* from taking a third-person LD antecedent, as in (18).

(15) John$_i$ thinks Peter$_j$ trusts himself$_{*i/j}$. (complex reflexive *himself*)

(16) Zhangsan$_i$ renwei Lisi$_j$ xiangxin taziji$_{*i/j}$ (complex reflexive *taziji*)
Zhangsan think Lisi trust himself
'Zhangsan thinks that Lisi trusts himself.'

(17) Zhangsan$_i$ renwei Lisi$_j$ xiangxin ziji$_{i/j}$ (simplex reflexive *ziji*)
Zhangsan think Lisi trust himself/him
'Zhangsan thinks that Lisi trusts himself/him.'

(18) Zhangsan$_i$ renwei wo$_j$ xiangxin ziji$_{*i/j}$. (*ziji* under blocking effects)
Zhangsan think I trust self
'Zhangsan thinks that I drew myself.'

Ziji has been researched extensively in theoretical syntax (e.g., Cole, Hermon, & L. Sung, 1990; C.-T. J. Huang & J. Tang, 1991; Cole, Hermon, & C.-T. J. Huang, 2006), though purely syntactic analyses have not been satisfactory. Building on prior work, C.-T. J. Huang and C.-S. L. Liu's (2001) non-uniform dual approach to *ziji* settles the issue by proposing two kinds of *ziji*: a syntactic anaphor (which has the local reading) and a pragmatic logophor (which typically has the LD reading). The anaphor is subject to syntactic constraints while the logophor is subject to pragmatic constraints at the syntactic-discourse interface. This non-uniform dual approach draws a line between a syntactic anaphor and a pragmatic logophor based on the size of the binding domain, in contrast with Reinhart and Reuland's (1993) approach, which draws a line based on co-argumenthood. Using a TVJT, Su (2017a) found that adult NSs of Taiwanese Mandarin preferred LD readings of "*ziji-de* NP" over just "*ziji*" as the object of the verbs, which supports Reinhart and Reuland's analysis. Following Reinhart and Reuland, only *ziji* in *ziji-de* NP, but not the direct object *ziji*, is a logophor since the antecedents are not the co-argument of the predicate. However, Su (2017a) only had two tokens per condition (possibly due to working with children), and Zeng (2010) did not find any differences in *ziji* vs. *ziji-de*. Thus, more research is needed to draw a firm conclusion on whether *ziji* as a direct object is a logophor or not.

Like RCs, LD reflexives are much studied in psycholinguistics to tease apart different theories (see e.g., Jäger, Engelmann, & Vasishth, 2015). More relevant to the current dissertation is that processing local reading of reflexives is easier for NSs than processing LD readings (see

Dillon, 2014, for an overview of reflexive processing). This finding has been supported by several psycholinguistic techniques, including ERPs (Xiaoqian Li & Zhou, 2010), eye-tracking (Jäger et al., 2015), self-paced reading (Z. Chen, Jäger, & Vasishth, 2012; Dillon, Chow, & Xiang, 2016; X. He & Kaiser, 2016), and the multiple-response speed-accuracy tradeoff (MR-SAT) paradigm (Dillon et al., 2014), though see Lu (2011) for an opposite finding. The difficulty in processing may affect the performance of speakers even in offline tasks, thus HSs and L2ers in my study may choose the easier way to process the reflexives, i.e., choose local readings of reflexives.[24]

In terms of frequency of LD vs. local readings of *ziji*, the only two corpus studies I am aware of are contradictory. Extracting 852 examples of *ziji* from two Chinese novels (written before 1950), L. Liu (2010) found that 10% of *ziji* involved local readings, 58% involved LD readings within the same sentence (different clauses), and 32% involved cross-sentential LD readings. He also found that when local readings are intended, *ziji* occurs more often than *taziji*; [25] when LD readings are intended within the same sentence, *ziji* occurs more often than pronouns.

In contrast, Lu (submitted) found that local readings of *ziji* are much more frequent than LD readings. Out of the randomly chosen 1000 sentences containing *ziji* from a Taiwan-based corpus (1981-2007), almost 80% of *ziji* involved local reading, 10% involved LD reading, and the remaining 10% had generic use or zero anaphora. However, as Lu did not differentiate bi-clausal sentences from mono-clausal sentences (which almost always have a local reading, unless the antecedent is outside of the sentences), the percentage of LD readings should exceed 10% in bi-

[24] Hawkins's "Minimize domains" proposal in computing syntactic dependencies (1994, 2004) seems compatible. It is used to explain the difficulty HSs have in acquiring/maintaining LD reading of reflexives in Turkish (Gračanin-Yuksek et al., 2020) and Icelandic (Putnam & Arnbjörnsdóttir, 2015).

[25] Jiang (2009, p. 484, footnote 5) also reports a corpus count and notes that *ziji* is much more frequent than *taziji*, but does not indicate whether the search separates the anaphor vs. intensifier usage of *ziji*. As an intensifier (Hole, 2008) or an adverbial (Tsai, 2019), *ziji* can appear after a proper name or a pronoun in the subject position, such as *Zhangsan ziji* and *ta ziji*. *Taziji* can also appear after a proper name as an intensifier (Tsai, 2019).

clausal sentences.[26] If all occurrences of *ziji* are considered, I argue that local readings are likely more frequent than LD readings, since mono-clausal sentences are presumably more frequent than bi-clausal sentences in most contexts, especially in oral speech. More studies are needed to support this claim.

2.4.2. Acquisition of Mandarin anaphors, with a focus on long-distance reflexives

In L1 acquisition, children under age eight predominantly choose local readings for *ziji* even in contexts that favor LD readings (Chien & Lust, 2006; Chien, Wexler, & H. Chang, 1993; Su, 2004, 2017a, 2017b). Chien and Lust (2006) explain such local preferences in two ways. The set-inclusion parametric approach (Wexler & Manzini, 1987) argues that there are five values to the Governing Category (or binding domain) parameter. Mandarin has an expanded binding domain outside of a clause while English has a binding domain that is a clause. Applying the Subset Principle to binding, children initially allow only local readings as the unmarked default option. After receiving positive evidence that LD readings are allowed in a given language, children switch to a less restrictive parameter setting.

The other explanation applies the non-uniform approach in syntax (C.-T. J. Huang & C.-S. L. Liu, 2001) to L1 acquisition. Chien and Lust (2006) hypothesize that children initially acquire syntactic anaphors (locally bound) and only acquire logophoric anaphors after gaining pragmatic knowledge. While studies on locality cannot differentiate between these two approaches, findings on another property of *ziji* (i.e., subject-orientation) support the non-uniform approach rather than the parametric

[26] Differences between the corpus studies might result from different criterions (e.g., whether mono-clausal sentences are considered) and/or different corpora (different dialects, genres, time periods, etc.).

approach.[27] In addition, given that the binding domain of Mandarin *taziji* is a clause, the parametric approach to binding domains cannot be set for each language. This is a serious critique, and the parametric approach to binding domains is no longer accepted in syntax and language acquisition literature.

In L2 acquisition, most studies have found that L2ers have difficulties acquiring LD reflexives (D. Chen, 1995; Christie & Lantolf, 1998; Dugarova, 2007; Sperlich 2013, 2016, 2017; Yuan, 1993, 1994, 1998; Zeng, 2010); see Table 2.10 for a selective summary. If available, test results on *taziji* are included to examine if transfer effects apply to both *ziji* and *taziji*. If the studies examine more than one context type for *ziji*, only results from experimental conditions where *ziji* is a direct object (or inside the direct objects) in embedded finite clauses in neutral contexts are included. A note is made if a developmental pattern is observed across proficiency groups.

Table 2.10. Selected studies on Mandarin reflexives by L2ers and HSs

Study	L1s or dominant languages	Test format	Pattern summary or key finding
Yuan (1994)	English	Acceptability judgement tasks	L2ers differ from NSs in LD readings; a mild U-shape on the acceptance of LD readings among the five L2 groups
Yuan (1998)	English & Japanese	A multiple-choice comprehension test	L1-Japanese L2ers patterned with NSs and outperformed (proficiency-matched and even more proficient) L1-English L2ers (the Intermediate group accepted more LD readings than the Advanced group)
Ying (1999)	English	A sentence interpretation task	L2ers predominantly only allowed local readings.

[27] While an English reflexive can take either a subject or an object as its antecedent, as in (i), Mandarin reflexive *ziji* is subject-oriented, as in (ii) (DE is a possessive marker).

(i) John$_i$ gave Bill$_j$ a photograph of himself$_{i/j}$.

(ii) Zhangsan$_i$ gei-le Lisi$_j$ yi-zhang ziji $_{i/*j}$ de zhaopian
Zhangsan give-perf Lisi one-CL self DE photo
'Zhangsan gave Lisi a photograph of him.'

Table 2.10. Selected studies on Mandarin reflexives by L2ers and HSs (cont'd)

Study	L1s or dominant languages	Test format	Pattern summary or key finding
Dugarova (2007)	Russian & English	Same as Yuan (1998)	L1-English L2ers (a mild U-shape among the three proficiency groups) outperformed L1-Russian L2ers at every proficiency level
Kong (2009)	English	A multiple-choice comprehension task	*Ziji*: Unlike NSs, L2ers accepted more LD readings than local readings. The Elementary group accepted fewer local readings than the Intermediate group. *Taziji*: L2ers accepted more local readings than LD readings. The Elementary group accepted fewer LD readings than the Intermediate group.
Zeng (2010)	English	A grammaticality judgment task and a context-based sentence judgment task	*Ziji*: L2ers differed from NSs in LD readings; a mild U-shape on the acceptance of LD readings among the three proficiency groups *Taziji*: The less proficient groups accepted LD readings while the most proficient group patterned with Mandarin NSs
Sperlich (2013)	English & Korean	An interpretive judgement test and a TVJT	*Ziji*: L1-Korean L2ers patterned with NSs and outperformed L1-English L2ers (though not on every measure).
C. Chen (2019a; C. Chen & Ionin, in preparation)	English & Korean	A picture-based TVJT	*Ziji*: L2ers predominantly only allowed local readings. *Taziji*: L1-English L2ers under-accepted local readings (and numerically over-accepted LD readings).
C. Chen (2019b)	English (of HSs and L2 learners in the United States)	A picture-based TVJT	*Ziji*: both HSs and L2ers predominantly allowed only local readings. *Taziji*: HSs are native-like, while L2ers over-accepted LD readings and under-accepted local readings.

A few studies have tested two L2-Mandarin groups with different L1s to examine L1 transfer, but the results are not very consistent. Korean speakers (Sperlich, 2013, see also 2016, 2017) and Japanese speakers (Yuan, 1998) were found to outperform English speakers in acquiring Mandarin

ziji due to similar forms in Korean (*caki*) and Japanese (*zibun*).[28] However, D. Chen (1995) found that neither the English nor the French group acquired LD readings of *ziji*, so there is no transfer from French *soi* that usually has LD readings. Using the same tasks as Yuan (1998), Dugarova (2007) found English speakers to outperform Russian speakers even though neither language has LD reflexives. While only testing one language group, Zeng (2010) clearly found a lack of L1 transfer since English speakers did not allow local readings of *ziji* and *taziji* consistently, as would be expected under English transfer. In C. Chen (2019a; C. Chen & Ionin, in preparation), I found that both Korean and English speakers predominantly allow only local readings of *ziji*, showing no transfer from Korean (even though Korean speakers do show transfer of local reading of pronouns, following recent findings in E. H. Kim, 2018, 2019).

In heritage language acquisition, LD reflexives have been examined in Korean (J.-H. Kim, Montrul, & Yoon, 2009, 2010; S. Y. Lee, 2012) and Turkish (Gračanin-Yuksek et al., 2020). Using an offline antecedent selection task and an online self-paced reading task, Gračanin-Yuksek et al. (2020) found that HSs of Turkish in Germany maintained the LD readings of *kendi* and *kendisi* without simplifying the binding system to match their dominant language German (the focus of their study is on simplification rather than on language transfer). Both J.-H. Kim et al. (2010) and S. Y. Lee (2012) found that HSs of Korean and L1-English L2-Korean learners allowed LD readings of simplex reflexives *caki*, though not to the same extent as NSs did in J.-H. Kim et al. (2010). Adapting J.-H. Kim et al.'s (2009) picture-based TVJT in Korean, I tested both English dominant HSs and L1-English L2 learners in C. Chen (2019b). Given that this dissertation in turn adapted the design in C. Chen (2019b), I discuss that design in detail but focus on four (out of six)

[28] Sperlich (2013, see also 2016, 2017) did not conduct his study under the generative framework. He posited that Korean, like Mandarin, is a pragmatic language while English is a syntactic language. Korean speakers have an advantage over English speakers because they were able to transfer their pragmatic strategies to L2-Mandarin. One critique of this study is that he only had two tokens per condition.

test conditions that crossed the factor 'reflexive type' (two levels: *ziji* and *taziji*) with the factor 'picture type' (two levels: local and LD readings). Results showed that both HSs and L2ers predominantly only allowed local readings of *ziji*, hence no heritage advantage. There are many explanations for local preferences, including dominant language transfer from English, local binding as the default option, which could in turn be due to ease of processing (Dillon, 2014), or to higher input frequency (Lu, submitted, though see L. Liu 2010, for an opposite finding). However, given that Korean transfer is not found in a similar TVJT (C. Chen, 2019a; C. Chen & Ionin, in preparation), English transfer cannot be the only reason why HSs and L1-English L2ers only allow local readings of *ziji*.

Another possibility is that, under the Interface Hypothesis (Sorace, 2011; Tsimpli & Sorace, 2006), LD reflexives are difficult to acquire because they are at the syntactic-discourse interface under C.-T. J. Huang and C.-S. L. Liu's non-uniform approach. However, the updated Interface Hypothesis (Sorace, 2011) argues that only structures at the external interface, rather than the internal interface, are difficult to acquire. Yet, it is not always clear how to distinguish the syntax-semantics interface from the syntax-discourse/pragmatics interface (see e.g., Slabakova, 2011). It is uncertain whether LD readings of *ziji* tested in C. Chen (2019a) lie at the internal syntax-semantics interface or the external syntax-discourse interface; see Reinhart and Reuland (1993) and Su (2017a) mentioned above.

Regarding *taziji*, while HSs in C. Chen (2019a) patterned with NSs in correctly accepting only local readings, L2ers over-accepted the LD readings and under-accepted the local readings, possibly due to indeterminacy in judgments or misanalysis of *taziji* as the pronoun *ta* 'he/she/him/her'.[29] Following Binding Principle B, pronouns cannot refer to the antecedent in its

[29] In some contexts, *taziji* can indeed be grammatically interpreted as *ta* followed by an intensifying *ziji* (or called 'emphatic reading' or 'focus reading'). However, C.-C. J. Tang (1989, p. 98) argued that without context, it is more

binding domain. Thus, in (19), *ta* is coreferential with *Zhangsan* or someone else not mentioned in the sentence.

(19) Zhangsan$_i$ renwei Lisi$_j$ hua-le ta$_{i/*j/k}$. (pronoun *ta*)
Zhangsan think Lisi draw-le him/her
'Zhangsan thinks that Lisi drew him.'

Among the phenomena tested in this dissertation, LD reflexives represent the interface between syntax and semantics instead of the syntax-discourse interface because, under the current design, *ziji* is an anaphor even with LD readings (following Reinhart & Reuland, 1993) and pragmatics is neither tested nor manipulated. By testing *ta*, *ziji*, and *taziji* within the same task, this dissertation examines whether proficiency-matched participants distinguish between different anaphors, and examines whether L2ers indeed misinterpret *taziji* as a pronoun rather than a reflexive. **LD reflexives are expected to be the least likely to yield any HS advantage, because they are a very late-acquired phenomena in L1 acquisition:** by age eight, HSs are typically in English-speaking schools and English-dominant, with reduced Mandarin input. Indeed, C. Chen (2019) did not find any HS advantage for LD readings of *ziji*, even though HSs were more proficient than L2ers in that study.

2.5. Pedagogical note

Given that all L2ers and most HSs tested in this dissertation were at some point instructed learners, a note on the pedagogy of these phenomena is in order. Tones are usually introduced in the first class, along with Romanization and tone marks. According to a recent survey (C. Yang &

difficult to get the intensifying reading in object position. Previous studies (e.g., Chien et al., 1993) have shown that adult Mandarin NSs predominantly only allow local readings when *taziji* is in the object position or as a possessive inside the object position such as *taziji-de* NP (*de* is the possessive marker).

W. Jin, 2018), 86% of instructors taught the T3 sandhi rule, while only 53% of them taught the half-T3 rule. However, while these rules are usually taught, tone errors are not necessarily corrected in class. For aspect markers, based on Y. Shi's (2013) summary, *-le* is taught early in L2 classrooms, usually during the first year. Instructors often emphasize that *-le* is not English past tense, but errors remain even in highly proficient learners. The difficulty or confusion experienced by L2ers is perhaps heightened by the fact that *-le* has the same form with a sentence-final particle *LE*. Other aspect markers are not focused on as much. For *-le*, see C. Zhu (2019) for a guide for English-speaking learners and Y. Xu (2020) for consciousness-raising among L2ers in the classroom. See also Q. Zhang (2016) for comparing a Grammar-Translation approach and a communicative approach in the teaching of the four Mandarin aspect markers. For RCs, at least in some institutions, after teaching learners the pre-nominal adjectival modifiers, instructors then teach learners that clausal modifiers (i.e., RCs) precede Mandarin head nouns as well; ORCs are often taught earlier or more heavily emphasized than SRCs. For *ziji*, students are simply taught the basic translation 'self'; LD properties and blocking effects are not taught. In fact, many Mandarin instructors do not know LD properties and blocking effects explicitly.

2.6. Research questions and hypotheses

The current dissertation asks two broad research questions (RQs) and several specific RQs on each linguistic phenomenon. The two broad RQs are the following:

(20) Broad RQ 1: Can HSs and L2ers of Mandarin whose dominant language is English fully acquire the properties of Mandarin that are different from or absent in English?

(21) Broad RQ 2: Do HSs have selective advantages over proficiency-matched L2 learners, and does this vary by linguistic domain?

To answer these questions, I look at four different linguistic phenomena by using three offline tasks: a Tone Identification Task, an Acceptability Judgement Task (AJT), and a Truth Value Judgement Task (TVJT) with pictures; see Table 2.11.

Table 2.11. Summary of the four target phenomena in Mandarin

Phenomena	Linguistic domain	Age of acquisition in monolingual children	Tested in …
Tone 3 sandhi	Phonology	2-3	Tone Identification Task
Aspect	Morpho-semantics interface	most at 3, some by 5	AJT
RCs	Syntax	3-4, stabilize at 5	AJT and TVJT
LD reflexives	Syntax-semantics interface	At least after 8	TVJT

In answer to the first question, learners' (HSs' and L2ers') ability to acquire these different phenomena are likely to be influenced by multiple considerations, including **transfer from English**, **domain vulnerability**, **frequency**, and **processing considerations**. Given that the tasks employed in this dissertation are offline, I will leave **processing** aside. I also cannot address **frequency** directly given that it is difficult to quantify the input that L2ers vs. HSs have received. Thus, I focus on **vulnerable domains** and **dominant language transfer from English** in this dissertation, though frequency and processing will be briefly mentioned .

Recall that the four phenomena examined in this dissertation are either absent or differentially instantiated in English. Yet not having counterparts in English does not mean that all phenomena will be equally difficult. Considering the **vulnerable domains**, the Bottleneck Hypothesis (L2ers: Slabakova, 2008, 2014; HSs: Montrul, 2018) predicts special difficulty with aspectual marking, if it is considered to be morphology-related (but not when it is considered to lie at the syntax-semantics interface, following S. Zhang, 2018). The updated Interface Hypothesis (Sorace, 2011;

HSs: Montrul & Polinsky, 2011) predicts special difficulty with LD reflexives, if they are considered at the syntax-discourse interface, but not if they are at the syntax-semantics interface. Processing considerations predict special difficulty with ORCs (relative to SRCs) (e.g., Jäger et al., 2015, but see e.g., Gibson & H.-H. I. Wu, 2013) and LD reflexives (relative to locally bound reflexives) (e.g., Jäger et al., 2015; Dillon et al., 2016). ORCs are difficult to process due to the long distance between the head noun and the gap/verb; LD reflexives are difficult to process due to long distance between the antecedent and the anaphor. Potential processing difficulty may also occur with T3 sandhi and non-prototypical combinations of lexical and grammatical aspect. C. Zhang et al., (2015) found processing difficulty on T3 sandhi (relative to the non-sandhied counterpart). Yap et al. (2004, cited in Yap et al. 2009) found that processing *-le* with accomplishments and *zai* with activities is easier. For frequency considerations, LD reflexives might be especially difficult to acquire, given their low frequency.

For **dominant language transfer from English**, given that T3 sandhi is absent in English, there is nothing to transfer. For the other three phenomena (aspectual marking, RCs, and LD reflexives), I hypothesize that both groups would exhibit **transfer from English**, which should result in difficulty with LD reflexives, and errors with grammatical aspect markers and RC interpretations in those contexts where English and Mandarin do not match. (Additionally, English transfer might be more pronounced in L2ers than HSs, especially for those features that are early-acquired in monolingual children, e.g., some pairing of grammatical and lexical aspect.)

Since only one language combination (Mandarin/English) is being tested, it cannot be definitively concluded that any of the errors are due to cross-linguistic influence rather than to some other source, such as some form of grammar simplification, due to **reduced input** and/or **low frequency** (e.g., LD reflexives are relatively infrequent); at the same time, transfer predicts

specific patterns of errors with aspect markers and RCs, which are not likely to stem from another source.

To summarize the hypothesis for RQ1 (complete acquisition), T3 sandhi is likely to be fully acquired by HSs, but not by L2ers. Aspectual marking is likely not fully acquired by either group, due to English transfer and/or morphology-related difficulty. RCs are likely to be fully acquired by both groups since they involve only syntax and transfer of word order is easy to overcome. LD reflexives are likely to be the most difficult for both groups due to English transfer and/or potential difficulty with the external interface.

In answer to the second question, **HSs would have an advantage over L2ers (i.e., would be more target-like) for those phenomena that are early acquired in L1 acquisition**.[30] Based on the Ages of acquisition (AoAs) in Table 2.7, HSs are most likely to have an advantage over L2ers on tone sandhi, followed by aspect markers. While RCs are acquired not much later than (some) aspect markers, I hypothesize that L2ers would have more difficulty with aspect markers (morphology) than with RCs (syntax), per the Bottleneck Hypothesis (L2ers: Slabakova 2008, 2014; HSs: Montrul, 2018). Finally, LD reflexives, which are particularly late-acquired, are expected to confer no HS advantage. Both HSs and L2ers are expected to have problems with LD reflexives, given the multiple factors that conspire to make them difficult to acquire (local reading of reflexives is cross-linguistically the default and is the most frequent, while LD reading is not possible in English, is likely less frequent even in Mandarin, and may involve discourse-based

[30] The possibility that the HSs in our study acquired the relevant phenomena as children but later lost them due to attrition cannot be completely excluded. Without direct child HS/adult HS comparisons (as in Polinsky, 2011; Montrul & Sánchez-Walker, 2013; see Polinsky, 2018a, for a review) or longitudinal research (Silva-Corvalán, 2018), it is not possible to tease apart attrition from incomplete acquisition. While beyond the scope of this dissertation, this issue can be addressed indirectly by collecting detailed background information concerning participants' use of Mandarin, on the hypothesis that attrition is less likely among HSs who use the language regularly.

knowledge, which has been claimed to be problematic for all types of bilingual populations – cf. Sorace, 2011).

While the selective advantages of HSs are hypothesized to be due to early AoA, it is possible that such advantages could be due to HSs getting more input than L2ers. Without quantifying the input HSs and L2ers have, this question is difficult to address. Even if the heritage advantage is only about input, the existence of a selective HS advantage would mean that input has a differential effect across domains.

To summarize the hypothesis for RQ2 (selective advantage), T3 sandhi is the most likely to reveal a HS advantage, followed by aspectual marking. RCs are hypothesized to be fully acquired by both groups, thus revealing no HS advantage. I also hypothesize no HS advantage with LD reflexives, as neither group is predicted to acquire LD reflexives.

Ultimately, the primary focus of this dissertation is on HS/L2 comparisons, examining (a) whether complete acquisition is possible for either group, and (b) whether HSs have selective advantages over L2ers. Whatever the source of error (e.g., transfer vs. reduced input), if HSs outperform proficiency-matched L2ers, this points to the importance of early AoA in that particular domain. Furthermore, while this dissertation cannot measure exactly how much input HSs and L2ers have received, as long as both groups report using primarily English rather than Mandarin in their daily lives, they will have received reduced input in Mandarin relative to NSs.

Given that more proficient learners in both groups are expected to be more target-like than lower-proficiency learners, the HS/L2 comparison is only meaningful when Mandarin proficiency is matched; otherwise, proficiency alone might be the significant predictor in accounting for the HS/L2 differences. Therefore, in this dissertation, the two groups will be matched for Mandarin proficiency.

Chapter 3. Identification of tone sandhi in HS/L2 Mandarin

This chapter reports on Mandarin tone 3 (T3) sandhi that was tested in the Tone Identification Task. Recall that the broad RQs in this dissertation are concerned with (in)complete language acquisition and HS advantages, repeated below. RQs 1a and 2a, also listed below, are specific to T3 sandhi and are instantiations of the broad RQs 1 and 2.

- Broad RQ 1: Can HSs and L2ers of Mandarin whose dominant language is English fully acquire the properties of Mandarin that are different from or absent in English?
- Broad RQ 2: Do HSs have selective advantages over proficiency-matched L2ers, and does this vary by linguistic domain?
- RQ 1a: Can HSs and L2ers acquire the T3 sandhi rule in Mandarin, despite lack of tones or tone sandhi in English?
- RQ 2a: Do HSs have an advantage over L2ers in T3 sandhi in the domain of phonology?

Given that Mandarin T3 sandhi is acquired before age three in monolingual children, HSs are expected to have acquired it because most HSs are still regularly exposed to Mandarin at home by that age. In college-level Chinese classes, the full T3 sandhi rule, but not the half-T3 rule, is often explicitly taught (see C. Yang & W. Jin, 2018, for a survey on teachers' practice on teaching Mandarin pronunciation in the United States). Even though L2ers are taught the full T3 sandhi rule, they may still have difficulty acquiring it since phonology is known to be difficult for adult L2ers. Thus, a HS advantage is expected for T3 sandhi. In the Tone Identification Task employed here, I specifically tested whether HSs and L2ers understand that the underlying T3T3 disyllabic sequence should be pronounced as T2T3.

3.1. Procedure

All tasks were administered on Qualtrics, a web-based survey tool. Besides a few participants who chose not to finish all tasks, participants in this dissertation completed, in this order, the background questionnaire, the Mandarin proficiency test, the Tone Production Task (which is not

included in this dissertation), the Tone Identification Task, the Truth Value Judgement Task (TVJT), the Acceptability Judgment Task (AJT), and an additional Mandarin proficiency test (which only some participants took and is not included in this dissertation). The participants' linguistic backgrounds were collected using the questionnaire. The proficiency test was administered to independently measure participants' Mandarin proficiency so that the proficiency of heritage and L2 groups could be matched. However, the test required participants to read Chinese characters, so one L2er did not complete it and was subsequently excluded. All tasks were untimed. On average, NSs spent about 12 minutes on the Tone Identification Task, while HSs and L2ers spent about 15 minutes on the Tone Identification Task.

Compared to the two judgment tasks (the AJT and the TVJT) discussed in later chapters, a smaller subset of the participants is reported in this chapter. After the first version of the Tone Production and Tone Identification Tasks were administered to 44 participants (20 NSs, 11 HSs, and 13 L2ers), two problems were discovered that made the tasks unusable. Due to the undesired task effects, the format of the first version was modified (more details below after the corrected procedure is introduced) and administered to a different group of 95 participants; I report only the results of the second, corrected version of the Tone Identification Task here. To preview the results, this modification helped remove the undesired task effect to some degree but not completely, which I will address in the discussion section.

The production task preceded the perception task to avoid biasing participants' production and raising their awareness of the T3 sandhi phenomena. Some participants realized that the tasks were about the T3 sandhi phenomenon while doing the production task or later while doing the perception task. Those who recognized the purpose of the tasks might wonder whether they should choose T2T3 or T3T3 for the T3T3 condition, depending on how they interpreted the prompt.

Whether they recognized the purpose of the tasks was not systematically documented, so no further analyses were performed to see the relationship between such a realization and their performance.

Within each trial, participants listened to two sound files, each with one monosyllable. The two sound files were placed on one page, with the first monosyllable above the second. After participants listened to the two sound files of monosyllables, they clicked "Next" to move to the next page with four sound files of disyllabic sequences, made up of the two monosyllables but with different tonal combinations. The four sound files were placed on the webpage horizontally (in random order except for the critical condition, discussed below), from left to right. Participants needed to choose the disyllabic sequence (one out of four) that corresponded to the way that a NS would pronounce the previous two individual monosyllables together. They were told that none of the disyllabic sequences in any tonal combinations were real words, though most (if not all) individual monosyllables were real words.[31]

There were two practice trials (yuan2la1, zai4ni3 – these two tonal combinations were not tested). Given that participants were not able to return to the previous page to listen to the monosyllables again, in the practice trials, I emphasized that they had to "remember the sounds [the monosyllables they listened to]" before clicking "Next" and until they had chosen the disyllabic sound files in that trial. I indicated that they could just repeat the monosyllables in their heads without saying them out loud, but some still repeated them quietly to themselves. They were not allowed to take notes on the monosyllables, because they might have written down tone

[31] The concept of "words" in Chinese is not always straightforward. For more discussion on "words" in Chinese, see Packard (2000, Chapter 2). Most Chinese words are monosyllabic historically but disyllabic in modern times (Duanmu, 1999). In the experiment, participants only hear the syllables, so they may or may not think of a particular morpheme. Even if they do, they might think of different morphemes, which correspond to different characters, to map the sounds, e.g., yi3 can be 以, 已, or 椅, but 以 and 已 are probably the two characters that most Mandarin speakers would first think of due to their higher frequency. Out of all syllables (with four tones) in the stimuli, only yi3 and (perhaps you1) is used predominantly as a bound morpheme in Modern Chinese though it was used freely in classical Chinese. Most of the time, 以 and 已 require another syllable to be meaningful in Modern Chinese.

notations such as “33” or “✓✓” for the T3T3 condition (the critical condition - more explanation below), which would have likely encouraged them to choose the unnatural-sounding T3T3 sound files.

They were allowed to listen to the sound files as many times as they wanted, but whether they actually listened to all of the sound files was not controlled. Some NSs chose the correct disyllabic file after identifying the correct one; some did not necessarily start with the leftmost sound files. Most HSs and L2ers did listen to all four sound files (even multiple times) before making a decision.

For completeness, I report how the tasks were modified. In the first version, after reading two Chinese characters, participants had to choose the correct sound file (one out of four) for the two monosyllables. Then, they were asked to choose the “correct” pronunciation of the disyllabic sequence created by the two monosyllables. Some participants, even NSs of Mandarin, chose the unnatural-sounding T3T3 sound files after two T3 monosyllables. Based on post-task feedback, some participants chose the T3T3 sound files because they thought T2T3 was just a “colloquial” pronunciation while T3T3 was the “correct” way; others chose the T2T3 sound files and were able to articulate the full-T3 sandhi rule to me. An investigation of their production data from the Tone Production Task also confirmed that all NSs indeed pronounced T3T3 as T2T3. For some of the HSs and L2ers, the fact that they needed to read the beginner-level Chinese characters (in isolation) in the first version made it too difficult.[32] Thus, the tasks were modified in two ways: first, participants were asked to “[c]hoose one which you believe is the native pronunciation” instead of

[32] I believe reading the Chinese characters “in isolation” was the major problem for HSs and L2ers, because they were used to reading Chinese words or phrases (in context) instead of as single characters. For example, the Mandarin word for ‘easy’ is made up of two syllables: rong2yi4, one HS pronounced rong2 while seeing the character yi4.

the "correct" pronunciation; second, the Chinese character reading component was removed and they did not have to choose sound files for monosyllables.

3.2. Participants in the tone task

Data from 42 Mandarin NSs, 21 HSs, and 25 L2ers were included in the Tone Identification Task, after excluding seven participants based on the language background questionnaire. The seven participants that were excluded included one participant who grew up in Singapore and self-reported that he felt more comfortable speaking Chinese, but reading English; one L2er who did not complete the proficiency test, as he self-reported not being able to read much; one HS who considered Mandarin, English, and Indonesian all to be native languages; and four HSs who had a non-Mandarin Chinese language/dialect as one of their native languages (two with Cantonese, one with Taiwanese – aka Taiwanese Southern Min, and one with Shanghainese). These HSs were excluded because extensive experiences with other tonal languages since childhood may differentiate them from other HSs of Mandarin.[33] For example, Hao (2012) found that L1-Cantonese L2-Mandarin learners have additional difficulty with T1 and T4 in Mandarin, which is not found with L1-English L2-Mandarin learners. Similarly, Tsukada et al. (2015) found that L1-English L2-Mandarin learners outperformed HSs of Cantonese in differentiating Mandarin T1 vs. T4.

In terms of testing locations, NSs were tested in Beijing (n=22) and Taipei (n=20) in a private setting. All NSs were born and raised in mainland China or Taiwan, had not spent more than one

[33] Some HSs did not include non-Mandarin Chinese languages/dialects as native language(s) but did include them later in the questionnaire when I asked if they knew any non-Mandarin Chinese languages/dialects and/or when I asked what languages they were exposed to during childhood (I listed possible sources, including parents, relatives in the same household). They were kept in the dataset, since experience with primarily Mandarin and some non-Mandarin Chinese languages/dialects was not uncommon among my HS participants.

year abroad, had not been immersed in a bilingual environment such as an English-speaking international school, and were not students of linguistics or Chinese pedagogy. The reason to recruit NSs outside of the United States was to minimize English influence and/or L1 attrition, which is known to influence at least one of the target structures investigated in this dissertation (see S. Zhang, 2018, who found L1 attrition on LD readings of reflexives). While T3 sandhi occurs in all Mandarin varieties, recruiting NSs from two varieties of Mandarin helped to account for possible dialectal differences should any differences occur. All HSs and L2ers reported in this chapter completed the tasks in the United States or Taiwan in a private setting. HSs tested in Taiwan grew up in the United States or Canada but moved to Taiwan as adults or were visiting Taiwan at the time of testing. The participants' relevant background information and scores on the Mandarin proficiency test are summarized in Table 3.1. While the proficiency was matched ($p = 0.4674$), the nature of the test (vocabulary and cloze test in written Chinese characters) might have underestimated some participants' Mandarin proficiency. Ideally, the proficiency test should have included a listening component. This is a limitation of the present dissertation that I will address later (in Chapter 8).

3.3. Materials

Seven conditions testing disyllabic tones were created; see Table 3.2. Conditions refer to the underlying/citation tones. While T3T3 must be realized as T2T3 due to tone sandhi when pronounced together, the condition is named T3T3, referring to T3T3 disyllabic sequence in underlying tones, even though the correct answer would be T2T3. For example, after listening to da3 and yi3 (da3 being placed above yi3, but on the same page), the answer for da3yi3 is da2yi3 (out of the four possible choices: da1yi3, da2yi3, da3yi3, and da4yi3) due to T3 sandhi. T3T1,

Table 3.1. Information about the participants for the Tone Perception Task

	NSs (*n* = 42)	HSs (*n* = 21)	L2ers (*n* = 25)
Age of testing	Mean 22.3 (range 19 - 37)	Mean 21.5 (range 19 - 33)	Mean 27 (range 20 - 46)
Age of Acquisition of Mandarin (in years)	N/A	Mean: 0.05 (range 0-1) 20 since birth 1 at 1	Mean: 19.3 (range 8-31) 3 at ages 8 through 12 4 ages 14 through 17 18 ages 18 and up
Age of Acquisition of English (in years)	N/A	Mean: 1.6 (range 0-5) 13 since birth 8 before or at age 5	N/A
Average years of Mandarin classes	N/A	Mean: 6.2 (range 0.7-14)	Mean: 3.8 (range 0.4-13)
Age of arrival in the United States	N/A	16 born in the United States 1 at age 2 1 at age 4 3 born in Canada	19 born in the United States, but 1 grew up in the United Kingdom 6 born outside the United States[a]
Proficiency test (max score = 40 for HSs and L2ers, but 16 for NSs)	Mean 15.9 (range 15-16)	Mean 26.7 (range 12-39)	Mean 28.4 (range 13-38)

Note. [a]1 born in Canada; 1 born in Bahamas, but grew up in Canada; 4 born in the United Kingdom

Table 3.2. Conditions in the tone identification task (correct answer **bolded**)

Underlying/citation tone	Sound files	Purpose of the condition(s)
T3T3	T1T3, **T2T3**, T3T3, T4T3	Critical condition: T3 sandhi
T1T3	**T1T3**, T2T3, T3T3, T4T3	These three conditions have T3 as the second syllable; T1T3 and T4T3 conditions are included to demonstrate that the T3 sandhi rule only applies to T3 (as a first syllable) and not to T1 and T4 before T3; while the T3 sandhi rule does not apply in the T2T3 condition, the realization is the same as the T3T3 condition
T2T3	T1T3, **T2T3**, T3T3, T4T3	
T4T3	T1T3, T2T3, T3T3, **T4T3**	
T3T1	T1T1, T2T1, **T3T1**, T4T1	These three conditions have T3 as the first syllables to demonstrate that the T3 sandhi rule only applies before another T3 and not before T1, T2, or T4
T3T2	T1T2, T2T2, **T3T2**, T4T2	
T3T4	T1T4, T2T4, **T3T4**, T4T4	

T3T2, and T3T4 conditions are necessary to establish that the T3 sandhi rule only applies before another T3 and not before T1, T2, or T4. Similarly, T1T3 and T4T3 conditions are necessary to establish that the T3 sandhi rule only applies when T3 (as a first syllable) is before another T3 and does not apply to T1 and T4 before a T3. For the T2T3 condition, while T3 sandhi does not apply, the outcome is the same.

With five tokens per condition, there were 35 trials, each presented with four sound files. Five common Mandarin syllables (yi, ma, da, you, and ke) that have four different tones (not every Mandarin syllable can be pronounced in all four tones) were chosen. Vowels (such as the close front rounded vowel [y]) and consonants (such as the voiceless alveolo-palatal affricate [tɕ]) that are known to be difficult for English speakers were avoided. Before imposing tones, there were five disyllabic sequences: yima, dayou, yike, keda, and dayi. Then, each of them was imposed with the seven tonal combinations. For example, dayi was tested as da1yi3, da2yi3, da3yi3, da4yi3, da3yi1, da3yi2, and da3yi4. None of the disyllabic sequences were real words.

There were five blocks, each with seven conditions. Blocks were randomized, but the trials within blocks were not (as participants who saw the T3T3 and T2T3 conditions consecutively might have become aware of the purpose of the task). The four sound files for each trial were presented in random order except for the T3T3 condition. For the T3T3 condition, to avoid participants from just choosing T3T3 without having the chance to hear T2T3, the sound files of the unnatural-sounding T3T3 were placed before (i.e., on the left of) the sound files of T2T3 in two tokens (in two different blocks) and after the sound files of T2T3 in two other tokens. The fifth token had a random order. For the full list of blocks and stimuli, see Appendix A.

In each of the seven conditions, there were four types of sound files. Note that the second syllable of all four types of sound files in each condition were the same tone. (This design reduced

the choices participants had to make since having (4*4=)16 choices might have been overwhelming.) In four (out of the seven) conditions (T3T3, T1T3, T2T3, and T4T3; the top four rows in Table 3.2), the second syllable was always T3 in all sound files; for the other three conditions (T3T1, T3T2, and T3T4; the bottom three rows in Table 3.2), the second syllable for the corresponding conditions are T1, T2, and T4 respectively. Thus, participants can answer correctly as long as they can identify which tone the first syllable has, without identifying the second syllable, because all four sound files in a given trial have the same second syllable.

3.4. A note on the sound materials (recording)

The stimuli were recorded by a female NS of (Taiwanese) Mandarin in a sound-attenuated booth. While T3 in Taiwanese Mandarin is typically realized as half-T3 [21], even in isolation, she deliberately produced clear, full T3 [214], which is easier for listeners to identify. While it is generally very difficult for a NS to produce T3T3 sequences without any pause in between, the speaker was able to produce it without a pause. I later listened to the recordings and checked the pitch contour in Praat to confirm that the first syllable in all T3T3 sequences was indeed T3, not T2. The stimuli were not synthesized nor modified.

3.5. Predictions

Given that English lacks lexical tones, there might not be a transfer effect, particularly not with the critical condition of T3T3. However, the transfer effect might show up in T4 (high-falling tone) due to English intonation. I briefly discuss this possibility before addressing the critical T3T3 condition. T4 is easy to identify for English speakers in word-final position or isolation since English declarative sentences have the falling intonation. However, the T3T4 condition here does

not require participants to identify T4 in order to choose the correct sound file (out of T1T4, T2T4, T3T4, T4T4) – as all they need to select the correct answer is to identify the first syllable as T3. Thus, this condition is predicted to be difficult because of T3, which is easily confused with T2 (see below). When T4 is in word-initial position, previous studies have found that it becomes difficult to identify (e.g., Hao, 2018). If so, the T4T3 condition could present some difficulty for HSs and L2ers.

For the critical T3T3 condition, recall that T3 becomes T2 if followed by another T3. The prediction is that, when asked to choose a disyllabic sequence after hearing two T3 syllables, those who do not know T3 sandhi will choose the sound file of T3T3, which sounds unnatural to NSs. For HSs and L2ers, the T2T3 condition is expected to be slightly easier than T3T3, since they do not need to know the T3 sandhi rule. Given that T2 and T3 are mutually confusable, participants might have a difficult time differentiating among T2T3 (as well as T3T3, which is realized as T2T3), T3T2, and T2T2. However, the four sound files in the T3T3 condition are T1T3, T2T3, T3T3, and T4T3, and T3T2 and T2T2 are not among them. Thus, participants can correctly choose T2T3 as long as they can identify the first syllable is a T2. Of course, the unnaturalness of T3T3 might help HSs and L2ers to make the correct choice.

Next, I discuss what it means when participants correctly choose T2T3 for the T3T3 condition but appear to do so due to overgeneralization by making mistakes in other conditions. Note that the second row through the fourth row in Table 3.2 all have T3 as the second syllable (i.e., T1T3, T2T3, and T4T3), and the last three rows all have T3 as the first syllable (i.e., T3T1, T3T2, and T3T4). For the conditions ending with T3 syllables, if participants still chose T2T3 sound files for T1T3 and T4T3 conditions, this would indicate overgeneralization of T3 sandhi to other disyllabic sequences that have T3 as the second syllable. For the conditions starting with a

T3 (i.e., T3T1, T3T2, and T3T4), if participants correctly choose T2T3 for the T3T3 condition, but wrongly choose T2T1, T2T2, and T2T4 respectively for the T3T1, T3T2, and T3T4 conditions, this would indicate overgeneralization of T3 sandhi to other disyllabic sequences that have T3 as the first syllable or are simply misperceiving T3 as T2.

For the three conditions ending with T3 (T1T3, T2T3, and T4T3) vs. the three conditions starting with T3 (T3T1, T3T2, and T3T4), the latter is predicted to be more difficult. First, T2 and T3 are mutually confusable, so HSs and L2ers might choose T2T1, T2T2, and T2T4 respectively for the T3T1, T3T2, and T3T4 conditions. Second, given that T3 before T1, T2, and T4 is pronounced as half-T3 (21; low falling), it may be perceived as T4 (51; high falling), as found in H. Zhang (2017) with low-level learners, but not with advanced-level learners. If so, HSs and L2ers might incorrectly choose T4T1, T4T2, and T4T4 for the T3T1, T3T2, and T3T4 conditions. These predictions are summarized in Table 3.3.

Table 3.3. Predictions for the tone identification task (correct answer **bolded**)

Citation/underlying tone	Sound files	NSs	HSs and L2ers
T3T3	T1T3, **T2T3**, T3T3, T4T3	T2T3	T3T3 if they do not know the full T3 sandhi rule
T1T3	**T1T3**, T2T3, T3T3, T4T3	T1T3	T2T3 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence ending with a T3 syllable
T2T3	T1T3, **T2T3**, T3T3, T4T3	T2T3	T2T3
T4T3	T1T3, T2T3, T3T3, **T4T3**	T4T3	T2T3 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence ending with a T3 syllable
T3T1	T1T1, T2T1, **T3T1**, T4T1	T3T1	T2T1 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence starting with a T3 syllable or simply misperceive T3 as T2; T4T1 if they misidentify half-T3 as T4

Table 3.3. Predictions for the tone identification task (correct answer **bolded**) (cont'd)

Citation/underlying tone	Sound files	NSs	HSs and L2ers
T3T2	T1T2, T2T2, **T3T2**, T4T2	T3T2	T2T2 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence starting with a T3 syllable or simply misperceive T3 as T2; T4T2 if they misidentify half-T3 as T4
T3T4	T1T4, T2T4, **T3T4**, T4T4	T3T4	T2T4 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence starting with a T3 syllable or simply misperceive T3 as T2; T4T4 if they misidentify half-T3 as T4

Recall that participants can just pay attention to the first syllables of the four sound files they needed to choose from without paying attention to the second syllables. However, given that in previous studies (e.g., Hao, 2018) the initial syllables were more difficult to identify than the second syllables, this does not necessarily make the task much easier.

For HS/L2 comparisons, HSs are predicted to outperform L2ers in all conditions, including the T3T3 condition (T3 sandhi), which is acquired after the basic tones. Monolingual Mandarin-speaking children acquire T3 sandhi by age three, so HSs presumably also acquire it even when living in an English-speaking country since their Mandarin exposure should still be extensive by that time. Furthermore, they would never have heard the unnatural-sounding T3T3 sequences in their input.

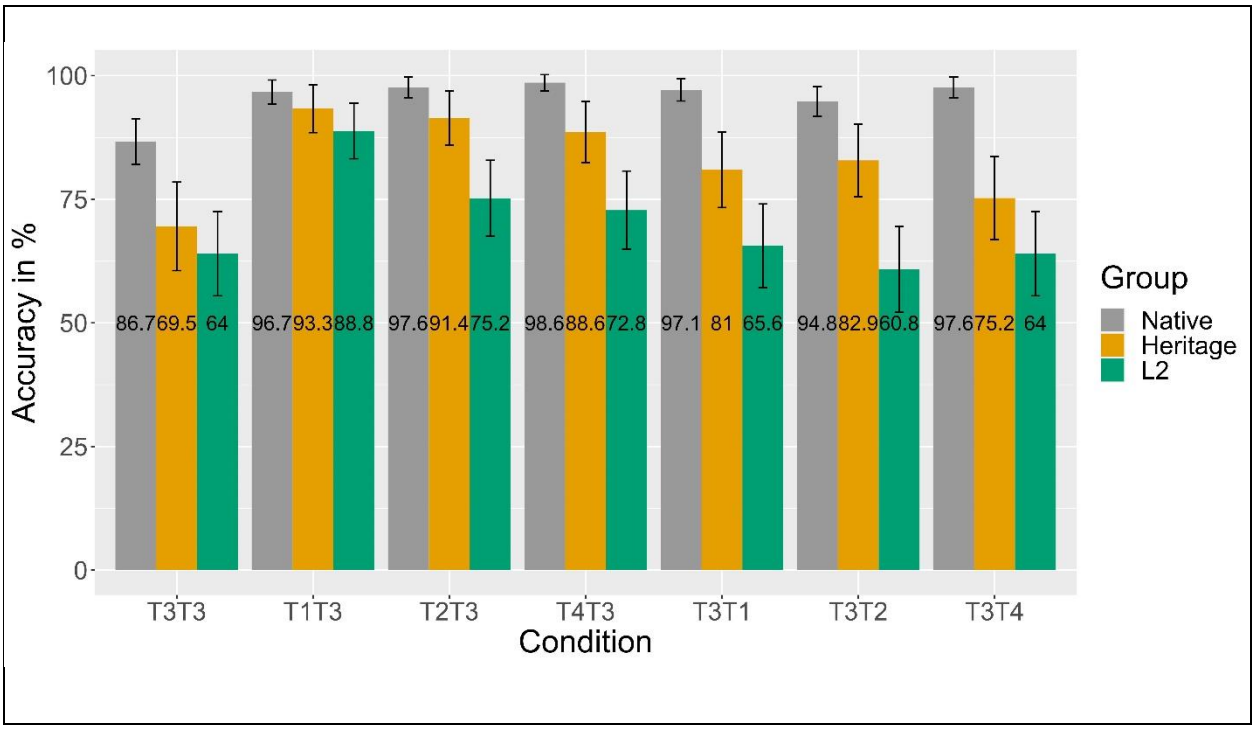
3.6. Results

3.6.1. Group analysis

A correct response was coded as "1" and an incorrect response as "0". (Note that T2T3, but not T3T3, was coded as the correct answer for T3T3 condition here; choices of T2T3 vs. T3T3 will be reported later.) Then, the raw scores in each condition (range 0-5 as there were five tokens

per condition) were averaged across the participants and converted to percentages. Figure 3.1 shows the mean accuracy of the group results.

Figure 3.1. Tone Identification Task results: Mean accuracy (in%)



Data were analyzed in a logistic mixed-effects model (Jaeger, 2008) using the *glmer()* function in the *lme4* package in R (R Core Team, 2019). The model included group (NSs, HSs, and L2ers) and condition (T3T3, T1T3, T2T3, T4T4, T3T1, T3T2, and T3T4) and their interaction as fixed effects; the random effects included a random intercept for subjects and a random intercept for items. Dummy coding was used. The reference level for group was the NSs, while the reference level for condition was the T3T3 condition. Using the *Anova()* function in the *car* package to assess the overall effect, the model output is presented in Table 3.4. There is a main effect of group, suggesting that NSs were more accurate than HSs, who in turn, were more accurate than L2ers. There is also a main effect of condition, with overall lower accuracy in the T3T3 condition and

higher accuracy in the T1T3 condition. The two-way interaction of condition and group is significant, with HSs and particularly L2ers being more accurate on T1T3 condition than other conditions.

Table 3.4. Results from the logistic mixed-effects model on tone

	Chisq	Df	Pr(>Chisq)	
(Intercept)	51.879	1	5.902e-13	***
condition	47.523	6	1.472e-08	***
Group	14.863	2	0.0005923	***
condition:Group	34.371	12	0.0005894	***
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Pairwise comparisons were conducted via *emmeans* (Lenth, 2019) following the significant interactions; the p-values are significant at the Tukey-adjusted alpha level of .05. See Table 3.5. I will discuss the between-group differences before the within-group differences. The only condition in which all three groups patterned similarly was the T1T3 condition. For the T3T3 condition, HSs patterned with NSs and L2ers; NSs were much more accurate than L2ers (87% vs. 64%), although the difference is marginal ($p = 0.06$). NSs were more accurate than HSs in four (out of seven) conditions (significant on T4T3, T3T1, and T3T4; marginally on T3T2) and were more accurate than L2ers in six (out of seven) conditions (significant on T2T3, T4T3, T3T1, T3T2, and T3T4; marginally on T3T3). HSs patterned with L2ers in all conditions. For within-group differences which made comparisons to the T3T3 condition, NSs were more accurate on all conditions except for T3T2, which had similar rates of accuracy with the T3T3 condition. HSs’ T3T1, T3T2 and T3T4 conditions had similar accuracy rates with the T3T3 condition. However, compared to the T3T3 condition, HSs were significantly more accurate on the T1T3 and T2T3 conditions, and marginally more accurate on T4T3 conditions. L2ers were more accurate on the T1T3 condition than the T3T3 condition; the T3T3 condition had scores similar to the other five conditions.

Table 3.5. Pairwise comparison results from the logistic mixed-effects model on tones

contrast	estimate	SE	df	z.ratio	p.value	
T3T3,Native - T3T3,Heritage	1.651448	0.537	Inf	3.077	0.2058	
T3T3,Native - T3T3,L2	1.801175	0.514	Inf	3.505	0.0615	
T1T3,Native - T1T3,Heritage	1.702519	0.753	Inf	2.260	0.7841	
T1T3,Native - T1T3,L2	2.033945	0.691	Inf	2.944	0.2804	
T2T3,Native - T2T3,Heritage	2.446500	0.782	Inf	3.127	0.1816	
T2T3,Native - T2T3,L2	3.629467	0.718	Inf	5.052	0.0001	
T4T3,Native - T4T3,Heritage	3.440107	0.862	Inf	3.989	0.0111	*
T4T3,Native - T4T3,L2	4.437029	0.820	Inf	5.408	<.0001	*
T3T1,Native - T3T1,Heritage	3.209784	0.712	Inf	4.505	0.0013	*
T3T1,Native - T3T1,L2	3.982449	0.683	Inf	5.833	<.0001	*
T3T2,Native - T3T2,Heritage	2.211414	0.635	Inf	3.483	0.0660	
T3T2,Native - T3T2,L2	3.405373	0.596	Inf	5.713	<.0001	*
T3T4,Native - T3T4,Heritage	3.836824	0.735	Inf	5.217	<.0001	*
T3T4,Native - T3T4,L2	4.320142	0.714	Inf	6.051	<.0001	*
T3T3,Heritage - T3T3,L2	0.149727	0.539	Inf	0.278	1.0000	
T1T3,Heritage - T1T3,L2	0.331427	0.671	Inf	0.494	1.0000	
T2T3,Heritage - T2T3,L2	1.182967	0.616	Inf	1.921	0.9424	
T4T3,Heritage - T4T3,L2	0.996923	0.590	Inf	1.689	0.9850	
T3T1,Heritage - T3T1,L2	0.772664	0.556	Inf	1.389	0.9987	
T3T2,Heritage - T3T2,L2	1.193959	0.559	Inf	2.134	0.8575	
T3T4,Heritage - T3T4,L2	0.483319	0.545	Inf	0.887	1.0000	
T3T3,Native - T1T3,Native	-2.063049	0.518	Inf	-3.982	0.0114	*
T3T3,Native - T2T3,Native	-2.518954	0.586	Inf	-4.301	0.0031	*
T3T3,Native - T4T3,Native	-3.169063	0.708	Inf	-4.478	0.0014	*
T3T3,Native - T3T1,Native	-2.274576	0.548	Inf	-4.150	0.0058	*
T3T3,Native - T3T2,Native	-1.420588	0.441	Inf	-3.218	0.1431	
T3T3,Native - T3T4,Native	-2.518975	0.586	Inf	-4.298	0.0031	*
T3T3,Heritage - T1T3,Heritage	-2.011978	0.465	Inf	-4.330	0.0027	*
T3T3,Heritage - T2T3,Heritage	-1.723902	0.429	Inf	-4.020	0.0098	*
T3T3,Heritage - T4T3,Heritage	-1.380404	0.395	Inf	-3.495	0.0635	
T3T3,Heritage - T3T1,Heritage	-0.716240	0.350	Inf	-2.044	0.8995	
T3T3,Heritage - T3T2,Heritage	-0.860622	0.358	Inf	-2.403	0.6836	
T3T3,Heritage - T3T4,Heritage	-0.333599	0.334	Inf	-0.998	1.0000	
T3T3,L2 - T1T3,L2	-1.830278	0.373	Inf	-4.903	0.0002	*
T3T3,L2 - T2T3,L2	-0.690661	0.317	Inf	-2.179	0.8333	
T3T3,L2 - T4T3,L2	-0.533209	0.313	Inf	-1.704	0.9834	
T3T3,L2 - T3T1,L2	-0.093302	0.305	Inf	-0.305	1.0000	
T3T3,L2 - T3T2,L2	0.183611	0.303	Inf	0.606	1.0000	
T3T3,L2 - T3T4,L2	-0.000007	0.304	Inf	0.000	1.0000	

Note: * indicates p<.05

3.6.2. Error analyses for the T3T3, T3T1, T3T2, and T3T4 conditions

Because NSs did not uniformly choose T2T3 in the T3T3 condition, further analysis was conducted on their responses to such conditions. Additionally, it is informative to see whether HSs and L2ers also chose T3T3 or if they just chose the other two entirely incorrect options. To examine what tones participants chose for the T3T3 condition, error analysis was conducted. Error analysis was conducted for other conditions as well, but only the T3T1, T3T2, and T3T4 conditions were discussed further, since specific error patterns were precited. Error analysis conducted with the three conditions ending with T3 (i.e., T1T3, T2T3, and T4T3) revealed that the errors were evenly distributed among the three incorrect choices and/or participants had low error rates.[34] Due to lack of obvious patterns, I do not discuss them further.

Figures 3.2 and 3.3 below each presents a percentage breakdown of the answer choices for the T3T3 and the T3T1 condition by group. In Figure 3.2, NSs incorrectly chose the T3T3 sound file 11% of the time while HSs and L2ers did so over 25% of the time. L2ers additionally chose both T1T3 and T4T4 about 5% of the time. In Figure 3.3, as predicted, due to overgeneralization and/or mutual confusion between T3 and T2, HSs and L2ers each incorrectly chose T2T1 about 20% of the time. Additionally, about 10% of the answers made by L2ers mistook half-T3 as T4, thus resulting in the incorrect choosing of T4T1. HSs appear to have correctly identified half-T3 as T3. Similar patterns were found with T3T2 (Figure 3.4) and T3T4 (Figure 3.5). Both HSs and

[34] For the T1T3 condition, 3.8% of HS answers were incorrect on T2T3, 1% on T3T3, 1.9% on T4T3; 5.6% of L2er answers were incorrect on T2T3, 0.8% on T3T3, and 4.8% on T4T3. For the T2T3 condition, 5.7% of HS answers were incorrect on T1T3, 1.9% on T3T3, and 1% on T4T3; 10.4% of L2er answers were incorrect on T1T3, 8.8% on T3T3, and 1.4% on T4T3. For the T4T3 condition, HSs incorrectly chose T1T3 1.9% of the time, T2T3 7.6% of the time, and T3T3 1.9% of the time; L2ers incorrectly chose T1T3 11.2% of the time, T2T3 8% of the time, and T3T3 8% of the time.

L2ers incorrectly chose T2 when it was T3, with L2ers having higher error rates. In addition, only L2ers incorrectly identified half-T3 as T4 in disyllabic sequences starting with T3.

Figure 3.2. Tone task results: Error analysis for the T3T3 condition by group

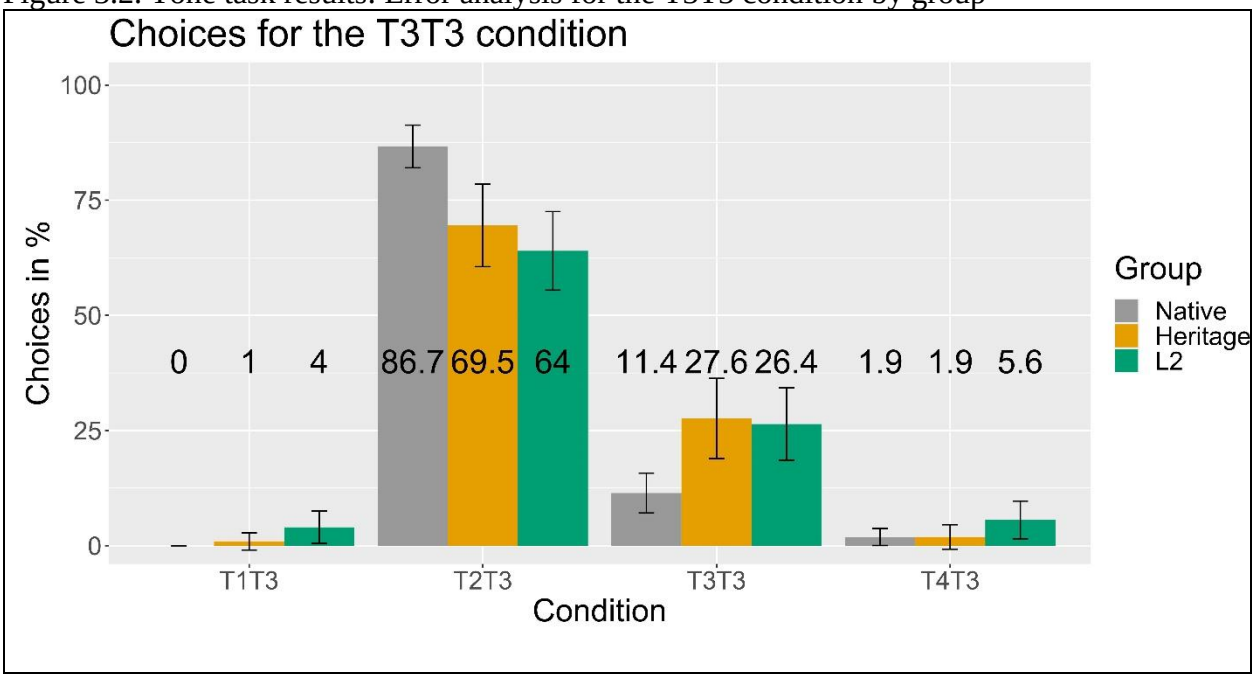


Figure 3.3. Tone task results: Error analysis for the T3T1 condition by group

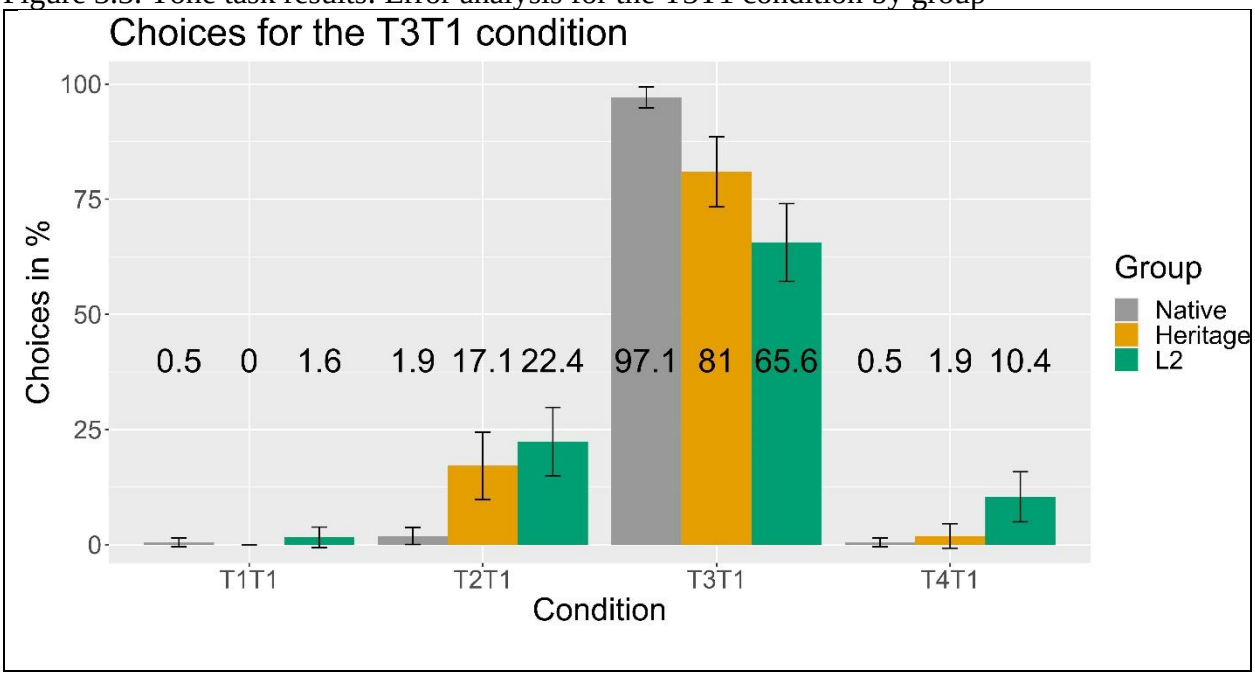


Figure 3.4. Tone task results: Error analysis for the T3T2 condition by group

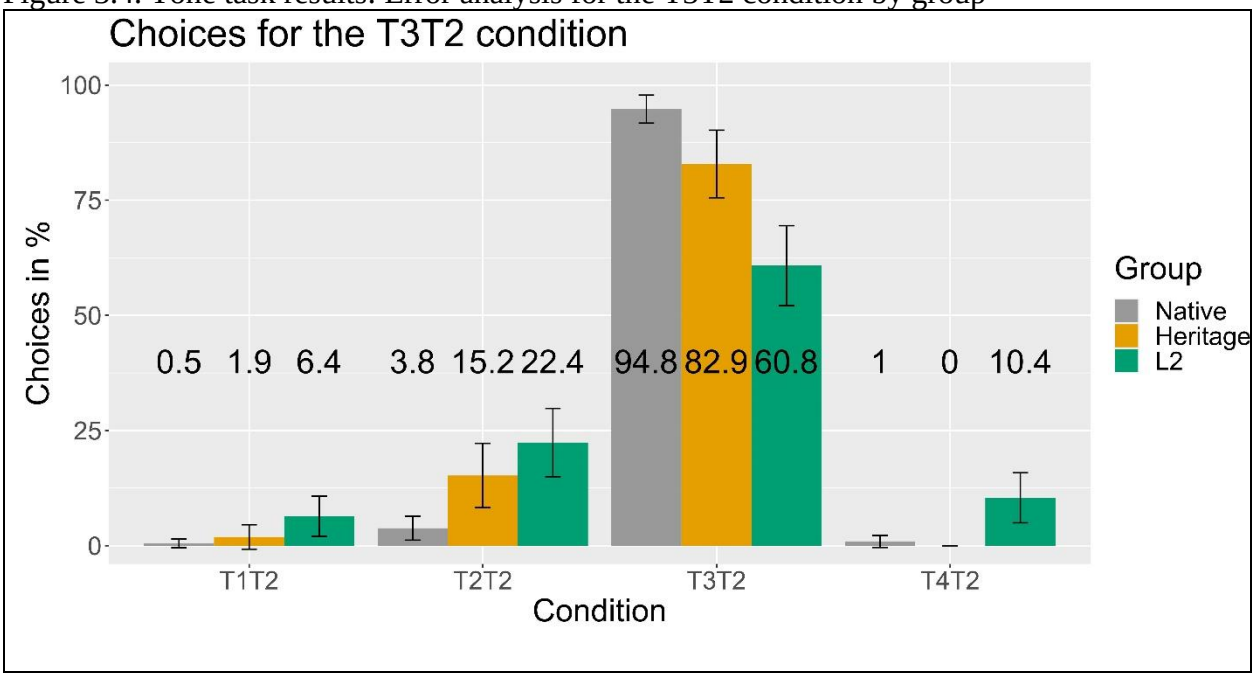
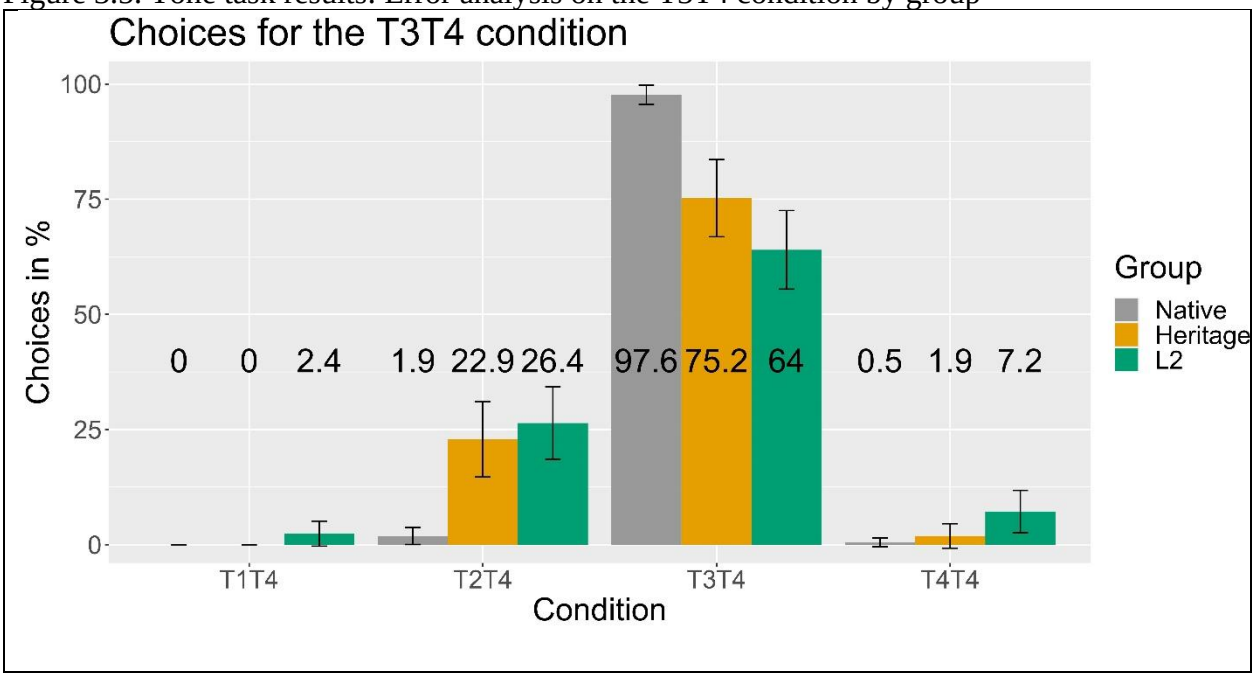


Figure 3.5. Tone task results: Error analysis on the T3T4 condition by group



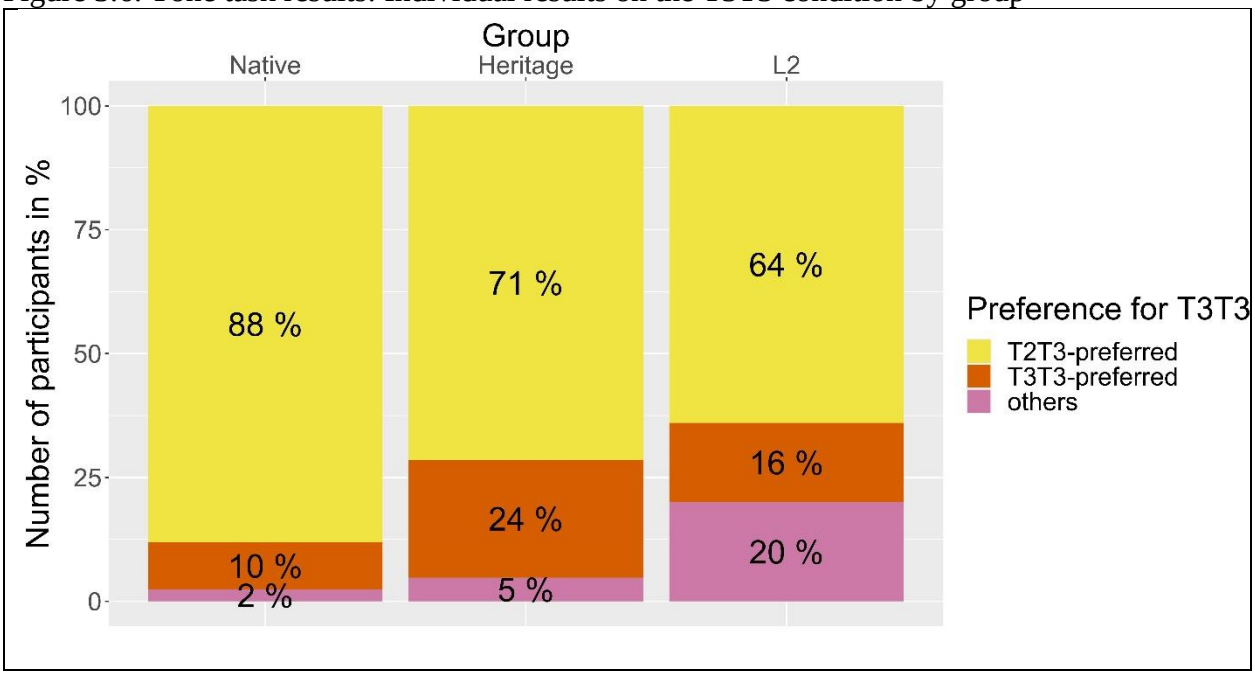
3.6.3. Individual subjects' analysis for the T3T3 condition

To examine what tones participants chose for the T3T3 condition, I coded their preference into three categories: T2T3-preferred, T3T3-preferred, and others. If a participant chose T2T3 or

T3T3 more than three times (>=3) out of five, they were classified as T2T3-preferred or T3T3-preferred respectively. The rest were classified as others.

Figure 3.6 below shows the preferred choices for the T3T3 condition by each group. Eighty-eight percent of NSs, 71% of HSs and 64% of L2ers correctly chose T2T3. However, 10% of NSs, 24% of HSs, and 16% of L2ers (incorrectly) chose T3T3. Additionally, 20% of L2ers incorrectly chose one of the other two sound files - either T1T3 or T4T3. This is consistent with the group results where L2ers showed the lowest accuracy (when I coded only T2T3, but not T3T3, to be "accurate").

Figure 3.6. Tone task results: Individual results on the T3T3 condition by group



3.7. Discussion

Recall that the broad RQs asked whether proficiency-matched HSs and L2ers can acquire language phenomena that are absent in or different from English and whether HSs have an advantage over proficiency-matched L2 learners. The specific RQs asked whether they correctly

choose T2T3 for the T3T3 condition. In the case of tone sandhi, there is nothing to be transferred from English. However, possible transfer of English intonation might influence the perception of T4 differently in different positions. While T4 in the final position is easier to identify than in other positions due to English intonation, the difficulty of the T3T4 condition here lies in T3, since both T2T4 and T3T4 are in the sound files for participants to choose from (out of T1T4, T2T4, T3T4, T4T4). In contrast, T4 in the beginning position is harder to identify. In this dissertation, L2ers did perform marginally worse on the T4T3 condition (at 72.8%) than on the T1T3 condition (at 88.8%) ($p = 0.078$), but performed similarly on the T2T3 condition (at 75.2%). By contrast, HSs perform similarly on the T1T3, T2T3, and T4T3 conditions, with over 88% accuracy. Thus, only L2ers show tentative evidence of being influenced by English intonation. Given that the test conditions did not include all possible T4 combinations, I do not further discuss English intonation transfer on T4 but move on to the critical condition of T3T3 below.

The unexpected results from NSs on the T3T3 condition are discussed before moving on to the performances by HSs and L2ers on all conditions. In the previous version of the tone task, participants were asked to choose the "correct" pronunciation, and many NSs choose T3T3. In this version where I asked them to "[c]hoose one which you believe is the native pronunciation". As in Figure 3.6, most (88%) NSs indeed chose T2T3 at least three out of five tokens but 10% still chose T3T3. It could be that these 10% of the NSs indeed think T3T3 is how it should be pronounced, which is largely contradictory to what we know from previous studies (and in the production task, which is not reported in this dissertation).[35] Prior studies have found that NSs cannot reliably differentiate between a T2 and a T3 that undergoes T3 sandhi, though some production studies have found slight differences between the two (some also found differences

[35] Another possibility is to check the five T3T3 sound files to see if there is a specific sound file that made NSs more likely to choose T3T3 rather than T2T3.

between real words and non-words, which is related to whether the T3 sandhi rule is computed on the fly or stored in the lexicon, see e.g., C. Zhang et al., 2015). But in any case, a T3 before another T3 cannot be naturally pronounced as a T3 (as occurs in the current task), so the fact that 10% of the NSs in this study did not apply the T3 sandhi rule has to be due to some undesired task effects.

Another possible source of the task effect is the over-attention to the first syllables. As mentioned above, participants could choose the correct answer by only paying attention to the first syllable. If they did so, it may explain why they chose T3T3 for the T3T3 condition. The experiment was set up in a way to minimize the possibility of choosing T3T3. Participants listened to one sound file (first syllable) and the second sound file (the second syllable) before clicking "next" to move on to the next page and were unable to return to the previous page. (If the six sound files were on the same page, participants could directly compare the first syllables in the four disyllabic sequences with the first monosyllables, which would encourage them to choose the T3T3 sound files for the T3T3 condition.) By having two separate pages, participants had to memorize the two monosyllables. If they pronounced the two monosyllables together (out loud or silently in their minds), the sequence should be T2T3 (since it is very difficult to pronounce the T3T3 sequence without any pause in between). Admittedly, some participants might have memorized the tone notation number (i.e., T1, T2, T3, T4), which would likely encourage them to choose the T3T3 sound file.

Given that about 10% of the NSs choose T3T3 for the T3T3 condition, some might question whether it is indeed "incorrect". However, there is no evidence for knowing T3 sandhi if T3T3 is counted as correct. While HSs patterned with NSs in the T3T3 condition, L2ers did not. For HSs and L2ers, the former chose T2T3 70% of the time and the latter chose T2T3 64% of the time. Using a three-token cutoff (out of five), 71% of HSs and 64% of L2ers choose T2T3 instead of

T3T3. Thus, it is safe to say that the majority of the HSs and over half of the L2ers did show clear evidence of knowing T3 sandhi. Data from the tone production task (not discussed in this dissertation) will provide complementary evidence for this. Recall that previous studies with L2 production found that some L2ers were able to correctly pronounce T2T3 when the underlying tones were T3T3 (based on perceptual judgements of NSs of Mandarin, e.g., W. Jin, 2019; H. Zhang, 2017). However, S. Chen et al. (2019), the first acoustic study on T3 sandhi by L2ers, found that L2ers' production was not native-like.

After discussing the T3T3 condition, I now turn to the other six conditions by HSs and L2ers. Both groups patterned with NSs in the T1T3 condition, but HSs additionally patterned with NSs in the T2T3 condition (as well as the T3T3 condition) while L2ers did not. Thus, while HSs patterned with L2ers in all conditions, HSs were more native-like than L2ers were. The three conditions starting with T3 (T3T1, T3T2, and T3T4) are predicted to be more difficult than the conditions ending with T3 (i.e., T1T3, T2T3, and T4T3) due to difficulty differentiating between T2 and T3/half-T3. This prediction is fully supported with HSs but not with L2ers, who struggled with all conditions except for the T1T3 condition. L2ers' difficulty with disyllabic tones is consistent with previous studies (e.g., Hao, 2012, 2018; Pelzl et al. 2019). For example, Hao (2012) found that disyllabic tones are more difficult than monosyllabic tones for L2ers. Additionally, first syllables were more difficult than second or final syllables. The difficulty of the first syllable relative to the second also happens with monolingual-speaking children (e.g., Wong & Strange, 2017). Among the three conditions ending with T3, HSs and particularly L2ers were more accurate on the T1T3 condition than the T2T3 condition, which in turn were slightly more accurate than the T4T3 condition. This is probably due to T4 being difficult to identify as a first syllable for

English speakers (Hao, 2018), because they are used to the falling intonation (T4 is the high-falling tone) at the end of an English utterance, as mentioned before.

In terms of frequency, tones and T3 sandhi are presumably very frequent, so both HSs and L2ers should have received much input from the very beginning (because acquiring tones and T3 sandhi does not require knowing vocabulary items). One exception is learners who focus on reading Chinese for comprehension and rarely listen to spoken Mandarin, but this is not the case for most of my L2 participants, given that 15 out of 25 L2ers reported here were tested in Taiwan. Despite ample input, L2ers have greater difficulty than HSs in identifying disyllabic tones and T3 sandhi, which is consistent with previous findings that age effects play an important role in the acquisition of phonology. In terms of processing, T3 sandhi is more difficult to process than the T2T3 sequences (C. Zhang et al., 2015). It is possible that the processing difficulty of T3T3 sequences relative to that of T2T3 sequences contributed to the significantly higher accuracy on the T2T3 condition than the T3T3 condition by NSs and HSs (but not L2ers), but this remains speculative in an offline task like the present one.

Returning to the broad RQs, while HSs and L2ers do not differ significantly, HSs were more native-like than L2ers were in the T3T3, T2T3, and T3T2 conditions. Thus, HSs still show a slight HS advantage over L2ers in tones and T3 sandhi in the domain of phonology, even though the effect is not as strong as hypothesized. It is probably because the critical condition of T3T3 has an undesired task effect that led even some NSs to choose the unnatural T3T3 pronunciation. While the performances of HSs and L2ers were largely non-native-like, HSs' greater difficulty with the conditions starting with T3 (T3T1, T3T2, and T3T4) rather than the conditions ending with T3 is compatible with prior findings (difficulty with T2 vs. T3 and difficulty with half-T3). L2ers also scored lower on the three conditions starting with T3, but not significantly lower than other

conditions (except for the T1T3 condition). L2ers' difficulties with disyllabic sequences seem to be more general, consistent with prior studies (e.g., Hao, 2012, 2018; Pelzl et al., 2019).

Chapter 4. **Judgment tasks: overall methodology**

This chapter introduces the overall procedure and format of the Acceptability Judgement Task (AJT) and the Truth Value Judgement Task (TVJT) that tested three linguistic phenomena: grammatical aspect, relative clauses (RCs), and anaphors (with a focus on long-distance (LD) reflexives). The AJT tested grammatical aspect and RCs while the TVJT tested RCs and anaphors. Chapters 5, 6, and 7 will discuss grammatical aspect, RCs, and anaphors, respectively.

4.1. Procedure

Similar to the Tone Identification Task, the AJT and the TVJT were administered on Qualtrics, a web-based survey tool. Participants completed the background questionnaire, the Mandarin proficiency test, the Tone Production Task (which is not included in this dissertation), the Tone Identification Task, the TVJT, and the AJT in that order. Additionally, some participants took a second proficiency test after the AJT, which is not included in this dissertation. The TVJT preceded the AJT because both tasks tested RCs, and the AJT was more explicit than the TVJT. For the proficiency test, native speakers (NSs) only completed the cloze portion, but not the vocabulary portion which required participants to choose English translations for Mandarin words.

The questionnaire was administered to gain a general understanding of the participants' linguistic backgrounds. The proficiency test independently measured participants' Mandarin proficiency so that the proficiency of heritage and L2 groups could be matched. The Mandarin proficiency test was a multiple-choice test, consisting of 24 vocabulary items and 16 cloze items. Participants needed to choose one out of five English translations for the Chinese words and one out of five Chinese words for a missing word in isolated Chinese sentences. The questions were

from the unused question pools taken from the Chinese placement test (only the reading portion) at UIUC.

All tasks were untimed. Together with the Tone Production Task and the Tone Identification Task, NSs typically spent 60-70 minutes while HSs and L2ers spent 1 hour 40 minutes to 2.5 hours.

4.2. Participants in the AJT and TVJT

Data from 62 Mandarin NSs, 39 HSs, and 36 L2ers were analyzed for the AJT, while data from 62 Mandarin NSs, 44 HSs, and 41 L2ers were analyzed for the TVJT, since some participants did not finish the AJT. The aforementioned numbers were obtained after excluding 13 participants based on the language background questionnaire: one L2er from the Philippines who was a NS of both English and Filipino, one HS of Mandarin and Korean, and 11 HSs of non-Mandarin Chinese languages/dialects, such as Cantonese, or HSs of both Mandarin and other non-Mandarin Chinese languages/dialects. These 11 HSs were excluded because grammatical aspect and properties of LD reflexives in some non-Mandarin Chinese languages/dialects are likely different from Mandarin. For example, Cole, Hermon, and C. L. Lee (2001) found different discourse properties of LD reflexives in Singapore Mandarin and Singapore Teochew. Two t-tests which examined the AJT and the TVJT revealed no significant differences between HSs and L2ers on the Mandarin proficiency test ($p > 0.05$); therefore, no HS or L2er was excluded based on proficiency.

In terms of testing, as already introduced in Chapter 3, NSs were tested in China (Beijing) (n=22) and Taiwan (Taipei and Hsinchu) (n=40) in a private setting. They were all born and raised in China or Taiwan, had not spent more than one year abroad, had not been immersed in a bilingual environment such as an English-speaking international school, and were not students of linguistics or Chinese pedagogy. The reason for recruiting NSs outside of the United States was to minimize

English influence and/or L1 attrition (see S. Zhang, 2018, who found L1 attrition on LD readings of reflexives after an average of 13 years living in the United Kingdom.). In addition, the recruitment of NSs from two varieties of Mandarin helped to account for possible dialectal differences. All HSs and L2ers completed the study in the United States or in Taiwan in a private setting, except for three L2ers who completed the study remotely on their own (for the TVJT only, and not the AJT). HSs tested in Taiwan had either grown up in the United States or Canada but moved to Taiwan as adults or were visiting Taiwan at the time of testing. The participants' relevant background information and scores on the Mandarin proficiency test are summarized in Table 4.1 for both the TVJT and AJT, though five of the HSs and five of the L2ers who completed the TVJT did not complete the AJT. I did not separate HSs further into HSs of Taiwanese Mandarin vs. HSs of mainland Mandarin.

4.1. Overview of the AJT

The AJT includes 16 conditions that tested aspect marking (in four separate experiments) and four conditions that tested RCs; see Table 4.2. A Likert scale from 1 (completely unacceptable) to 4 (fully acceptable) was used because some sentences could have been difficult to judge on a binary scale, even for NSs.

With five tokens per condition, there were (5*20=)100 target sentences per participant. With 16 conditions on aspect, ideally there would have been 16 filler conditions not on aspect, but it would have made the task too long. Thus, only ten filler conditions (six ungrammatical; four grammatical) were added. Fillers spanned the full range of (un)grammaticality, including different structures and did not contain aspect markers. In total, each participant read 150 sentences, including 100 target sentences and 50 fillers.

Table 4.1. Information about the participants for the TVJT

	NSs (*n* = 62)	HSs (*n* = 44)	L2ers (*n* = 41)
Age of testing	Mean 22.2 (range 19-37)	Mean 21 (range 18-33)	Mean 27.6 (range 18-55)
Age of Acquisition of Mandarin (in years)	N/A	Mean 0.05 (range 0-1) 42 at birth 2 at age 1	Mean: 19.6 (range 8-49) 3 at ages 8 through 12 12 ages 13 through 17 26 ages 18 and up
Age of Acquisition of English (in years)	N/A	Mean: 1.5 (range 0-5; two NAs)	N/A
Average years of Mandarin classes	N/A	Mean: 5.9 (range 0-14)	Mean: 3.6 (range 0.4-13)
Age of arrival in the United States	N/A	30 born in the United States 2 at age 1 2 at age 2 1 at age 3 2 at age 4 3 at age 5 4 born in Canada	30 born in the United States, but 1 grew up in the United Kingdom 11 born outside the United States[a]
Proficiency test (max score = 40 for HSs and L2ers, but 16 for NSs)	Mean 15.9 (range 15-16)	Mean 27.1 (range 11-40)	Mean 29.5 (range 10-40)

Note. [a]1 born in Australia; 1 born in Canada; 1 born in Bahamas, but grew up in Canada; 4 born in the United Kingdom; 1 born in India; 1 born in Saint Vincent and the Grenadines; 1 born in Germany with U.S. parents, and moved to the U.S. at age 6; 1 born in Honduras with U.S. parents, and moved to the United States at age 17

Table 4.2. Conditions in the AJT

	Conditions	Number of (un)grammatical conditions
Aspect	In each of the four separate experiments (4 lexical aspect), there were 4 conditions (4 levels of grammatical aspect)	7 ungrammatical; 9 grammatical
RCs	4 conditions = 2 (RC type) * 2 (position of head noun)	2 ungrammatical; 2 grammatical
Fillers		6 ungrammatical; 4 grammatical

To avoid priming (i.e., preventing participants from overaccepting sentences when they see similar sentences), four lists were created using a Latin-square design. Items were arranged into five blocks and randomized within each block, following standard procedure (Cowart, 1997); no trials from a single condition occurred sequentially. To prevent participants from answering the questions strategically or analytically, participants could not return to the previous block (the previous page on the web-based survey) to change their answers.

4.2. Overview of the picture-based TVJT

Each TVJT item consisted of a picture and a sentence. Participants were instructed to choose the TRUE response if the picture and sentence matched and the FALSE response if the picture and sentence did not match. The TVJT included four conditions testing RCs and six conditions testing anaphors; four filler conditions were added to balance the number of TRUE and FALSE responses (see Table 4.3). With six tokens per condition, there were 60 target sentences and 24 fillers. (Similar to the AJT, ideally, there should have been more fillers, but the TVJT would have been too long and demanding for HSs and L2ers.)

There were two lists.[36] Items were arranged into six blocks and pseudorandomized within each block, following standard procedure (Cowart, 1997); no trials from a single condition occurred sequentially. Using only two lists might potentially give rise to priming effects; therefore, in order to lessen priming effects and/or discourage participants from comparing different anaphor forms, the blocks were randomized for each participant. Additionally, participants were not allowed to

[36] To fully counterbalance the TVJT design, ideally there should be four lists on RCs, and six lists on anaphors. However, combining four lists and six lists is harder. Additionally, many self-directed verbs that are necessary to test local readings of anaphors are not very picturable (e.g., trusting oneself, be fond of oneself) and some verbs might be too difficult for HSs and L2ers. Thus, only verbs that were picturable and not too difficult for HSs and L2ers were used (by checking a L2-Mandarin vocabulary list, see L. Chang, 2012), resulting in only two lists.

return to previous test items (that is, previous pages on the web-based survey) to change their answers.

Table 4.4.3. Conditions in the TVJT

	Conditions	Number of TRUE and FALSE conditions
RCs	4 conditions = 2 RC type * 2 picture type	2 TRUE; 2 FALSE
Anaphors	6 conditions = 3 anaphors * 2 antecedent type	4 TRUE; 2 FALSE
Fillers		1 TRUE; 3 FALSE

4.3. Notes on test version (different scripts and Romanization)

Mandarin can be written in characters as well as alphabetically. Thus, four different versions were created which were identical except for the scripts used and a few lexical differences. There are two systems of Chinese characters: traditional Chinese characters used in Taiwan, Hong Kong and Macau and simplified Chinese characters used in mainland China. To my knowledge, many major U.S. colleges offer both systems of Chinese characters, though not usually throughout all levels. As Chinese characters are non-alphabetic, many HSs have difficulty acquiring Chinese literacy, even if they are fluent in speaking (e.g., Y. Xiao, 2008). To minimize literacy issues, Romanization (i.e., Pinyin) was added to the AJT and TVJT for the HS/L2 groups, but not to the proficiency test. For all the tasks, they could choose the characters (traditional or simplified Chinese) with or without Romanization. NSs from Taiwan and China were presented with traditional and simplified Chinese characters respectively. A few lexical differences exist in the versions of simplified vs. traditional Chinese characters due to dialectal differences. When constructing the stimuli, efforts were made to reduce lexical differences and to use basic vocabulary (based on a L2-Mandarin vocabulary list, see L. Chang, 2012). The stimuli were mainly constructed by the researcher, a NS of Taiwanese Mandarin, with the help of undergraduate research assistants who were NSs of either Taiwanese or mainland Mandarin. The stimuli were

reviewed by several NSs of Taiwanese Mandarin and mainland Mandarin who were not participants of the study to ensure the authenticity and naturalness of the sentences before collecting data.

Chapter 5. **Judgments of aspect in HS/L2 Mandarin**

This chapter reports on Mandarin aspect tested in the AJT (see Chapter 4 for the overall design and participant details). Recall that the broad research questions (RQs) for this dissertation pertain to (in)complete acquisition and selective HS advantages across domains, listed below. RQs 1b and 2b, also listed below, specifically target aspect and are instantiations of the broad RQs 1 and 2.

- Broad RQ 1: Can HSs and L2ers of Mandarin whose dominant language is English fully acquire the properties of Mandarin that are different from or absent in English?
- Broad RQ 2: Do HSs have selective advantages over proficiency-matched L2 learners, and does this vary by linguistic domain?
- RQ 1b: Can HSs and L2ers acquire the interaction between grammatical aspect and lexical aspect in Mandarin, despite differences from English?
- RQ 2b: Do HSs have an advantage over L2ers in aspect in the domain of morpho-semantics interface?

Given that Mandarin aspect is acquired before age five in monolingual children, HSs are expected to acquire aspect, since most are still regularly exposed to Mandarin by that time. For L2ers, given that the type of aspect marking tested in the present dissertation is in the domain of morpho-semantics interface and that morphology may present a particular challenge under the Bottleneck Hypothesis (Slabakova 2008, 2014), L2ers might have some difficulties in those contexts where English and Mandarin differ. Thus, a slight HS advantage is expected, especially on those early-acquired combinations (interaction between grammatical and lexical aspect) (e.g., -le with accomplishments), if not those later-acquired ones (e.g., *zai* with accomplishments). In the AJT employed here, I specifically tested whether HSs and L2ers correctly accept only those grammatical combinations.

5.1. AJT conditions on aspect

The purpose of this task is to examine whether participants know the interaction between lexical aspect and grammatical aspect in Mandarin. The AJT on aspect consisted of four separate experiments, each on one lexical aspect predicate (states, activities, accomplishments, and achievements, Vendler, 1967). Four target conditions, corresponding to four grammatical aspect markers (*-le, -guo, zai,* and *-zhe*), were created for each experiment. In total, there were 16 conditions on aspect. Under a grammatical-ungrammatical dichotomy, nine conditions were 'grammatical' and seven were 'ungrammatical'. However, three conditions actually had marginal acceptability, marked with a '?'; see Table 5.1. Participants rated acceptability on a 1-to-4 scale. The expected ratings are 'fully acceptable'' (rating: 4) for grammatical conditions (without the '?' marks) and 'completely unacceptable' (rating: 1) for ungrammatical conditions (without the '?' marks). The three conditions with marginal acceptability are expected to be rated in between.

Table 5.1. Conditions testing aspect in the AJT (✓= grammatical; ✗= ungrammatical)

	Perfective *-le*	Experiential *-guo*	Progressive *zai*	Durative *-zhe*
States[a]	✗ (incomplete)	✓	✗	✗
Activities	✗/? (incomplete)	✓	✓	✓/? (incomplete)
Accomplishments	✓	✓	✓	✗/?
Achievements	✓	✓	✗	✗

Note: [a]States here include only individual-level states

The target sentence frames are presented in (16). All were simple declarative sentences with pronouns as subjects, mostly *ta* 他 (with a human radical) 'he/she' and *ta* 她 (with a female radical) 'she'. For accomplishments, the direct objects were quantified NPs, so that accomplishments were classified as such without controversy (per Soh & J. Kuo, 2005). For states,

activities, and achievements, objects were in principle optional; whenever possible, however, an object was inserted after the verb to ensure that *-le* was interpreted as verb-final aspectual *-le* instead of sentence-final *LE*. All achievements included objects in the experimental stimuli.

(16) Four sentence types testing aspect marking in the AJT

A. States: Subject *zai* Verb-*le/-guo/-zhe* (Object).
B. Activities: Subject *zai* Verb-*le/-guo/-zhe* (Object).
C. Accomplishments: Subject *zai* Verb-*le/-guo/-zhe* Object.
D. Achievements: Subject *zai* Verb-*le/-guo/-zhe* Object.

There were four aspect conditions in each experiment, hence four lists (instead of 16 lists). With five tokens per condition, 20 token sets (VPs) were created per experiment; a sample token set exemplifying the four conditions testing aspect marking on states is provided in (17), on activities in (18), on accomplishments in (19), and on achievements in (20). The predicates used here are listed in Appendix B. Some state verbs were repeated with different objects, e.g., *love that woman* (in (17)) and *love that man*, because there are few individual-level state verbs that are both consistently classified as such and are known by HSs and L2ers.

(17) The four conditions testing aspect marking on states in the AJT

a) * With *-le* (incomplete):
Tā ài-le nàge nǚrén (他愛了那個女人)
He love-perf that woman
'He love-*le* that woman.'

b) With *-guo*:
Tā ài-guo nàge nǚrén (他愛過那個女人)
He love-exp that woman
'He love-*guo* that woman.'

c) *With *zai*:
Tā *zai* ài nàge nǚrén (他在愛那個女人)
He prog love that woman
'He *zai* love that woman.'

d) * With *-zhe*:
Tā ài-zhe nàge nǚrén (他愛著那個女人)
He love-dur that woman
'He love-*zhe* that woman.'

(18) The four conditions testing aspect marking on activities in the AJT

a) */? With *-le* (incomplete):
   Tā chàng-le gē (他唱了歌)
   He sing-perf song
   'He sing-*le* song.'

b) With *-guo*:
   Tā chàng-guo gē (他唱過歌)
   He sing-exp song
   'He sing-*guo* song.'

c) With *zai*:
   Tā zai chàng gē (他在唱歌)
   He prog sing song
   'He *zai* sing song.'

d) ? With *-zhe* (incomplete):
   Tā chàng-zhe gē (他唱著歌)
   He sing-dur song
   'He sing-*zhe* song.'

(19) The four conditions testing aspect marking on accomplishments in the AJT

a) With *-le*:
   Tā xiě-le sān-fēng xìn (她寫了三封信)
   She write-perf three-CL letter
   'She write-*le* three letters.'

b) With *-guo*:
   Tā xiě-guo sān-fēng xìn (她寫過三封信)
   She write-exp three-CL letter
   'She write-*guo* three letters.'

c) With *zai*:
   Tā zai xiě sān-fēng xìn (她在寫三封信)
   She prog write three-CL letter
   'She *zai* write three letters.'

d) */? With *-zhe*:
   Tā xiě-zhe sān-fēng xìn (她寫著三封信)
   She write-dur three-CL ii letter
   'She write-*zhe* three letters.'

(20) The four conditions testing aspect marking on achievements in the AJT

a) With *-le*:
   Ann dào-le měiguó (她到了美國)
   Ann arrive-perf the United States
   'Ann arrive-*le* the United States.'

b) With *-guo*:
   Ann dào-guo měiguó (她到過美國)
   Ann arrive-exp the United States
   'Ann arrive-*guo* the United States.'

(20) The four conditions testing aspect marking on achievements in the AJT (cont'd)

c) *With *zai*:
Ann zai dào měiguó (她在到美國)
Ann prog arrive the United States
'Ann *zai* arrive the United States.'

d) *With *-zhe*:
Ann dào-zhe měiguó (她到著美國)
Ann arrive-dur the United States
'Ann arrive-*zhe* the United States.'

5.2. Predictions

For conditions labeled ✓ or ✗ in Table 5.1, NSs are expected to give relatively uniform judgments. For those marked with a "?"in Table 5.1, however, even NSs are expected to give variable judgments. If HSs and L2ers map English past tense *-ed* to Mandarin *-le* and *-guo* and map English progressive *-ing* to *zai* and *-zhe*, three (out of the 16) conditions have cross-linguistic differences between Mandarin and English: states with *-le*, achievements with *zai*, and achievements with *-zhe*. Table 5.2 summarizes the interactions between *-ed/-ing* and lexical aspect in English: while *-ed* is compatible with all four lexical predicates, *-ing* is largely incompatible with states. (Note that stative progressives such as *love/like* and *think* are increasingly accepted by young NSs of American English (Smiecinska, 2003) and British English (Freund, 2016).)

Table 5.2. Interaction between lexical aspect and *-ed/-ing in English*

	-ed	-ing
States	✓ (e.g., 'I resembled my father.')	✗ (e.g., *'I am knowing math.')
Activities	✓	✓
Accomplishments	✓	✓
Achievements	✓	✓ (e.g., 'I am arriving.')

Assuming English transfer, HSs and L2ers are predicted to incorrectly accept states with *-le*, and achievements with *zai/-zhe*. For the three conditions with marginal acceptability, marked with a '?' in Table 5.1, HSs and L2ers are expected to not notice the somewhat degraded acceptability and therefore give higher ratings than NSs. The predictions for HSs and L2ers are summarized in Table 5.3, based on the cross-linguistic differences between English and Mandarin (a result by comparing Table 5.1 and Table 5.2 above). However, as *-zhe* often provides background information and is less frequent in simple declarative sentences, HSs and L2ers may reject it with all four lexical aspect predicates. At the same time, if HSs have an advantage over L2ers, the advantage might be more pronounced in combinations acquired early by monolingual children (e.g., accomplishments with *-le* by age three) than later-acquired ones (e.g., accomplishments with *zai* by age five) (e.g., P. Li & Bowerman, 1998).

Table 5.3. Predictions for HSs and L2ers under English transfer (*-le/-guo = -ed; zai/-zhe = -ing*) on the AJT testing aspect (cells shaded in black are the three conditions with marginal acceptability; cells shaded in black and grey all indicate higher ratings as compared to NSs)

	Perfective -le	Experiential -guo	Progressive zai	Durative -zhe
States[a]	✓	✓	✗	✗
Activities	✓	✓	✓	✓
Accomplishments	✓	✓	✓	✓
Achievements	✓	✓	✓	✓

Next I will unpack the predictions made by each of the four lexical aspect predicates, in consideration with both English transfer (for both HSs and L2ers) and AoA in monolingual children (for adult HSs). An HS advantage is expected only when a specific combination is early acquired in monolingual children and the pattern is different from English, since HSs would be

more likely to have acquired it as children. L2ers, on the other hand, might have difficulty acquiring these types of combinations. For states and activities, non-native-like performance might appear in the *-le* condition since HSs and L2ers might incorrectly accept these two combinations if they equate *-le* and *-guo* with English *-ed*. Considering AoA in monolingual children, since perfective markers are typically acquired first with telic VPs (achievements and accomplishments) in L1 acquisition, HSs might not fully acquire the ungrammaticality of *-le* with atelic VPs (states and activities). Thus, the HS advantage might not be as pronounced as in T3 sandhi (which is acquired by age three) as discussed in Chapter 3. Regarding *zai* with states, neither HSs nor L2ers are predicted to accept it because it is not allowed in English and monolingual children were never found to misuse *zai* with states (besides in elicitation). Table 5.4 and Table 5.5 summarize these predictions.

Table 5.4. Predictions for the AJT conditions on states (✓= grammatical; ✗= ungrammatical) (non-target-like responses are shaded in gray)

	NSs	HSs and L2ers under English transfer (*-le/-guo = -ed*; *zai/-zhe = -ing*)	Is an HS advantage expected?
-le	✗	✓	Yes, but only slightly (this combination is acquired later than -le with telic VPs in monolingual children)
-guo	✓	✓	No (Mandarin and English have the same pattern)
zai	✗	✗	No (mistakes not observed in L1 children in natural production, except for experimental elicitation; Mandarin and English share the same pattern)
-zhe	✗	✗	No (Mandarin and English share the same pattern)

Table 5.5. Predictions for the AJT conditions on activities (✓= grammatical; ✗= ungrammatical) (non-target-like responses are shaded in gray)

	NSs	HSs and L2ers under English transfer (*-le/-guo* = *-ed*; *zai/-zhe* = *-ing*)	Is an HS advantage expected?
-le	✗/?	✓	Yes, but only slightly (this combination is acquired later than -le with telic VPs in monolingual children)
-guo	✓	✓	No (Mandarin and English have the same pattern)
zai	✓	✓	No (Mandarin and English have the same pattern)
-zhe	✓/?	✓	Yes, but only slightly (*-zhe* is misused with achievements and is acquired later than *-le* in monolingual children)

For accomplishments, if HSs and L2ers equate *-zhe* as *-ing*, both groups might misjudge accomplishments with -zhe as grammatical. But since *-zhe* in simple declarative sentences, as tested here, is not very common, it is possible that HSs and L2ers may correctly judge these combinations as ungrammatical simply because they do not know -zhe can be used in simple declarative sentences outside of verbs denoting posture or locations (see (6) in Chapter 2); see Table 5.6. Similarly, with achievements, if HSs and L2ers equate *zai* and *-zhe* as *-ing*, both groups might misjudge the two aspect markers to be compatible with achievements. However, since both *zai* and *-zhe* first emerge with activities before extending to accomplishments, the HS advantage might not be as pronounced as in T3 sandhi (which is acquired by age three) as discussed in Chapter 3. For some HSs, there may be no advantage in rejecting *zai* with achievements if they never acquired the ungrammaticality of *zai* with achievements as children; see Table 5.7.

Table 5.6. Predictions for the AJT conditions on accomplishments (✓= grammatical; ✗= ungrammatical) (non-target-like responses are shaded in gray)

	NSs	HSs and L2ers under English transfer (*-le/-guo = -ed*; *zai/-zhe = -ing*)	Is an HS advantage expected?
-le	✓	✓	No (Mandarin and English have the same pattern)
-guo	✓	✓	No (Mandarin and English have the same pattern)
zai	✓	✓	No (Mandarin and English have the same pattern)
-zhe	✗/?	✓	Yes, but only slightly (*-zhe* is first acquired with activities before extending to other lexical predicates in monolingual children)

Table 5.7. Predictions for the AJT conditions on achievements (✓= grammatical; ✗= ungrammatical) (non-target-like responses are shaded in gray)

	NSs	HSs and L2ers under English transfer (*-le/-guo = -ed*; *zai/-zhe = -ing*)	Is an HS advantage expected?
-le	✓	✓	No (Mandarin and English have the same pattern)
-guo	✓	✓	No (Mandarin and English have the same pattern)
zai	✗	✓	Yes, but only slightly (*zai* is first acquired with activities before extending to achievements in monolingual children)
-zhe	✗	✓	Yes, but only slightly (*-zhe* is first acquired with activities before extending to achievements in monolingual children)

5.3. Results

Data from 62 NSs, 39 HSs, and 36 L2ers were included in the AJT results (see Chapter 4 for participant details). The four lexical aspect predicates were analyzed and reported separately in the following four subsections. Four ordinal mixed regression models (Christensen, 2018), each on one lexical aspect (states, activities, accomplishments, and achievements) were conducted, using the *clmm()* function in R (R Core Team, 2019). The dependent variable was the participants' ratings from 1 to 4. The fixed effects were group (NSs, HSs, and L2ers), aspect marker (*-le*, *-guo*, *zai*, and *-zhe*), and their interaction; the random effects included a random intercept for subjects and a random intercept for items. The reference level for the variable group was the NSs, while the reference level for the variable condition was the experiential marker *-guo*, since it is fully compatible with all lexical aspect predicates. For each lexical aspect, I first created two models (one with interaction and one without) and compared the two models using the *anova()* function. Results showed that including the interaction significantly improved the model. Then, the global effects on condition and group were assessed with the *drop()* function, which I report in the output below. For each model, follow-up pairwise comparisons via the *emmeans()* function were conducted, with a Tukey-adjusted p-value of 0.05. Additional individual analyses were conducted for the three categories of interest, namely states with *-le*, achievements with *zai*, and achievements with *-zhe*, in which English transfer is expected to lead to non-native-like performance.

5.3.1. States

5.3.1.1. Group analysis

Figure 5.1 shows the mean ratings on the four aspect markers by different groups when combined with states. The model output of the ordinal mixed regression model is in Table 5.8. There was a

significant effect of aspect marker which indicates that overall different aspect markers have different degrees of acceptability, with *zai* being rated the lowest, followed by *-zhe*, *-le*, and *-guo* (*zai* < *-zhe* < *-le* < *-guo*, but *-le* and *-guo* are similar). While there is no significant effect of group, the interaction between aspect marker and group is significant: the source of this interaction is that NSs rated *-guo* higher than HSs and L2ers.

Figure 5.1. AJT results: Mean ratings on aspect (States)

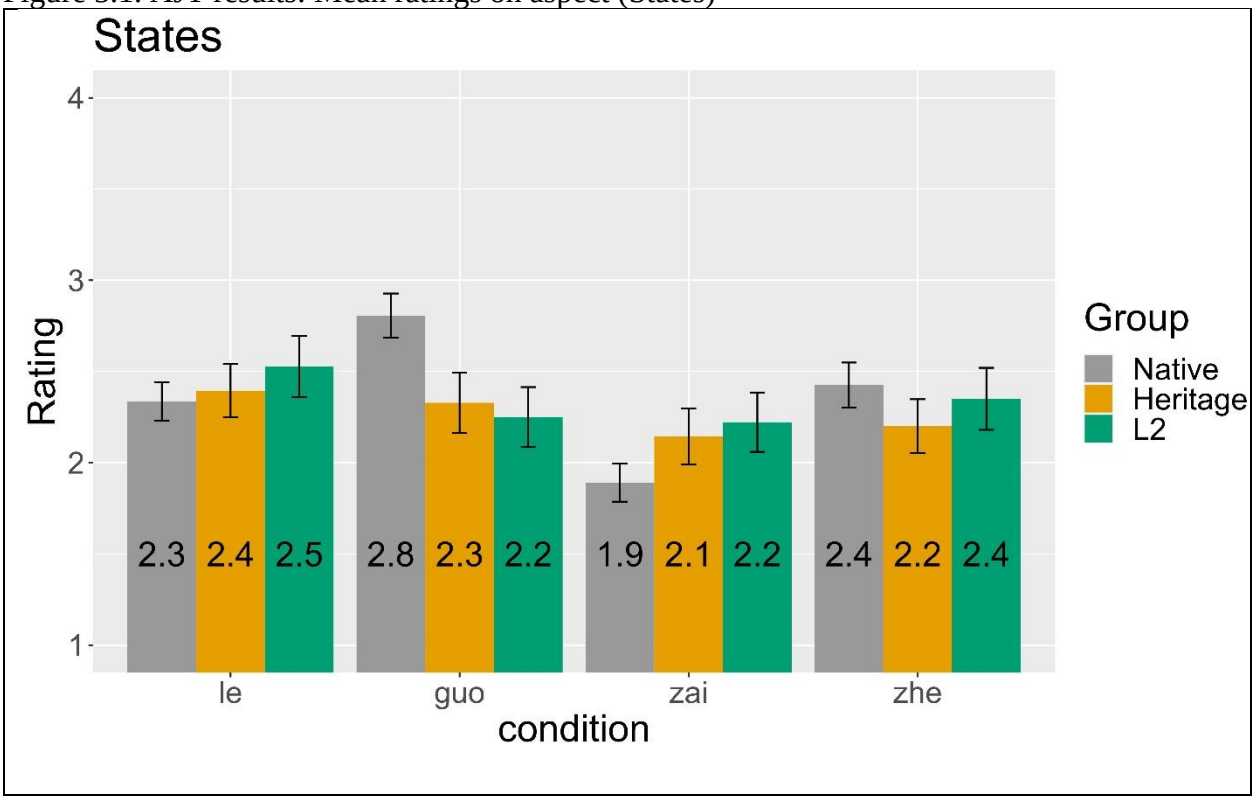


Table 5.8. Results from the ordinal mixed-effects model on aspect (States)

	Df	AIC	LRT	Pr(>Chi)
<none>		6204.6		
condition	3	7338.6	1139.98	< 2.2e-16 ***
Group	2	6212.5	11.93	0.002571 **
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Follow-up pairwise comparisons were made; see Table 5.9. I discuss the between-group differences before the within-group differences. All three groups pattern similarly on *-le*, *zai* and

-zhe, except for *-guo*. NSs rated *-guo* significantly above HSs and L2ers, while the HSs and L2ers patterned similarly. Among different aspect markers within groups, NSs rated *-guo* significantly above the other three aspect markers, while HSs and L2ers rated all aspect markers similarly.

Table 5.9. Pairwise comparison results from the ordinal mixed-effects model on aspect (States)

contrast	estimate	SE	df	z.ratio	p.value	
le,Native - le,Heritage	-0.15642	0.242	Inf	-0.647	1.0000	
le,Native - le,L2	-0.42863	0.251	Inf	-1.707	0.8656	
guo,Native - guo,Heritage	1.11980	0.250	Inf	4.485	0.0005	*
guo,Native - guo,L2	1.20480	0.256	Inf	4.713	0.0002	*
zai,Native - zai,Heritage	-0.54386	0.248	Inf	-2.195	0.5536	
zai,Native - zai,L2	-0.61225	0.255	Inf	-2.404	0.4030	
zhe,Native - zhe,Heritage	0.45986	0.243	Inf	1.889	0.7666	
zhe,Native - zhe,L2	0.29497	0.251	Inf	1.175	0.9908	
le,Heritage - le,L2	-0.27221	0.278	Inf	-0.980	0.9981	
guo,Heritage - guo,L2	0.08500	0.282	Inf	0.301	1.0000	
zai,Heritage - zai,L2	-0.06839	0.282	Inf	-0.243	1.0000	
zhe,Heritage - zhe,L2	-0.16489	0.278	Inf	-0.593	1.0000	
guo,Native - le,Native	1.05493	0.152	Inf	6.940	<.0001	*
guo,Native - zai,Native	2.04982	0.158	Inf	12.935	<.0001	*
guo,Native - zhe,Native	0.81739	0.152	Inf	5.369	<.0001	*
guo,Heritage - le,Heritage	-0.22128	0.195	Inf	-1.134	0.9932	
guo,Heritage - zai,Heritage	0.38616	0.199	Inf	1.938	0.7356	
guo,Heritage - zhe,Heritage	0.15745	0.197	Inf	0.799	0.9997	
guo,L2 - le,L2	-0.57850	0.206	Inf	-2.802	0.1790	
guo,L2 - zai,L2	0.23276	0.208	Inf	1.120	0.9939	
guo,L2 - zhe,L2	-0.09244	0.206	Inf	-0.448	1.0000	
zai,Native - zhe,Native	-1.23242	0.152	Inf	-8.120	<.0001	*
zai,Heritage - zhe,Heritage	-0.22871	0.195	Inf	-1.173	0.9910	
zai,L2 - zhe,L2	-0.32521	0.206	Inf	-1.576	0.9179	

Note: * indicates p<.05

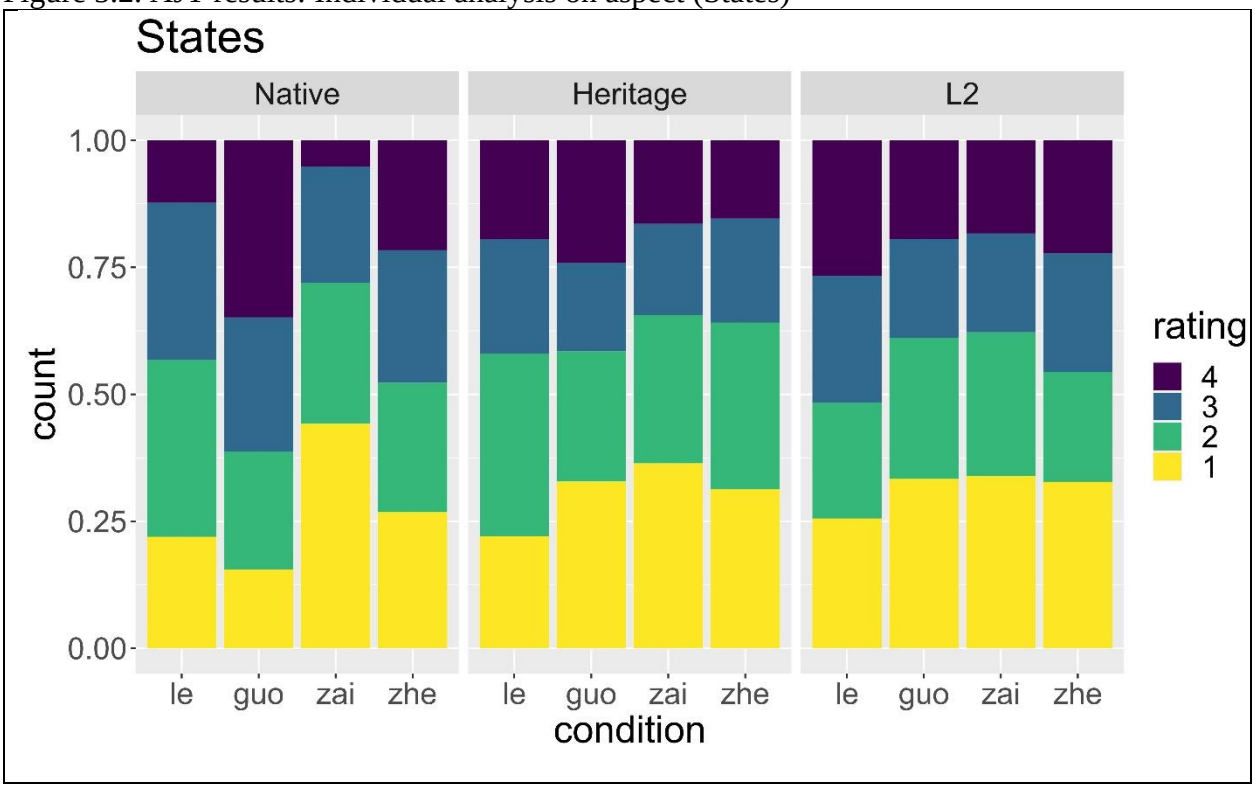
5.3.1.2. Individual subjects' analysis

While overall HSs and L2ers did not accept states with *-le*, an individual subjects' analysis was conducted to examine individual performance. For each individual, I calculated the mean rating for each of the four conditions (*-le, -guo, zai,* and *-zhe*). Given that the AJT used a four-point scale (1 = completely unacceptable; 4 = fully acceptable), the middle-point is 2.5, and ratings

between 2 and 3 may not show strong evidence of rejection or acceptance. Thus, participants who showed acceptance were defined as those who gave a mean rating (out of five tokens) above three (>3) to a given condition. Out of 62 NSs, 39 HSs, and 36 L2ers, 29 participants (ten NSs, five HSs, and 14 L2ers) accepted states with *-le*, with mean ratings of 3.2 for NSs, 3.3 for HSs, and 3.2 for L2ers. In terms of a proficiency effect, the five HSs had a mean score of 21.2 while the 14 L2ers had a mean score of 26.9 (the average scores were 27.1 for HSs and 29.5 for L2ers, see Table 4.1 in Chapter 4).

For completeness, Figure 5.2 below shows the proportion of the raw counts for each rating scale (1, 2, 3, and 4) for each grammatical aspect by each group. The pattern is similar across groups. For *-le*, L2ers gave proportionally more ratings of 4 than NSs and HSs did. This is consistent with the group results in that L2ers show numerically higher ratings on *-le*.

Figure 5.2. AJT results: Individual analysis on aspect (States)



5.3.1. Activities: group analysis

Figure 5.3 shows the mean ratings on how the four aspect markers were rated when combined with activities. The model output of the ordinal mixed regression model is given in Table 5.10. There is a main effect of aspect marker, indicating that *zai* received the highest rating, followed by *-le* and *-guo*, and the lowest *-zhe* (*zai* > *-le* = *-guo* > *-zhe*). There is a main effect of group, with NSs giving overall higher ratings than HSs, which in turn gave higher ratings than L2ers (besides on *-zhe*). There is also a significant interaction between aspect marker and group: the between-group difference is more pronounced in *-le*, *-guo* and *-zhe* than in *zai*.

Figure 5.3. AJT results: Mean ratings on aspect (Activities)

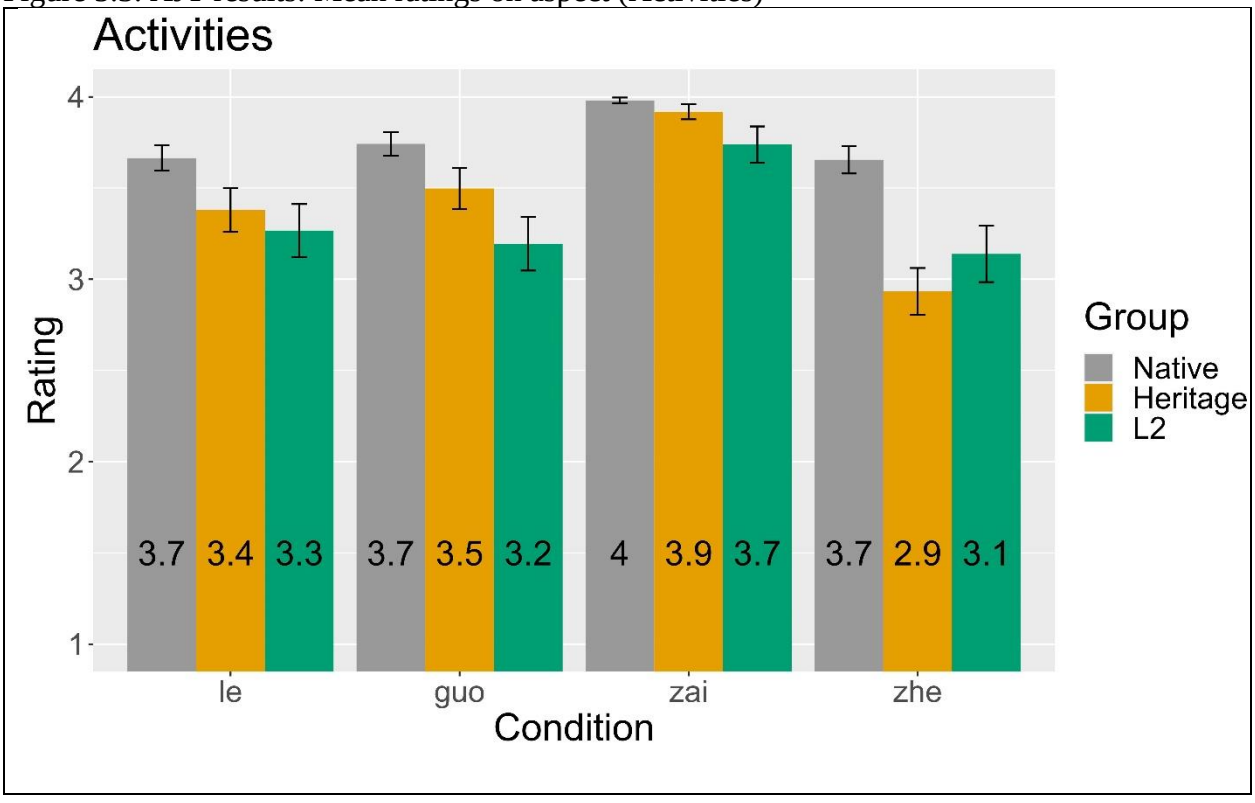


Table 5.10. Results from the ordinal mixed-effects model on aspect (Activities)

	Df	AIC	LRT	Pr(>Chi)	
<none>		4063.2			
condition	3	4403.7	346.53	< 2.2e-16	***
Group	2	4110.4	51.20	7.613e-12	***
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Follow-up pairwise comparisons are shown in Table 5.11. I start with the between-group differences before discussing the within-group differences. NSs rated *-le, -guo* and *-zhe* significantly higher than HSs. When compared with L2ers, NSs rated significantly higher than them in all four conditions. HSs and L2ers patterned similarly in all four conditions. For all three groups, the ratings on *-guo* were significantly below *zai*, but *-guo* and *-le* had similar ratings. While NSs and L2ers also rated *-guo* and *-zhe* similarly, HSs rated *-guo* significantly above *-zhe*.

To assess the incompleteness effect of *-zhe,* I additionally compared how participants rated *zai* vs. *-zhe,* to compare with the finding in Y. Guo (2020) which used a sentence completeness judgment task. All three groups rated *zai* significantly above *-zhe*; see the last three rows in Table 5.11. To check if *zai* was overall rated higher than *-zhe* in all lexical predicates, I compared how participants rated *zai* and *-zhe* in all four lexical predicates and report them here. NSs rated *-zhe* and *zai* similarly in achievements, but rated-*zhe* significantly higher than *zai* in states, significantly lower than *zai* in activities and accomplishments. HSs rated *zai* and *-zhe* similarly in states and achievements, but rated *zai* significantly higher than *-zhe* in activities and accomplishments. L2ers rated *zai* and *-zhe* similarly in states, accomplishments, and achievements, but rated *zai* significantly higher than *-zhe* in activities. Thus, the significantly lower ratings of *-zhe* with activities (as compared to *zai*) among the three groups is not an overall preference for *zai,* but may be an indication that they are sensitive to the incompleteness effect, or at least know that *-zhe* is not as perfectly compatible with activities as *zai* is.

Table 5.11. Pairwise comparison results from the ordinal mixed-effects model on aspect (Activities)

contrast	estimate	SE	df	z.ratio	p.value	
le,Native - le,Heritage	1.0253	0.292	Inf	3.510	0.0228	*
le,Native - le,L2	1.2812	0.299	Inf	4.290	0.0011	*
guo,Native - guo,Heritage	1.0043	0.304	Inf	3.301	0.0452	*
guo,Native - guo,L2	1.8217	0.305	Inf	5.972	<.0001	*
zai,Native - zai,Heritage	1.5649	0.550	Inf	2.847	0.1607	
zai,Native - zai,L2	2.6864	0.519	Inf	5.175	<.0001	*
zhe,Native - zhe,Heritage	2.1106	0.286	Inf	7.371	<.0001	*
zhe,Native - zhe,L2	1.5313	0.298	Inf	5.143	<.0001	*
le,Heritage - le,L2	0.2559	0.316	Inf	0.809	0.9997	
guo,Heritage - guo,L2	0.8174	0.319	Inf	2.560	0.3028	
zai,Heritage - zai,L2	1.1215	0.423	Inf	2.651	0.2514	
zhe,Heritage - zhe,L2	-0.5794	0.305	Inf	-1.899	0.7602	
guo,Native - le,Native	0.3891	0.209	Inf	1.859	0.7848	
guo,Native - zai,Native	-2.7564	0.449	Inf	-6.136	<.0001	*
guo,Native - zhe,Native	0.3637	0.212	Inf	1.719	0.8600	
guo,Heritage - le,Heritage	0.4102	0.220	Inf	1.865	0.7808	
guo,Heritage - zai,Heritage	-2.1957	0.326	Inf	-6.732	<.0001	*
guo,Heritage - zhe,Heritage	1.4701	0.212	Inf	6.948	<.0001	
guo,L2 - le,L2	-0.1514	0.219	Inf	-0.692	0.9999	
guo,L2 - zai,L2	-1.8916	0.263	Inf	-7.194	<.0001	*
guo,L2 - zhe,L2	0.0733	0.215	Inf	0.341	1.0000	
zai,Native - zhe,Native	3.1201	0.446	Inf	6.989	<.0001	*
zai,Heritage - zhe,Heritage	3.6658	0.320	Inf	11.466	<.0001	*
zai,L2 - zhe,L2	1.9650	0.262	Inf	7.497	<.0001	*

Note: * indicates p<.05

5.3.2. Accomplishments: group analysis

Figure 5.4 presents the mean ratings on the four aspect markers for accomplishments. The model output of the ordinal mixed regression model is given in Table 5.12. Similar to the model output on activities, the main effects on the aspect marker and group are both significant. The overall ratings for *-le* is higher than *-guo*, *zai*, and *-zhe*, in that order (*-le* > *-guo* > *zai* > *-zhe*). Overall, NSs gave higher ratings than HSs, which in turn gave higher ratings than L2ers. The interaction between aspect marker and group is also significant: the source of this interaction is

that the differences between NSs vs. HSs/L2ers are more pronounced in *-le* and *-guo* than in other aspect markers.

Figure 5.4. AJT results: Mean ratings on aspect (Accomplishments)

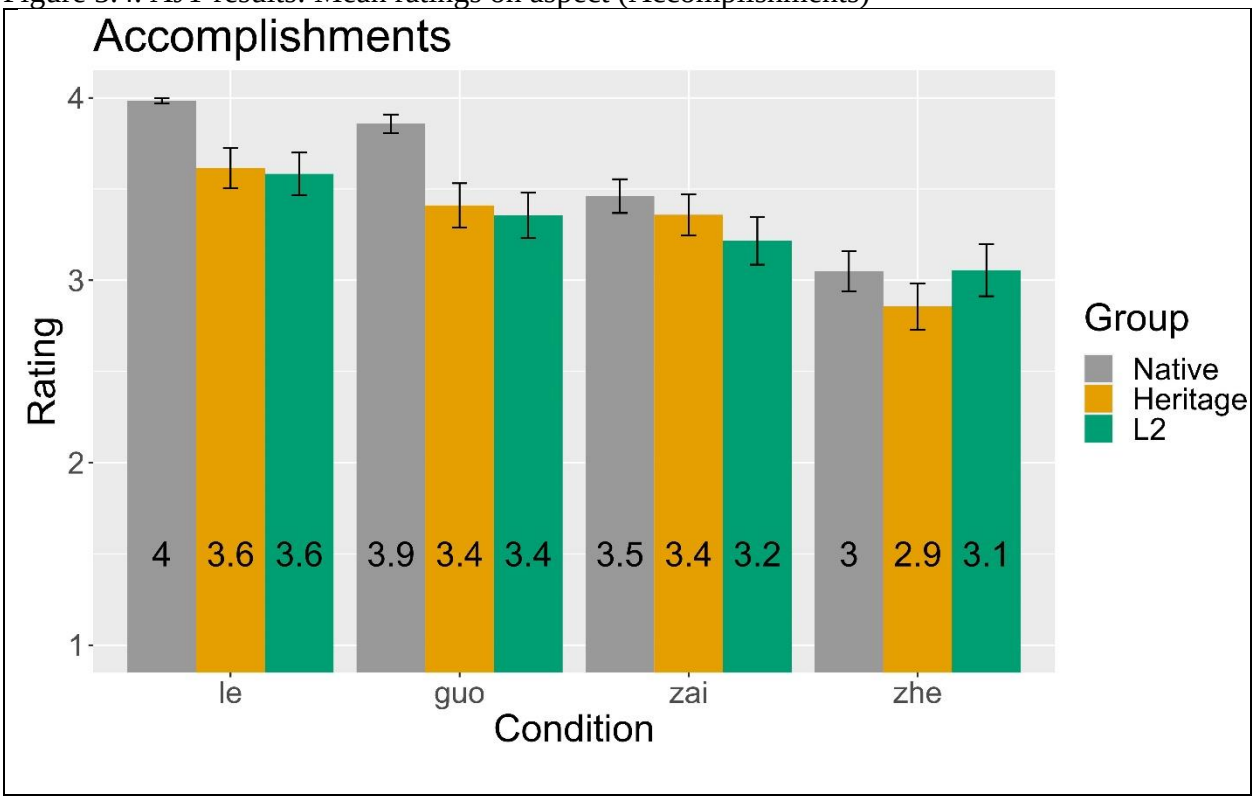


Table 5.12. Results from the ordinal mixed-effects model on aspect (Accomplishments)

	Df	AIC	LRT	Pr(>Chi)
<none>		4742.9		
condition	3	5229.6	492.64	< 2.2e-16 ***
Group	2	4767.8	28.87	5.378e-07 ***
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Follow-up pairwise comparisons are summarized in Table 5.13. I again discuss between-group differences before the within-group differences. While all groups patterned similarly in *zai* and *-zhe*, NSs rated *-le* and *-guo* significantly higher than HSs and L2ers did. HSs and L2ers patterned similarly on all four conditions. For all three groups, the ratings on *-guo* is significantly

higher than *-zhe*. While NSs also rated *-guo* significantly higher than *zai*, HSs and L2ers rated both *-guo* and *zai* similarly. Despite minimal numerical differences, the ratings for *-le* and *-guo* by NSs were significantly different from each other. The ratings for *-le* and *-guo* by both HSs and L2ers were marginally significant.

Table 5.13. Pairwise comparison results from the ordinal mixed-effects model on aspect (Accomplishments)

contrast	estimate	SE	df	z.ratio	p.value	
le,Native - le,Heritage	3.4765	0.543	Inf	6.404	<.0001	*
le,Native - le,L2	3.7393	0.544	Inf	6.871	<.0001	*
guo,Native - guo,Heritage	1.9957	0.317	Inf	6.302	<.0001	*
guo,Native - guo,L2	2.2350	0.319	Inf	7.001	<.0001	*
zai,Native - zai,Heritage	0.3751	0.267	Inf	1.404	0.9630	
zai,Native - zai,L2	0.6428	0.271	Inf	2.372	0.4253	
zhe,Native - zhe,Heritage	0.5067	0.255	Inf	1.991	0.7000	
zhe,Native - zhe,L2	-0.0170	0.264	Inf	-0.064	1.0000	
le,Heritage - le,L2	0.2628	0.330	Inf	0.797	0.9997	
guo,Heritage - guo,L2	0.2392	0.304	Inf	0.786	0.9998	
zai,Heritage - zai,L2	0.2677	0.295	Inf	0.907	0.9991	
zhe,Heritage - zhe,L2	-0.5237	0.289	Inf	-1.809	0.8133	
guo,Native - le,Native	-2.2454	0.511	Inf	-4.397	0.0007	*
guo,Native - zai,Native	1.9420	0.235	Inf	8.257	<.0001	*
guo,Native - zhe,Native	3.0442	0.233	Inf	13.049	<.0001	*
guo,Heritage - le,Heritage	-0.7646	0.235	Inf	-3.255	0.0521	
guo,Heritage - zai,Heritage	0.3213	0.211	Inf	1.525	0.9341	
guo,Heritage - zhe,Heritage	1.5552	0.207	Inf	7.520	<.0001	*
guo,L2 - le,L2	-0.7410	0.236	Inf	-3.146	0.0721	
guo,L2 - zai,L2	0.3498	0.212	Inf	1.648	0.8912	
guo,L2 - zhe,L2	0.7922	0.213	Inf	3.711	0.0112	*
zai,Native - zhe,Native	1.1022	0.165	Inf	6.667	<.0001	*
zai,Heritage - zhe,Heritage	1.2339	0.198	Inf	6.237	<.0001	*
zai,L2 - zhe,L2	0.4424	0.207	Inf	2.142	0.5919	

Note: * indicates p<.05

5.3.3. Achievements

5.3.3.1. Group analysis

Figure 5.5 shows the mean ratings on the four aspect markers when combined with achievements. The model output of the ordinal mixed regression model is in Table 5.14. Similar to the model output on activities and accomplishments, the main effects on the aspect marker and group are significant. The overall ratings for *-le* is higher than *-guo*, which in turn is much higher than *zai and zhe* (*-le* > *-guo* > *zai* > *-zhe*), with *zai* being slightly higher than *-zhe*. Overall, with states and activities, NSs gave higher ratings than HSs and L2ers, but for accomplishments and achievements, NSs gave lower ratings than HSs, which in turn gave lower ratings than L2ers. NSs typically gave either high or low ratings while HSs and L2ers gave more intermediate ratings. The interaction between aspect markers and group is also significant: the source of this interaction is that the NSs rated *-le* and *-guo* higher than and *zai* and *-zhe* lower than HSs/L2ers did.

Figure 5.5. AJT results: Mean ratings on aspect (Achievements)

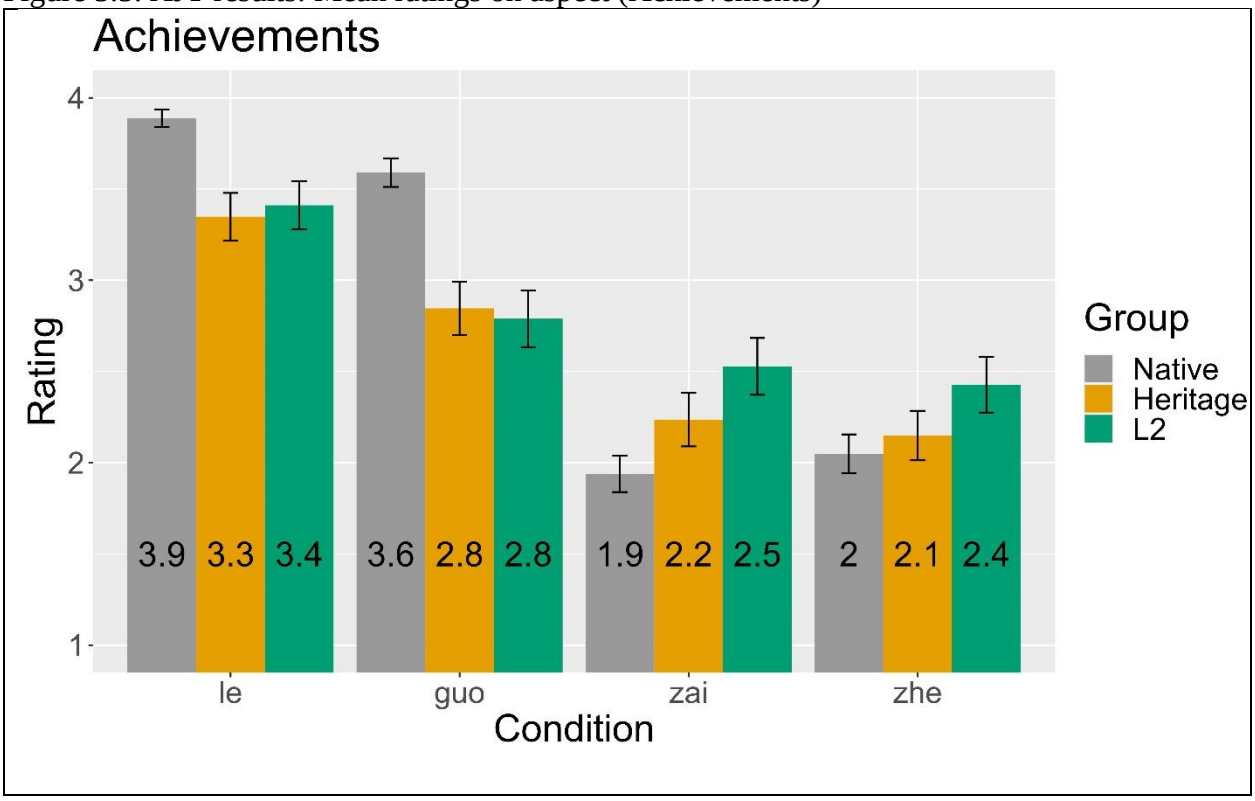


Table 5.14. Results from the ordinal mixed-effects model on aspect (Achievements)

	Df	AIC	LRT	Pr(>Chi)	
<none>		6204.6			
condition	3	7338.6	1139.98	< 2.2e-16	***
Group	2	6212.5	11.93	0.002571	**
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Follow-up pairwise comparisons were given in Table 5.15. The between-group differences are again discussed before within-group differences. On *-le* and *-guo*, NSs rated significantly higher than HSs and L2ers did. NSs rated significantly lower than L2ers (but not HSs) on *zai* and rated marginally below L2ers (but not HSs) on *-zhe.* HSs and L2ers patterned similarly on all four conditions. The ratings of NSs and HSs on *-guo* differ significantly from the other three aspect markers: compared to *-guo*, they have higher acceptance of *-le* and lower acceptance of *zai* and *-zhe*. For L2ers, they also rated *-guo* significantly below *-le* and significantly above *-zhe*, but the ratings on *-guo* and *zai* were similar.

Table 5.15. Pairwise comparison results from the ordinal mixed-effects model on aspect (Achievements)

contrast	estimate	SE	df	z.ratio	p.value	
le,Native - le,Heritage	2.5232	0.312	Inf	8.075	<.0001	*
le,Native - le,L2	2.3199	0.319	Inf	7.262	<.0001	*
guo,Native - guo,Heritage	2.0128	0.247	Inf	8.149	<.0001	*
guo,Native - guo,L2	2.0473	0.252	Inf	8.132	<.0001	*
zai,Native - zai,Heritage	-0.5901	0.230	Inf	-2.566	0.2991	
zai,Native - zai,L2	-1.1798	0.236	Inf	-4.994	<.0001	*
zhe,Native - zhe,Heritage	-0.2920	0.228	Inf	-1.279	0.9818	
zhe,Native - zhe,L2	-0.7596	0.235	Inf	-3.235	0.0553	
le,Heritage - le,L2	-0.2033	0.280	Inf	-0.725	0.9999	
guo,Heritage - guo,L2	0.0344	0.261	Inf	0.132	1.0000	
zai,Heritage - zai,L2	-0.5896	0.260	Inf	-2.264	0.5028	
zhe,Heritage - zhe,L2	-0.4676	0.259	Inf	-1.808	0.8141	
guo,Native - le,Native	-1.8210	0.259	Inf	-7.028	<.0001	*
guo,Native - zai,Native	3.7534	0.182	Inf	20.622	<.0001	*
guo,Native - zhe,Native	3.5527	0.181	Inf	19.666	<.0001	*
guo,Heritage - le,Heritage	-1.3106	0.203	Inf	-6.455	<.0001	*

Table 5.15. Pairwise comparison results from the ordinal mixed-effects model on aspect (Achievements) (cont'd)

contrast	estimate	SE	df	z.ratio	p.value	
guo,Heritage - zai,Heritage	1.1504	0.191	Inf	6.009	<.0001	*
guo,Heritage - zhe,Heritage	1.2479	0.190	Inf	6.552	<.0001	*
guo,L2 - le,L2	-1.5484	0.215	Inf	-7.214	<.0001	*
guo,L2 - zai,L2	0.5263	0.197	Inf	2.669	0.2418	
guo,L2 - zhe,L2	0.7459	0.197	Inf	3.780	0.0086	*
zai,Native - zhe,Native	-0.2006	0.147	Inf	-1.365	0.9701	
zai,Heritage - zhe,Heritage	0.0975	0.188	Inf	0.520	1.0000	
zai,L2 - zhe,L2	0.2196	0.196	Inf	1.122	0.9938	

Note: * indicates p<.05

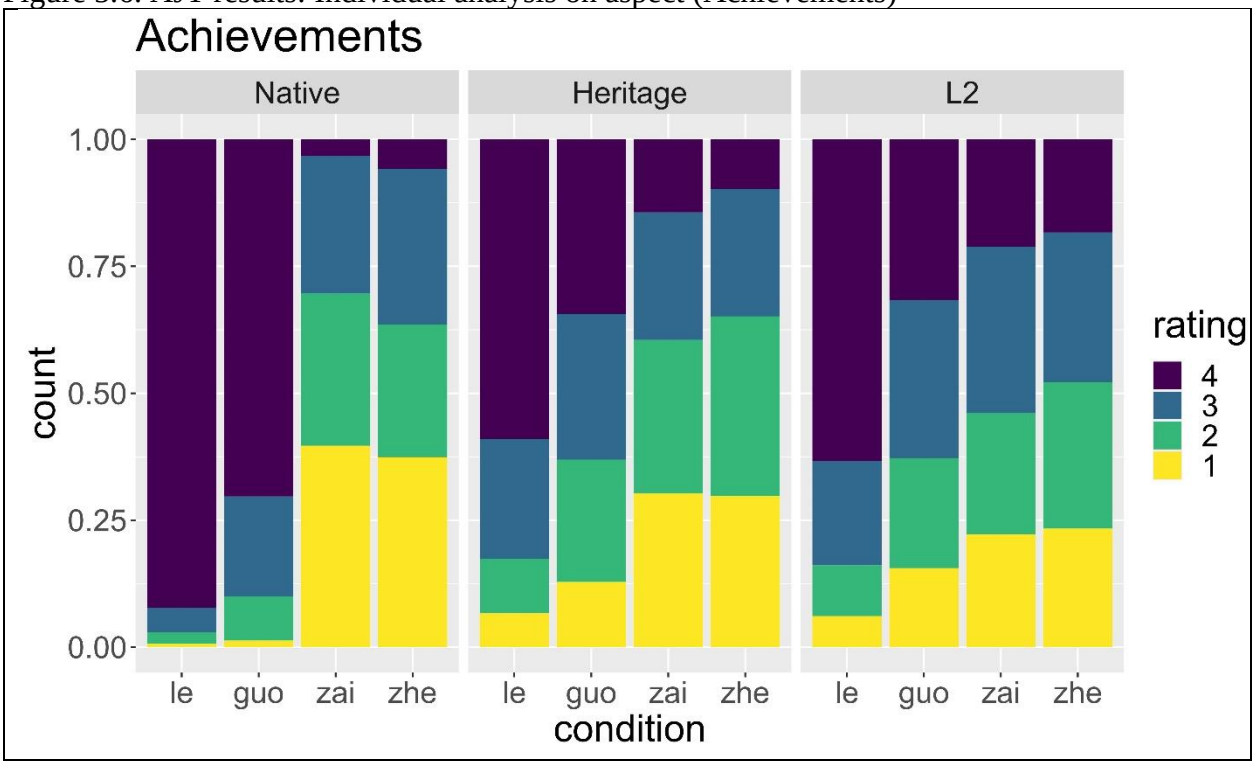
5.3.3.2. Individual subjects' analysis

Compared to NSs, L2ers over-rated achievements with *zai* and, to a lesser degree, *-zhe.* HSs did not do this. Similar to the individual subjects' analysis conducted for states, an individual subjects' analysis was conducted for achievements to assess individual performance. The coding procedure was the same. For each individual, a mean rating for each of the four conditions was calculated. Participants who gave a mean rating above three (>3) to a given condition (out of five tokens) were considered to show acceptance. Out of 62 NSs, 39 HSs, and 36 L2ers, 19 participants (seven NSs, three HSs, and nine L2ers) accepted achievements with *zai*, with mean ratings of 3 for NSs, 3.2 for HSs, and 3.1 for L2ers. There were 21 participants who accepted achievements with *-zhe*, including eight NSs, four HSs, and nine L2ers; the mean ratings were 3 for NSs, 3.2 for HSs, and 3.1 for L2ers. In terms of proficiency effect, the three HSs had a mean score of 19.3 while the nine L2ers had a mean score of 26.2 (again, the average scores are 27.1 for HSs and 29.5 for L2ers; see Table 4.1 in Chapter 4).

As Figure 5.6 shows, NSs tended to rate *-le* and *-guo* as 4 and *zai* and *-zhe* below 4. On both *zai* and *-zhe*, L2ers had proportionally more counts for ratings of 4 than NSs and L2ers. This was

consistent with the group results, in which the mean ratings from L2ers were higher than those from NSs and HSs.

Figure 5.6. AJT results: Individual analysis on aspect (Achievements)



5.4. Discussion

Recall that the RQs asked whether proficiency-matched HSs and L2ers acquire language phenomena that are absent in or different from English, and whether there is an HS advantage. The specific RQs asked whether they could correctly accept only grammatical combinations (interactions between grammatical and lexical aspect) tested in the AJT. I discuss some unexpected results from NSs before non-native like performances from HSs and L2ers. In Table 5.1, there are nine grammatical conditions (one is marginal with a "?" mark) and seven ungrammatical conditions (two with "?" marks). As Figure 5.1, Figure 5.3, Figure 5.4, and Figure 5.5 show, among

the nine grammatical conditions, eight are rated at 3.5 and above while one was rated unexpectedly low at 2.8: states with *-guo*. Among the seven ungrammatical conditions, five were rated at 2.4 and below. The remaining two ungrammatical conditions were the two with marginal acceptability (marked with "?"): activities with *-le* rated at 3.7 and accomplishments with *-zhe* at 3. I discuss the under-ratings of the grammatical condition (states with *-guo*) before the over-ratings of two ungrammatical or marginal conditions (activities with *-le*; accomplishments with *-zhe*).[37]

Recall that *-guo* is an experiential marker, but the previous state no longer exists. According to R. Xiao and McEnery (2004b, p. 143), *-guo* is compatible with all four lexical aspect predicates, though it does not occur very often with states (e.g., *you* 'exist'). A corpus study in Mandarin conducted by R. Xiao and McEnery (2004b, p. 143, Table 4.6) revealed that less than 10 % of states co-occur with *-guo* (8.33 % of individual-level states and 1.19% of stage-level states), compared to 29.76% of activities, 42.86% of accomplishments, and 16.67% of achievements. (The sum will be 100% after adding 1.19% for semelfactives, which is not discussed in this dissertation.) In the AJT, while the states with *-guo* condition was only rated at 2.8, NSs still rated it significantly above the other three aspect markers, which were ungrammatical conditions. (HSs and L2ers rated all aspect markers with states similarly.) The lowered acceptability is probably due to the fact that many state verbs do not typically take any aspect markers at all. Specifically, in the AJT stimuli, the lowest four ratings (out of 20 token sets) by NSs were all on the verb *xiang* 'resemble' with different objects (ranging from 1.6 to 2.2). While I included *xiang* 'resemble' as an individual-level state verb following R. Xiao and McEnery (2004b), Tai (1982, cited in Ross, 1991, p. 84)

[37] While previous studies on dialect differences on aspect do not address the combinations tested here, it is a possible explanation worthy of future investigation. Using a corpus method, Khoo and Lin (2018) examine the dialect differences on aspect marking in Singaporean, mainland, and Taiwanese Mandarin. While the focus is not on dialect, J. Li and Hsieh (2015, footnote 4, 5, and 22) report that some documented usage of *-zhe* was not found in their corpus study using Taiwanese Mandarin or was rejected by consultants who speak Taiwanese Mandarin.

actually uses this verb as an example that does not take aspectual suffixes *-le*, *-guo*, and *-zhe*. Thus, the generalization that individual-level states are compatible with *-guo*, as claimed by R. Xiao and McEnery (2004b), might need further investigation.[38]

Next, I turn to the two conditions that were originally classified as ungrammatical (though with "?" marks) but received ratings above 3: activities with *-le* rated at 3.7 and accomplishments with *-zhe* at 3. In the literature, accomplishments with *-zhe* is not consistently classified as grammatical or ungrammatical. In the corpus search in R. Xiao and McEnery (2004b, p. 188), less than 1% of the predicates marked with *-zhe* were accomplishments. I do not have a ready explanation as to why NSs rated accomplishments with *-zhe* at 3 despite their very low frequency. I turn to activities with *-le* and the associated incompleteness effect. Recall that activities with *-le* and *-zhe* (in simple declarative sentences) are sometimes considered incomplete sentences. I originally classified activities with *-le* as ungrammatical since the incompleteness effect is well documented, but classified activities with *-zhe* as grammatical since very few scholars have mentioned its incompleteness. Since both conditions were rated at 3.7 by NSs, they (as well as HSs and L2ers) clearly treated such so-called incomplete sentences as acceptable, at least in an out-of-context sentence-level AJT like the present one. It is possible that an AJT is not sensitive enough to elicit knowledge of the incompleteness effect, unlike Y. Guo (2020). Y. Guo (2020) tested this incompleteness effect of *-zhe* with activities by explicitly asking whether the sentences are complete or not after presenting the sentences word-by-word. She found that only NSs, but not L1-English L2-Mandarin learners, correctly judged that only *-zhe*, but not *zai*, exhibit such

[38] Given that Mandarin lacks (overt) tense and 'resemble' does not normally take an aspect marker, the Mandarin sentences for "*He resembled his father in childhood"* (past tense in English) vs. "*He resembles his father"* (present tense in English) differ only in the time phrase *in childhood*, but not in tense nor aspect.

incompleteness effect with activities. In this dissertation, all three groups rated *zai* significantly above *-zhe* with activities, indicating that they may know the incompleteness effect of *-zhe.*

Next I discuss results from HSs and L2ers, focusing on the three critical categories where English transfer was expected. Given that all three groups pattern similarly on *-le* with states, there is no evidence of English transfer of *-ed* by HSs and L2ers. For *zai* and *-zhe* with achievements, NSs pattern with HSs, but differ significantly from L2ers on *zai* and marginally on *-zhe*. Compared to *-le* and *-guo* with achievements, L2ers did give lower ratings for *zai* (at 2.5) and *-zhe* (at 2.4), but not as low as those by NSs and HSs. The higher ratings by L2ers, as compared to NSs and HSs, can be attributed to English transfer of *-ing*, consistent with prior findings (L. Jin & Hendriks, 2005; Y. Guo, 2020). Despite the same ratings by L2ers on *-le* with states and *zai* with achievements (both at 2.5), L2ers patterned with NSs on the former but differed significantly from NSs on the latter. If the statistical results are replicated with more L2 participants, the question then becomes why English transfer appears with progressive aspect but not for past tense. Given that *-le* is more frequent than *zai* (e.g., J. Chen & Shirai, 2010, based on parents speech in an L1 acquisition study; L. Jin & Hendriks, 2005, based on story-telling), L2ers might have received more corrections on *-le* than *zai* (also because *-le* shares the same form as the sentence-final particle *LE*). For HSs, another reason may be that *-le* is the first-acquired aspect marker, which helps adult HSs to judge the ungrammaticality of states with *-le*. While I do not focus on the Aspect Hypothesis due to lack of different proficiency groups (recall that the hypothesis focuses on emergence and makes predictions on developmental trajectory), the fact that L2ers were more native-like on *-le* than *zai* is compatible with Prediction B in the Aspect Hypothesis, which states that the perfective is acquired earlier than the imperfective.

While HSs and L2ers did not differ significantly on any conditions, HSs are more native-like than L2ers on *-zhe* with achievements (no differences between NSs and HSs, and marginal differences between NSs and L2ers) as well as activities with *-guo* (only marginal differences between NSs and HSs, and significant differences between NSs and L2ers). (The HS advantage on activities with *-guo* is not expected since *-guo* is the last-acquired aspect marker, and if HSs and L2ers map *-guo* with *-ed*, they would simply give high ratings for *-guo* with all lexical predicates.) This HS advantage is also found in Shi (2013), who only focuses on *-le*. The present dissertation shows that, even with proficiency controlled for, early exposure gives HSs a slight advantage on the early-acquired aspectual marking. The HS advantage cannot be a result of proficiency since HSs and L2ers were proficiency-matched. In fact, the individual subjects' analyses show that, for HSs and L2ers who have non-target-like patterns, L2ers had numerically higher proficiency scores than HSs.

In terms of frequency, recall that aspect markers occur in less than 40% of the predicates in elicited production (L. Jin & Hendriks, 2005) and even less in recorded conversations (less than 30% in child-directed speech and less than 5% in adult-to-adult speech in C.-C. Huang, 2006). Thus, Mandarin aspect markers are not as frequent as they may be in languages where tense/aspect is obligatory on all verbs. The uneven distribution of aspect markers on different lexical predicates makes its acquisition more challenging. I compare these results with the corpus frequency reported in R. Xiao & McEnery (2004b) based on newspaper texts. On *-le*, all three groups gave high ratings (>3) on activities, accomplishments, and achievements, while *-le* is heavily associated with achievement (50%) and accomplishments (30%) and less so with activities (13%). On *-guo*, all three groups gave high ratings (>3) on activities and accomplishments, but only NSs rated achievements above 3, consistent with the corpus study where *-guo* was heavily associated with

accomplishments (43%) and activities (30%), but less so with achievements (17%). On *zai*, all three groups gave high ratings (>3) on activities and accomplishments, but in the corpus, *zai* was heavily associated with activities (83%), but not with accomplishments (9%). On *-zhe*, all three groups gave relatively high ratings (>=2.9) on activities and accomplishments, but in the corpus, while *-zhe* is heavily associated with activities (55%), less than 1% is associated with accomplishments. It seems that the distribution frequency (in a written corpus) is largely, but not entirely, consistent with the ratings found in the present dissertation, which may be due to the specific genre of the corpus. That processing *-le* with accomplishments and *zai* with activities is easier for NSs (Yap et al., 2004, cited in Yap et al., 2009) seems to be consistent with the distribution frequency. As only an offline AJT is employed in the present dissertation, I leave processing considerations for future research with online methodologies.

Returning to the broad RQs, HSs show a slight HS advantage over L2ers in aspect in the domain of morpho-semantics, given that HSs were more native-like than L2ers when they were each compared to NSs. While the performances of HSs and L2ers are still not completely native-like, both groups did show sensitivity to the interaction between grammatical and lexical aspect. However, compared to HSs, L2ers seem to be more subject to dominant language transfer from English *-ing*.

Chapter 6. Judgments and interpretation of relative clauses in HS/L2 Mandarin

This chapter reports on relative clauses (RCs) tested in the Acceptability Judgement Task (AJT) (the same AJT that tested aspect, see Chapter 4 for the overall design and Chapter 5 for aspect) and the Truth Value Judgement Task (TVJT) (see Chapter 4 for the overall design). Recall that the broad research questions (RQs) focus on (in)complete acquisition and selective HS advantages, repeated below. RQs 1c and 2c, also listed below, are specifically for RCs and are instantiations of the broad RQs 1 and 2. Given that Mandarin RCs are acquired before age five in monolingual children and are in the domain of syntax (word order), both HSs and L2ers are expected to acquire RCs, despite differences between Mandarin and English. Thus, no HS advantage is expected. In the tasks employed here, I specifically tested whether participants know that Mandarin RCs are head-final (in the AJT) and whether they could correctly interpret Mandarin RCs (in the TVJT). In addition, both SRCs and ORCs were tested to examine whether participants have an SRC advantage. I describe the results from both tasks before a general discussion on RCs.

- Broad RQ 1: Can HSs and L2ers of Mandarin whose dominant language is English fully acquire the properties of Mandarin that are different from or absent in English?
- Broad RQ 2: Do HSs have selective advantages over proficiency-matched L2ers, and does this vary by linguistic domain?
- RQ 1c: Can HSs and L2ers acquire the head-final property of Mandarin RCs which is different from English?
- RQ 2c: Do HSs have an advantage over L2ers in RCs in the domain of (narrow) syntax?
- Additional RQ for RCs: Is there SRC/ORC asymmetry?

6.1. AJT on RCs

The goal of the AJT is to identify whether participants know that Mandarin RCs are head-final, which is a precondition to determine whether participants have correct interpretations in the TVJT. Given that both SRCs and ORCs were tested, I also compared SRCs vs. ORCs. However, this AJT

is not ideal for examining the SRC advantage (higher ratings of SRCs than ORCs in the context of this task), since the animacy of the heads differs between SRCs and ORCs, as discussed below.

6.1.1. Conditions

Four conditions were created in a 2 (RC type) x 2 (RC headedness) design, creating two grammatical conditions and two ungrammatical conditions; see Table 6.1.[39] On a 1-to-4 rating scale, the expected ratings are 'fully acceptable'' (rating: 4) and 'completely unacceptable' (rating: 1). With five tokens per condition, there were 20 target sentences per participant. There were 20 token sets (VPs) distributed across four lists.

Table 6.1. Conditions testing RCs in the AJT (✓= grammatical; ✗= ungrammatical)

	head-final	Head-initial
SRC	✓ [____chī píngguǒ *de*] **nánhái** 吃蘋果的**男孩** [____ eat apple *de*] **boy**	✗**nánhái** [____chī píngguǒ *de*]**男孩**吃蘋果的 **boy** [____eat apple *de*]
ORC	✓ [____nánhái chī *de*] **píngguǒ** 男孩吃的**蘋果** [____ boy eat *de*] **apple**	✗**píngguǒ** [____nánhái chī *de*]**蘋果**男孩吃的 **apple** [____ boy eat *de*]

The target sentence frames are provided in (21). The head nouns are common nouns and the RCs modify the matrix subjects. Note that the head nouns of ORC are always inanimate; due to the animacy constraint, the action is non-reversible. Using inanimate head nouns ensures that the participants correctly analyze the RC as SRC vs. ORC; since this task addresses headedness rather than interpretation, it is important for the RC to be unambiguous.

[39] I refer to head-initial Mandarin RCs as ungrammatical given their low frequency and marked status, especially in the written form, though head-initial RCs have occasionally been mentioned in the literature. Using a spoken corpus, F. Wang and F. Wu (2020) found that head-initial Mandarin RCs exist in spoken Mandarin, but are rare ("non-canonical") and mostly used as "afterthoughts" (one piece of evidence being that a pause often occurs between the head noun and the head-initial RC). Note that the "head-initial RCs" defined here are classified as appositive RCs by some scholars (e.g., Chao, 1968, cited in F. Wang & F. Wu, 2020, footnote 4). They found that, in both head-final and head-initial Mandarin RCs, SRCs are more frequent than ORCs. Head-initial Mandarin tend to modify sentential objects in spoken Mandarin, but for head-final Mandarin RCs, the tendency to modify sentential subjects or sentential objects seems to depend on genres.

(21) Four sentence types of RCs in the AJT, with examples

a) Grammatical head-final SRC:
[____ Verb-Object *de*] **HeadNoun** Predicate
[____ chī píngguǒ *de*] **nánhái** hěn gāo [吃蘋果的]**男孩**很高
[____ eat apple *de*] **boy** is tall
'**The boy** [that eats the apple] is tall.'

b) Grammatical head-final ORC:
[Subject-Verb ____ *de*] **HeadNoun** Predicate
[nánhái chī ____ *de*] **píngguǒ** hěn tián [男孩吃的]**蘋果**很甜
[boy eat ____ *de*] **apple** is sweet
'**The apple** [that the boy eats] is sweet'.

c) Ungrammatical head-initial SRC:
HeadNoun [____ Verb-Object *de*] Predicate
***nánhái** [____ chī píngguǒ *de*] hěn gā **男孩**[吃蘋果的]很高
boy [____ eat apple *de*] is tall
Intended meaning: '**The boy** [that eats an apple] is tall.'

d) Ungrammatical head-initial ORC:
HeadNoun [Subject-Verb _____ *de*] Predicate
***píngguǒ** [nánhái chī ____ *de*] hěn tián **蘋果**[男孩吃的]很甜
apple [boy eat ____ *de*] is sweet
Intended meaning: '**The apple** [that the boy eats] is sweet'.

6.1.2. Predictions

There are two possible misanalyses by HSs and L2ers. Due to English transfer of RC headedness, HSs and L2ers may incorrectly accept head-initial Mandarin RCs, and incorrectly reject head-final Mandarin RCs ("head error"); see prediction 1 in Table 6.2. Thus, if they consider all four conditions correct, there is no clear evidence of English transfer of head direction. Alternatively, if learners do not know Mandarin RCs and just rely on a Noun-Verb-Noun (NVN) word order strategy, they would consider head-initial SRC and head-final ORC grammatical, since both have a NVN sequence (ignoring the marker *de*); see prediction 2 in Table 6.2. Under such a word order strategy, learners will have an ORC advantage on head-final RCs, but an SRC advantage on head-initial RCs. If divergent RC advantages are found in the same task, it will provide evidence questioning the apparent ORC advantage reported in previous L2 studies that only tested head-final RCs.

Table 6.2. Predictions for the AJT conditions on RCs (✓= grammatical; ✗= ungrammatical) (non-target-like responses are shaded in gray)

	NSs	Prediction 1: HSs and L2ers under English transfer (RC headedness)	Prediction 2: HSs and L2ers under NVN word order strategies
Head-final SRC: [$_{RC}$ VNde]N	✓	✗	✗
Head-final ORC: [$_{RC}$ NVde]N	✓	✗	✓
Head-initial SRC: N[$_{RC}$ VNde]	✗	✓	✓
Head- initial ORC: N[$_{RC}$ NVde]	✗	✓	✗

6.1.3. Results: group analysis

Data from 62 NSs, 39 HSs, and 36 L2ers were included in the AJT results (see Chapter 4 for participant details). The mean ratings are shown in Figure 6.1. All groups rated head-final RCs higher than head-initial RCs.

The AJT results on RCs were analyzed with an ordinal mixed regression model (Christensen, 2018) for ordinal data using the *clmm()* function in R (R Core Team, 2019). The dependent variable was the participants' ratings from 1 to 4. The fixed effects were group (NSs, HSs, and L2ers), RC type (SRC vs. ORC), RC headedness (head-final vs. head-initial), and their interactions; the random effects included a random intercept for subjects and a random intercept for items. The reference level for the variable group was the NSs. Using the *summary()* function, the model output is given in Table 6.3. There is a main effect of RC headedness, with head-final RCs being rated higher than head-initial RCs. There is also a main effect of group, with HSs and L2ers providing lower ratings for all four conditions than NSs. RC type has no effect (and does not interact with RC headedness nor group), indicating no SRC advantage. There is a significant interaction between RC headedness and group, meaning that the ratings were more similar across groups on head-initial RCs than on head-final RCs.

Figure 6.1. AJT results: Mean ratings on RCs

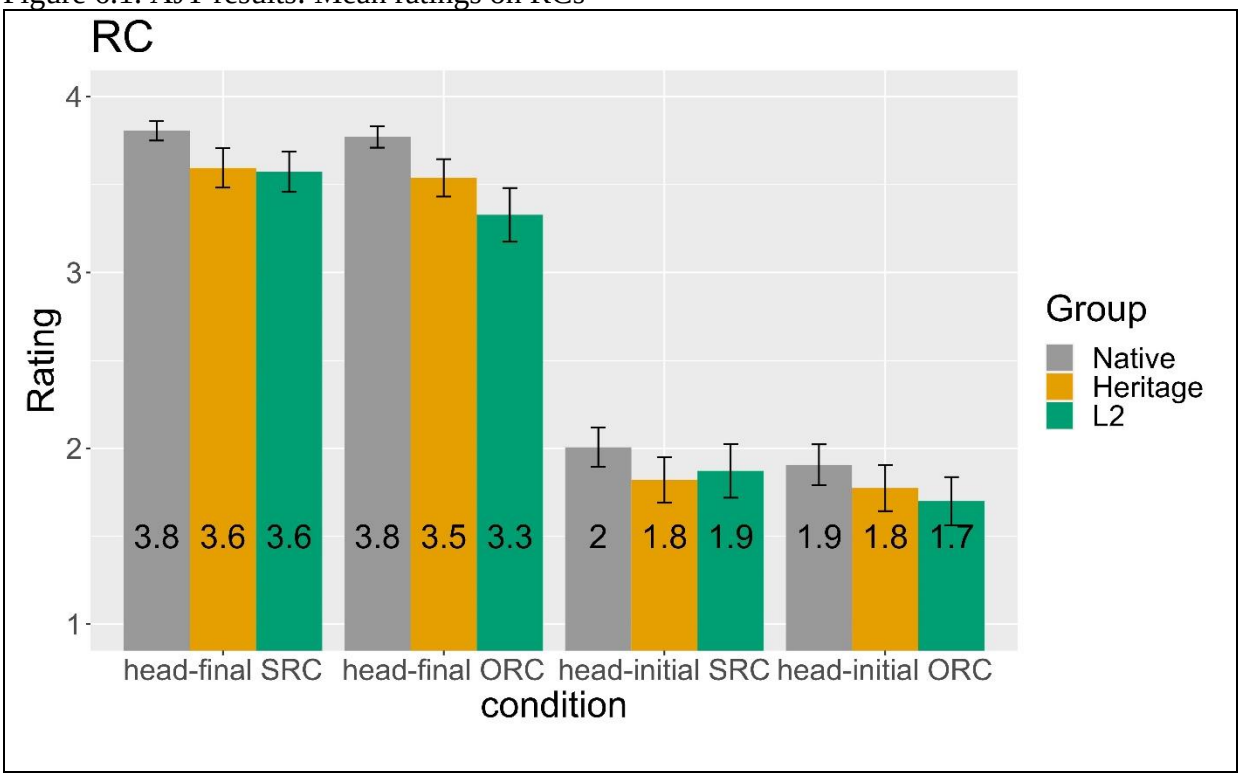


Table 6.3. Results from the ordinal mixed-effects model on RCs (AJT)

	Estimate	Std. Error	z value	Pr(>|z|)	
RC headedness	-4.15980	0.19910	-20.893	< 2e-16	***
RC type	0.24016	0.22105	1.086	0.277287	
Group (Heritage)	-0.83150	0.25017	-3.324	0.000888	***
Group (L2)	-1.23148	0.25414	-4.846	1.26e-06	***
RC headedness × RCtype	-0.01969	0.26771	-0.074	0.941370	
RC headedness × Group (Heritage)	0.62123	0.28203	2.203	0.027614	*
RC headedness × Group (L2)	0.78688	0.28843	2.728	0.006368	**
RC type × Group (Heritage)	0.08354	0.32197	0.259	0.795264	
RC type × Group (L2)	0.32823	0.32249	1.018	0.308782	
RC headedness × RCtype × Group (Heritage)	-0.25362	0.40611	-0.625	0.532291	
RC headedness × RCtype × Group (L2)	-0.23206	0.41219	-0.563	0.573433	
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Follow-up pairwise comparisons were made using *emmeans()*, with a Tukey-adjusted p-value of 0.05. While RC type did not interact significantly with other variables, pairwise comparisons including RC type were still conducted instead of being averaged out, since RC type

is of main interest for the RQs. Additionally, while it is more common to conduct post-hoc pairwise comparisons following significant interactions, some researchers argue that pairwise comparisons can follow non-significant interactions as well (J. Hsu, 1996). Below I report the between-group differences before the within-group differences; see Table 6.4. While HSs patterned with NSs and L2ers on head-final SRCs, NSs rated head-final SRCs significantly higher than L2ers (albeit a small difference, and despite HSs and L2ers having almost identical ratings). On head-final ORCs, NSs rated significantly higher than HSs and L2ers (albeit only a small difference between NSs and HSs), while HSs and L2ers rated similarly. All three groups patterned similarly in giving low ratings to head-initial RCs (both SRCs and ORCs). For all three groups, on both SRCs and ORCs, the ratings of head-final RCs were rated significantly higher than head-initial RCs.

Table 6.4. Pairwise comparison results from the ordinal mixed-effects model on RCs (AJT)

contrast	estimate	SE	df	z.ratio	p.value	
head-final,SRC,Native - head-final,SRC,Heritage	0.7480	0.268	Inf	2.792	0.1831	
head-final,SRC,Native - head-final,SRC,L2	0.9033	0.267	Inf	3.379	0.0352	*
head-final,ORC,Native - head-final,ORC,Heritage	0.8315	0.250	Inf	3.324	0.0420	*
head-final,ORC,Native - head-final,ORC,L2	1.2315	0.254	Inf	4.846	0.0001	*
head-initial,SRC,Native - head-initial,SRC,Heritage	0.3803	0.211	Inf	1.799	0.8189	
head-initial,SRC,Native - head-initial,SRC,L2	0.3484	0.219	Inf	1.591	0.9126	
head-initial,ORC,Native - head-initial,ORC,Heritage	0.2103	0.214	Inf	0.983	0.9981	
head-initial,ORC,Native - head-initial,ORC,L2	0.4446	0.223	Inf	1.994	0.6976	
head-final,SRC,Heritage - head-final,SRC,L2	0.1553	0.280	Inf	0.554	1.0000	
head-final,ORC,Heritage - head-final,ORC,L2	0.4000	0.264	Inf	1.518	0.9362	
head-initial,SRC,Heritage - head-initial,SRC,L2	-0.0319	0.243	Inf	-0.131	1.0000	
head-initial,ORC,Heritage - head-initial,ORC,L2	0.2343	0.245	Inf	0.955	0.9985	
head-final,SRC,Native - head-initial,SRC,Native	4.1795	0.206	Inf	20.270	<.0001	*
head-final,ORC,Native - head-initial,ORC,Native	4.1598	0.199	Inf	20.893	<.0001	*
head-final,SRC,Heritage - head-initial,SRC,Heritage	3.8119	0.233	Inf	16.352	<.0001	*
head-final,ORC,Heritage - head-initial,ORC,Heritage	3.5386	0.220	Inf	16.067	<.0001	*
head-final,SRC,L2 - head-initial,SRC,L2	3.6247	0.236	Inf	15.381	<.0001	*
head-final,ORC,L2 - head-initial,ORC,L2	3.3729	0.229	Inf	14.751	<.0001	*

Note: * indicates p<.05

6.1.4. Exclusion criteria for the purposes of the TVJT analysis

Recall that the TVJT tested whether participants correctly classified each RC as SRC or ORC. To do so, participants needed to understand both Mandarin RC headedness and the word order inside the RCs. Thus, before examining the results from the TVJT, I first excluded participants who incorrectly rejected head-final RCs or incorrectly accepted head-initial RCs in the AJT. For each individual, I calculated a mean rating for each of the four conditions. With a 1-to-4 rating scale (four being "fully acceptable"), 2.5 is considered a middle-point and ratings between 2 and 3 may not provide strong evidence for rejection or acceptance. Thus, participants who incorrectly accepted head-initial RCs were defined as those who gave a mean rating above three (>3) to either head-initial SRCs or ORCs. Among the five participants (three HSs and two L2ers) who incorrectly accepted head-initial RCs, four rated SRCs above three (range 3.2-3.4) but not ORCs (range 2.2-2.6); only one HS rated ORCs above three (at 3.2), but not SRCs (at 2.2).[40] The higher ratings of head-initial SRCs over head-initial ORCs are compatible with the NVN word order strategies outlined in Table 6.2. For head-final RCs, participants who incorrectly rejected these RCs were defined as those who gave a mean rating below two (<2) to either head-final SRCs or ORCs. No participant gave a mean rating below two to either head-final SRCs or head-final ORCs.[41] Based on the individual subjects' analysis outlined above, five participants in total were excluded in the subsequent analysis of the TVJT, discussed below.

[40] If the cutoff for incorrectly accepting head-initial RCs was set to >=3 (rather than >3), an additional six participants (two NSs, two HSs, and two L2ers) would have been excluded.
[41] If the cutoff for incorrectly rejecting head-final RCs was set to <=2 (rather than <2), no additional participants would have been excluded since the only one participant (a HS) who rated head-final SRCs at two (but rated head-final ORCs at 3.6) was already excluded because they had accepted head-initial SRCs at 3.2.

6.2. TVJT on RCs

The goal of the TVJT is to determine whether participants have correct interpretations of Mandarin RCs, which requires knowing the RC headedness and correct word order in SRCs and ORCs. Given that both SRCs and ORCs were tested, this TVJT also examines whether there is an SRC advantage (greater accuracy on SRCs than ORCs in the context of the task). Importantly, animacy is controlled for and simply using the SVO word order cue is insufficient to answer all four conditions correctly (more below).

6.2.1. Conditions

In the picture-based TVJT, four target conditions were created by crossing the factor 'RC type' (two levels: SRC vs. ORC) with the factor 'picture type' (two levels: matching vs. mismatching). With six tokens per condition, there were 24 target sentences per participant. The target sentence frames are provided in (22).

(22) Two sentence types of RCs in the TVJT

SRC: [_____Verb-Object] HeadNoun Predicate.
ORC: [Subject-Verb _____] HeadNoun Predicate.

Similar to the AJT, the TVJT tests RCs that modify the matrix subject.[42] Unlike the AJT, which tests whether learners know RC headedness, the TVJT tests interpretations. Furthermore, in the AJT, the head nouns of ORC items were always inanimate and the actions non-reversible, but the TVJT tests RCs with reversible actions. A sample token set is provided in Table 6.5: the target

[42] Subject-modifying RCs but not object-modifying RCs were tested because it is less complicated to have pictures for subject-modifying RCs like 'The woman [$_{RC}$ who looks at the man] holds a cup' than for object-modifying RCs like 'A cat is near a woman [$_{RC}$ who looks at the man]'.

responses are two TRUE and two FALSE responses. Both Picture A (on the left) and B (on the right) include the same three characters: Picture A depicts a man holding a cup looking at a woman, who in turn is looking at a man (without a cup); Picture B depicts a man (without a cup) looking at a woman, who in turn is looking at a man holding a cup. Note that the action is reversible so that the man can be an agent (looking at other people) or a patient (being looked at). The only difference between the two pictures lies in the direction of the action. This design requires participants to know that Mandarin RCs are head-final (tested in the AJT; see 6.1.4 for the exclusion criterion set for the subsequent TVJT analysis), and the word orders inside the RCs are Verb-Object in SRCs and Subject-Verb in ORCs. Crucially, unlike some previous studies, relying on linear word order or some other extra-syntactic strategy (e.g., animacy cues with non-reversible actions) will not yield the correct answer across all four conditions. Twelve predicates (but only 11 verbs) were used, with: *kanjian* 'see', *la* 'drag', *gen* 'follow', *qin* 'kiss', *mo* 'touch', *yao* 'bite', *bao* 'hug', *zhi* 'point to', *zhui* 'chase', *yadao* 'rest on', and *da* 'hit'.[43]

Table 6.5. Sample token set exemplifying the four conditions testing RCs in the TVJT

		Picture A	Picture B
			
SRC	[kànjiàn nǚrén de] **nánrén** názhe bēizi [See woman de] man hold cup The man [who sees a woman] holds a cup.	TRUE	FALSE
ORC	[nǚrén kànjiàn de] **nánrén** názhe bēizi [woman see de] man hold cup The man [who a woman sees] holds a cup.	FALSE	TRUE

[43] The verb "hug" was used twice, with two different subjects/objects. The Mandarin verb *ya*(dao) is literally "press (against)", but means "rest on/upon" or "be on top of" in the pictures I tested.

6.2.2. Predictions

NSs were expected to perform at ceiling. For HSs and L2ers, two predictions are outlined below. First, under English transfer, both HSs and L2ers might transfer their knowledge of English RC headedness to Mandarin. If they misread ORCs as head-initial SRCs ('head error') under English transfer, such as "a woman [$_{RC}$ who sees a man] holds a cup", they would incorrectly choose FALSE in both ORC conditions (since there is no woman holding a cup). However, if they misread ORCs as "a woman sees a man [$_{RC}$ who holds a cup]", they would accidentally give correct responses for ORCs for the wrong reason, marked with "?"; see Table 6.6. I use '??' for SRCs because verb-initial SRCs are harder to misanalyse as head-initial RCs (which are noun-initial); participants who do not understand head-final RCs might not be able to parse head-final SRCs at all and may just read the matrix predicates. If those who cannot parse RCs simply ignore the RCs and look at the matrix predicates, they would incorrectly choose TRUE for all four conditions (since there is a man holding a cup in all four), or at least to both SRC conditions (given that they are verb-initial). Another possibility for why participants might accept all four conditions is the use of the so-called "good-enough" reading strategy, known to be used by NSs and L2ers alike (e.g., Christianson, 2016; D.-B. Hsu, 2017; Lim & Christianson, 2013). Given that the matrix predicates ('holds a cup' in this example) match all four conditions, participants who adopt the good-enough reading strategy might give TRUE responses to all four. Overall, L2ers and HSs are expected to perform similarly, with greater accuracy with increased proficiency.

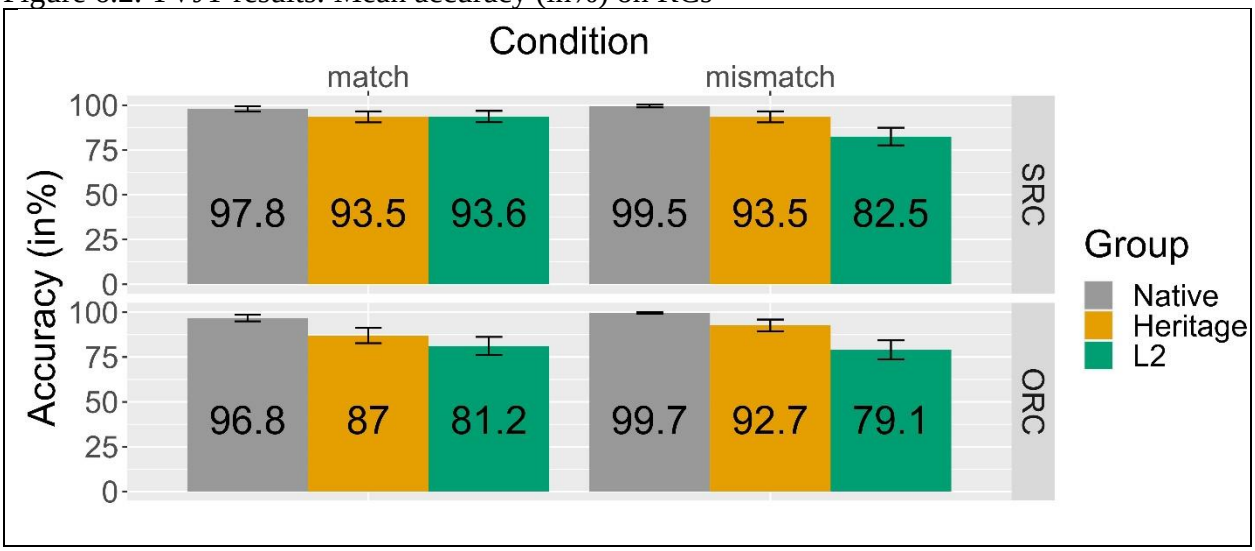
Table 6.6. Predictions for the TVJT testing RCs (non-target-like responses are shaded in gray)

	NSs	HSs and L2ers under English transfer of RC headedness
SRC; matching (Picture A)	TRUE	TRUE??
SRC; mismatching (Picture B)	FALSE	TRUE??
ORC; mismatching (Picture A)	FALSE	FALSE
ORC; matching (Picture B)	TRUE	FALSE?

6.2.3. Results: group analysis

Data from 62 NSs, 42 HSs, and 38 L2ers were analyzed (see Chapter 4 for participant details), after excluding three HSs and two L2ers who incorrectly accepted head-initial RCs in the AJT (see section 6.1.4). A correct response was coded as "1" and an incorrect response as "0". Then, the raw scores in each condition (range 0-6 as there were six tokens) were averaged across the participants and converted to percentages. Figure 6.2 shows the group results of the mean accuracy on RCs.

Figure 6.2. TVJT results: Mean accuracy (in%) on RCs



Note. Error bars show standard error.

The TVJT data on RCs were analyzed in a logistic mixed-effects model (Jaeger, 2008) using the *glmer()* function in the *lme4* package in R (R Core Team, 2019). The model included group (NSs, HSs, and L2ers), RC type (SRC vs. ORC), picture type (matching vs. mismatching) and their interactions as fixed effects; the random effects included a random intercept for subjects and a random intercept for items. For the variable group, dummy coding was used; the reference level was the NSs. Using the *Anova()* function in the *car* package to assess the overall effect, the model

output is presented in Table 6.7. There is a main effect of group, suggesting that NSs were more accurate than HSs, who in turn were more accurate than L2ers. There is a marginal effect of picture type, with overall slightly more accurate responses to matching pictures than to mismatching pictures. There is no effect of RC type (and no interactions with other variables), indicating no SRC advantage. The two-way interaction of picture type and group is significant, with L2ers being more accurate on matching pictures than mismatching pictures.

Table 6.7. Results from the logistic mixed-effects model on RCs

	Chisq	Df	Pr(>Chisq)	
(Intercept)	105.9295	1	< 2.2e-16	***
RC type	0.8731	1	0.350098	
Picture type	3.1437	1	0.076220	.
Group	6.1324	2	0.046598	*
RC type × Picture type	0.7607	1	0.383107	
RC type × Group	2.8676	2	0.238396	
Picture type × Group	12.8057	2	0.001657	**
RC type × Picture type × Group	0.3937	2	0.821305	
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Pairwise comparisons were conducted via *emmeans* (Lenth, 2019) following the significant interactions; the p-values are significant at the Tukey-adjusted alpha level of .05. Similar to the AJT, the RC type did not interact significantly with other variables. Given that RC type is of main interest for the RQs and that some researchers consider post-hoc pairwise comparisons valid even when the interactions are not significant, pairwise comparisons including RC type were conducted; see Table 6.8. Differences between groups are reported before differences within groups. All three groups patterned similarly on the SRC-matching condition. For the other three conditions, NSs scored significantly higher than L2ers; NSs also scored significantly higher than HSs on the ORC-matching condition, but the differences are only marginal in the two mismatching conditions. HSs and L2ers performed similarly on three (SRC-matching, ORC-matching, SRC-mismatching)

conditions, but HSs scored marginally higher than L2ers on the ORC-mismatching condition. In terms of SRC/ORC asymmetry, both NSs and HSs performed similarly on both SRCs and ORCs in both matching and mismatching conditions. For the matching (but not mismatching) condition, L2ers performed significantly better on SRCs than ORCs.

In addition, L2ers performed significantly better on the SRC-matching condition than the SRC-mismatching condition, but the difference between the ORC-matching condition and the ORC-mismatching condition is not significant. Such an over-acceptance of the matching (but not mismatching) conditions is an indication of Yes-bias. By contrast, HSs demonstrated no Yes-bias, since they performed similarly on matching and mismatching conditions (even performing numerically better on the ORC-mismatching condition than the ORC-matching condition). Given that Yes-bias is not of particular interest in this dissertation, it will not be discussed further.

Table 6.8. Pairwise comparison results from the logistic mixed-effects model on RCs

contrast	estimate	SE	df	z.ratio	p.value	
SRC,match,Native - SRC,match,Heritage	1.2366	0.539	Inf	2.296	0.4793	
SRC,match,Native - SRC,match,L2	1.1370	0.545	Inf	2.088	0.6314	
ORC,match,Native - ORC,match,Heritage	1.6422	0.463	Inf	3.548	0.0200	*
ORC,match,Native - ORC,match,L2	2.0779	0.456	Inf	4.553	0.0003	*
SRC,mismatch,Native - SRC,mismatch,Heritage	2.6260	0.817	Inf	3.216	0.0586	
SRC,mismatch,Native - SRC,mismatch,L2	3.8284	0.793	Inf	4.826	0.0001	*
ORC,mismatch,Native - ORC,mismatch,Heritage	3.5025	1.077	Inf	3.253	0.0523	
ORC,mismatch,Native - ORC,mismatch,L2	4.8396	1.061	Inf	4.561	0.0003	*
SRC,match,Heritage - SRC,match,L2	-0.0995	0.496	Inf	-0.201	1.0000	
ORC,match,Heritage - ORC,match,L2	0.4357	0.409	Inf	1.065	0.9960	
SRC,mismatch,Heritage - SRC,mismatch,L2	1.2024	0.451	Inf	2.664	0.2446	
ORC,mismatch,Heritage - ORC,mismatch,L2	1.3371	0.436	Inf	3.070	0.0895	
SRC,match,Native - ORC,match,Native	0.4411	0.472	Inf	0.934	0.9988	
SRC,mismatch,Native - ORC,mismatch,Native	-0.7091	1.230	Inf	-0.576	1.0000	
SRC,match,Heritage - ORC,match,Heritage	0.8467	0.341	Inf	2.480	0.3523	
SRC,mismatch,Heritage - ORC,mismatch,Heritage	0.1674	0.376	Inf	0.446	1.0000	
SRC,match,L2 - ORC,match,L2	1.3819	0.339	Inf	4.082	0.0026	*
SRC,mismatch,L2 - ORC,mismatch,L2	0.3020	0.264	Inf	1.145	0.9926	
SRC,match,Native - SRC,mismatch,Native	-1.4176	0.800	Inf	-1.773	0.8329	
ORC,match,Native - ORC,mismatch,Native	-2.5678	1.048	Inf	-2.451	0.3713	
SRC,match,Heritage - SRC,mismatch,Heritage	-0.0281	0.385	Inf	-0.073	1.0000	
ORC,match,Heritage - ORC,mismatch,Heritage	-0.7075	0.331	Inf	-2.139	0.5947	
SRC,match,L2 - SRC,mismatch,L2	1.2738	0.341	Inf	3.740	0.0100	*
ORC,match,L2 - ORC,mismatch,L2	0.1940	0.261	Inf	0.744	0.9999	

Note: * indicates p<.05

To address the effect of proficiency, an additional analysis was conducted on HSs and L2ers, since NSs did not complete the entire proficiency test (see Chapter 4 for participant details). The model is similar to the original model in Table 6.7, but with proficiency as an added covariate (no interaction with other variables). The model output is provided in Table 6.9. There is a significant main effect of proficiency, indicating that accuracy improves as proficiency increases. There is also a main effect of RC type, which indicates that SRCs were answered more correctly than ORCs. There is no effect of picture type nor group. There is significant interaction between picture type and group (but not other interactions), indicating that it is primarily L2ers, but not HSs, who were more accurate in the matching conditions. Differing from the results of the original model reported in Table 6.7, in the current model, RC type is significant while picture type and group are not. Pairwise comparisons (not reported here) indicated that, with only HSs and L2ers in the current model, HSs now performed significantly better than L2ers on both mismatching conditions.

Table 6.9. Results from the logistic mixed-effects model on RCs, which includes only HSs and L2ers and adds proficiency as a covariate

	Chisq	Df	Pr(>Chisq)
(Intercept)	0.2726	1	0.601575
RC type	5.8625	1	0.015466 *
Picture type	0.0014	1	0.970384
Group	0.5140	1	0.473416
Proficiency	42.4522	1	7.243e-11 ***
RC type × Picture type	1.5160	1	0.218228
RC type × Group	1.0478	1	0.306013
Picture type × Group	6.8901	1	0.008667 **
RC type × Picture type × group	0.3373	1	0.561391
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

6.2.4. Results: individual subjects' analysis

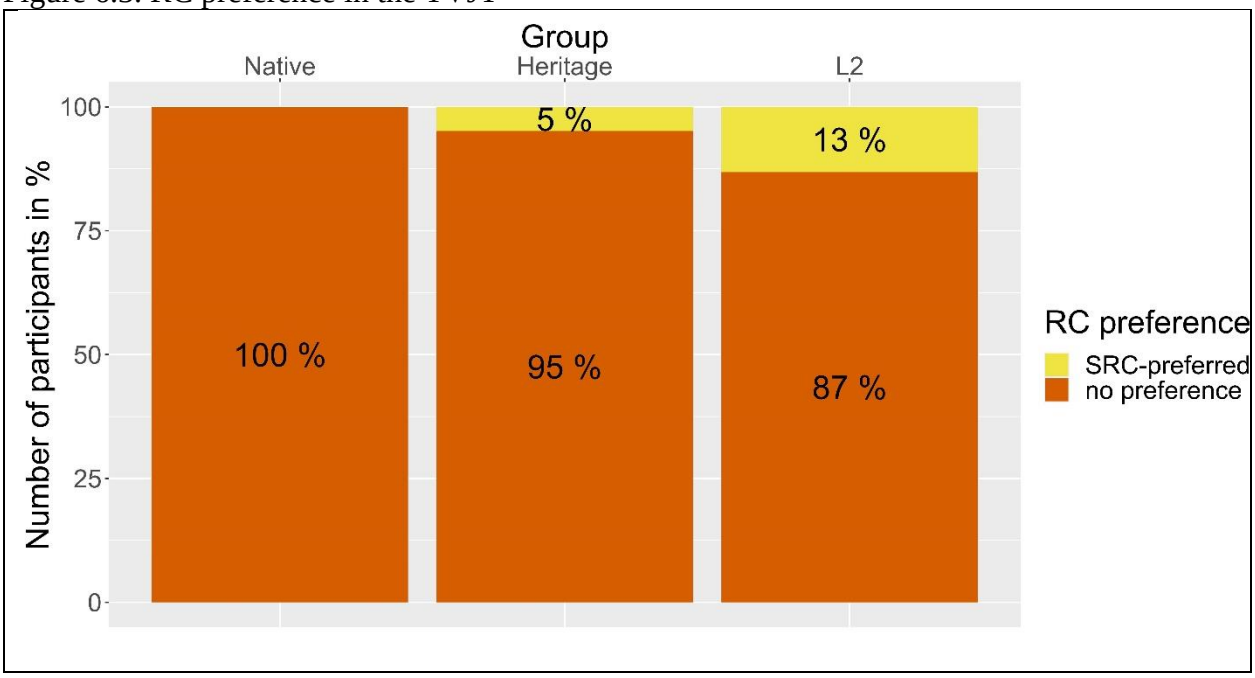
At the group level, only L2ers, but not HSs, show an SRC advantage on the matching, but not mismatching, condition. To confirm whether such an SRC advantage holds at the individual level, an individual subjects' analysis was conducted by classifying participants into one of the three patterns: those having an SRC preference, no preference, or an ORC preference. Before this classification, I first checked if any participants adopted an all-acceptance/TRUE strategy or an all-rejection/FALSE strategy.

If, for all four conditions, a participant accepted (i.e., answered TRUE to) at least four tokens out of six (>=4), they were classified as having an all-acceptance strategy, which indicates a lack of correct understanding of Mandarin RCs or the use of a good-enough strategy (e.g., Christianson, 2016; D.-B. Hsu, 2017; Lim & Christianson, 2013). If, for all four conditions, a participant accepted (i.e., answered TRUE to) at most two tokens out of six (<=2), they were classified as having an all-rejection strategy, which also indicates a lack of correct understanding of Mandarin RCs. Only one L2er demonstrated an all-acceptance pattern while neither NSs nor HSs showed an all-acceptance pattern. No one showed an all-rejection pattern.

After excluding the one participant who showed an all-acceptance pattern, the remaining participants were classified into one of the three patterns: having an SRC preference, no preference, or an ORC preference. In Lee-Ellis (2011), the SRC/ORC advantage was calculated by two-token differences (out of six tokens per condition). Given that there were 12 tokens for SRCs and 12 for ORCs here (after combining matching and mismatching conditions), I set the criterion to be four-token differences. If a participant correctly accepted more SRCs than ORCs by four tokens (e.g., 12 vs. 8), they were classified as having an SRC preference. If a participant correctly accepted more ORCs than SRCs by four tokens, they were classified as having an ORC preference. If a

participant accepted the same number of tokens of SRCs and ORCs or if the differences were fewer than four tokens, they were classified as having no preference. Only five L2ers (out of 38; 13%) and two HSs (out of 41; 5%) showed an SRC preference; most participants showed no preference using such criterion, as shown in Figure 6.3. The SRC preference, albeit with only a few participants, is consistent with the group results that only L2ers showed an SRC advantage in the pairwise comparisons.

Figure 6.3. RC preference in the TVJT



6.3. General discussion

Recall that the broad RQs asked whether proficiency-matched HSs and L2ers acquire language phenomena that are absent in or different from English, and whether HSs have selective advantages over proficiency-matched L2 learners. Using two different tasks, the specific RQs investigated whether HSs and L2ers could correctly accept head-final (but not head-initial) Mandarin RCs and have the correct interpretations of head-final RCs. In addition, SRC/ORC asymmetry is examined.

I discuss HS/L2 comparisons, dominant language transfer, SRC/ORC asymmetry, and word order strategies/cues below. In the AJT (see Figure 6.1) HSs and L2ers performed similarly, but HSs were more native-like than L2ers in head-final SRC. In the TVJT (see Figure 6.2), HSs were numerically (but not significantly) more accurate than L2ers on three conditions, and significantly (albeit only marginally) more accurate than L2ers on the ORC-mismatching condition. Compared to NSs, HSs were more native-like than L2ers in the mismatching conditions.

Given that RC acquisition is stabilized in monolingual children by age five (C. Hsu, 2014), HSs were expected to acquire RCs by about that time, or slightly later than monolingual children (This "protracted acquisition" was found in R. Jia & Paradis, 2020, which compared both child HSs and their monolingual peers). L2ers were also expected to acquire RCs since purely syntactic phenomena are relatively easy for L2ers (as compared to morphology, e.g., the Bottleneck Hypothesis, Slabakova, 2008, 2014). The lack of a clear HS advantage is found in studies with Korean RCs (O'Grady et al., 2001; T. Lee, 2016); even though they had similar accuracy, HSs and L2ers exhibited different error patterns in T. Lee (2016). The present dissertation finds a slight HS advantage given that HSs were more native-like than L2ers in some but not all conditions, though the statistical significance is either small or marginal.

In terms of dominant language transfer, unlike some head errors found in child HSs or English-Mandarin bilingual children (R. Jia & Paradis, 2020; Tsoi et al., 2019), HSs and L2ers in this dissertation correctly rejected head-initial and only accepted head-final Mandarin RCs in the AJT. There is no clear evidence of misinterpretations in the TVJT, though the TVJT in this dissertation is not ideal for checking whether participants have made head errors (incorrectly assuming RCs as head-initial) or reversal errors (incorrectly assuming ORCs as SRCs, and vice versa), as there are multiple reasons for giving a TRUE or FALSE response.

The lack of English transfer in this dissertation differs from studies on Korean RCs (e.g., O'Grady et al., 2001; T. Lee, 2016). There are at least three possible explanations: proficiency, task effect, and structural complexity (for English speakers). First, participants in this dissertation were likely more proficient in Mandarin than those in the Korean studies were in Korean. For example, HSs and some L2ers in O'Grady et al. (2001) were just in their second-semester of Korean classes, while in this dissertation HSs had taken an average of almost six years of Mandarin classes while L2ers had taken an average of almost four years (see Chapter 4 for details). The second possible explanation is of task effect. Studies such as O'Grady et al. (2001) and T. Lee (2016) used a listening picture-selection task to test Korean RCs while this dissertation used a written AJT and TVJT (both tasks having Romanization). The written tasks are presumably less demanding since participants could re-read and highlight the words (on the web page) to help parse the sentences. Third, Korean RCs may indeed be more complex than Mandarin RCs for English speakers. To learn Mandarin RCs, English speakers only need to learn that Mandarin RCs are head-final and that the word order inside the RCs remains SVO (SV and VO). However, to learn Korean RCs, English speakers need to learn two new properties: RCs are head-final and common nouns require case marking (note that English only marks case on pronouns such as *he* vs. *him*). The fact that both Korean SRCs and ORCs have NVN sequences might have led HSs and L2ers of Korean to ignore the case markers and incorrectly assume that Korean RCs are head-initial like English. In Mandarin, however, the fact that SRCs are verb-initial forces HSs and L2ers to notice that Mandarin has head-final RCs.

Next, I discuss the SRC/ORC asymmetry. Only L2ers showed an SRC advantage on the matching (but not the mismatching) conditions, supporting the Noun Phrase Accessibility Hierarchy (NPAH; Keenan & Comrie, 1977) and the expectation/frequency-based theories (e.g.,

Reali & Christiansen, 2007). Given that the HSs' overall accuracy was fairly high (between 87% and 93%), they might have been too advanced to show any asymmetry. Recall that, unlike previous studies, the TVJT used in this dissertation removed the animacy cues and, importantly, SVO word order cues. While an ORC advantage has been reported in some L2 studies (e.g., the multiple-choice questionnaire in Cui, 2013), I speculate that this may be a result of L2ers' use of SVO word order strategy rather than complete acquisition of Mandarin RCs. The SRC advantage found here, even though only with L2ers, provides evidence for that speculation; once knowing SVO word order is not enough, the ORC advantage is not found. Given that the SRC advantage in Mandarin processing studies is still debated (Jäger et al., 2015; Mansbridge et al., 2017, among many others), it is uncertain how processing plays a role in the acquisition of RCs. In terms of frequency, Mandarin SRCs are more frequent than ORCs, which is consistent with the finding from L2ers, despite the fact that RCs with two animate nouns (tested in the TVJT) are rare (for corpus findings, see F. Wu et al., 2012, for a summary).

While an SRC advantage is not always found in Mandarin (and other languages with head-final RCs such as Cantonese, Japanese, and Basque) and sometimes an ORC advantage is reported, this does not mean that Mandarin has a unique acquisition/processing preference. Rather, it is that multiple factors (NPAH and expectation-based theories vs. memory-based theories and word order) play a role in the acquisition and processing of RCs (as is the conclusion in some recent L1 processing studies, e.g., Mansbridge et al., 2017). In languages with head-initial RCs such as English, multiple factors conspire to make SRCs easier to acquire and process than ORCs. In languages with head-final RCs such as Mandarin, multiple factors point to opposing directions/ tendencies and render an SRC advantage sometimes undetectable.

Lastly, a note on the use of word order cues is in order. To distinguish between SRCs and ORCs in Mandarin, participants need to rely on word order, but simply using the SVO word order strategy is not enough to answer all conditions correctly in the TVJT reported here. Participants also had to know that Mandarin RCs are head-final and that the word order inside the RCs are SVO. To address strategies (or "cues" as in the Unified Competition Model, MacWhinney, 2012) used by speakers of different languages, I additionally tested L1-Korean L2-Mandarin learners using the same TVJT and compared them with proficiency-matched L1-English L2ers (C. Chen, 2019c). L1-English L2-Mandarin learners were significantly more accurate than proficiency-matched L1-Korean L2-Mandarin learners in the SRC-matching condition, and numerically (but not significantly) more accurate in the other three conditions. L1-Korean L2ers' accuracy was between 78% and 83% but showed no asymmetry. Crucially, the fact that L1-English L2ers outperformed L1-Korean L2ers is better explained by the use of different strategies or cues. When comprehending RCs in their L1s, Korean NSs rely on case-marking cues while Mandarin and English NSs rely on word order cues. Thus, L1-English L2ers have an advantage over L1-Korean L2ers in comprehending Mandarin RCs. Alternatively, if RC headedness is transferred from Korean to Mandarin (both head-final), L1-Korean L2ers should outperform L1-English L2ers, but this was not the case.

Returning to the broad RQs, HSs showed a slight HS advantage over L2ers in this domain in being more native-like than L2ers, though the statistical significance is either small or marginal. While the performances of HSs and L2ers are not completely native-like, both groups overcame dominant language transfer from English in correctly rejecting head-initial RCs and were able to correctly interpret Mandarin head-final RCs most of the time.

Chapter 7. Interpretation of anaphors in HS/L2 Mandarin

This chapter reports on Mandarin anaphors, particularly long-distance (LD) reflexives, tested in the same picture-based TVJT that tested relative clauses (see Chapter 4 for the TVJT design). The broad RQs in this dissertation are again listed below, as are the RQs for anaphors, 1d and 2d, which are instantiations of broad RQs 1 and 2.

- Broad RQ 1: Can HSs and L2ers of Mandarin whose dominant language is English fully acquire the properties of Mandarin that are different from or absent in English?
- Broad RQ 2: Do HSs have selective advantages over proficiency-matched L2ers, and does this vary by linguistic domain?
- RQ 1d: Can HSs and L2ers acquire the LD reflexives in Mandarin, despite lack of such phenomena in English?
- RQ 2d: Do HSs have an advantage over L2ers in LD reflexives in the domain of syntax-semantics interface or syntax-discourse interface?

Given that LD reflexives are acquired late (at least after age eight) by monolingual children and that LD reflexives are at the interface of syntax-semantics or syntax-discourse, no HS advantage was expected. In addition to testing LD readings of *ziji*, the TVJT employed here examined all three anaphors in Mandarin: pronoun *ta*, simplex reflexive *ziji*, and complex reflexive *taziji*. Specifically, I tested whether participants correctly allowed both LD and local readings for *ziji*, only local readings for *taziji*, and only LD readings for *ta*.

7.1. TVJT conditions

The picture-based TVJT was adapted from C. Chen (2019b), which in turn was adapted from J.-H. Kim et al.'s (2009) study on Korean reflexives. Three anaphors were tested: *ta*, *ziji*, and *taziji*. In Chinese characters, *ta* 他 (with the 'human' radical) is the default third-person singular pronoun that can refer to both 'he/him' and 'she/her' while *ta* 她 (with the 'female' radical) can only be

used for ‘she/her’. To avoid any influence from orthography, the two antecedents (matrix subjects and embedded subjects) agreed in gender (by using typical male or female English names). Half of the sentences used male names and half used female names. The target sentence frames are provided in (23); the matrix verb was always ‘say’ (with a speech bubble in the picture). All the embedded anaphors were direct objects.[44] Twelve verbs were used: *kan* ‘see’, *qiao(dao)* ‘hit/knock’, *mo* ‘touch’, *yao* ‘bite’, *bao* ‘hug’, *guancha* ‘observe’, *xiao* ‘laugh at’, *geshang* ‘cut’, *zhi* ‘point at ’, *tang* ‘burn’, *da* ‘hit/slap’, and *hua* ‘draw’.[45]

(23) Three sentence types testing binding in the TVJT
   a) Name1 say Name2 Verb *ta*.
   b) Name1 say Name2 Verb *ziji*.
   c) Name1 say Name2 Verb *taziji*.

Six target conditions were created by crossing anaphor type (three levels: *ta* vs. *ziji* vs. *taziji*) with picture type (two levels: LD vs. local readings of the anaphor). The *ta* conditions served as the control because there are no cross-linguistic differences between Mandarin and English. There were six tokens per condition, resulting in 36 target sentences per participant. A token set is provided in Table 7.1: the target responses are four TRUE and two FALSE responses. Pictures A and B have the same two characters (Peter and John) and the speaker outside of the speech bubble is the same. In Picture A, Peter is drawing John, so this picture depicts the LD reading of the anaphor. In Picture B, Peter is drawing a self-portrait, so this picture depicts the local reading of

[44] With possessive objects such as ‘John said Peter took his book’ and its Mandarin equivalent, both John and Peter are possible antecedents. This is not ideal, since both conditions for possessive pronouns would be TRUE. The Mandarin possessive marker *de* directly follows *ta*, *ziji*, and *taziji*, i.e., *ta-de* ‘his/her,’ *ziji-de* ‘self’s’ (ungrammatical in English), and *taziji-de* ‘himself/herself’s’ (ungrammatical in English).

[45] While “to (literally) knock someone” is unnatural in English, in Mandarin it is acceptable and can mean “to hit (someone) (e.g., with a hammer).” While it sounds more natural with a body part expressed by a possessive, e.g., “*ta/ziji/taziji de* head”, for reasons mentioned in footnote 44, “knock *ta/ziji/taziji*” was used.

the anaphor. Picture A thus elicits a TRUE response when the embedded object is *ta* or *ziji* while Picture B elicits a TRUE response when the embedded object is *ziji* or *taziji*.

Table 7.1. Sample token set exemplifying the six conditions testing anaphors in the TVJT

	Picture A (LD readings of the anaphor)	Picture B (local readings of the anaphor)
	 LD	 Local
John$_i$ shuō Peter$_j$ huà-le tā $_{i/*j/k}$. John say Peter draw-ASP self 'John said that Peter drew him.'	TRUE	FALSE
John$_i$ shuō **Peter**$_j$ huà-le zìjǐ $_{i/j}$. John say Peter draw-ASP self 'John said that Peter drew himself/him.'	TRUE	TRUE
John$_i$ shuō **Peter**$_j$ huà-le tāzìjǐ$_{*i/j}$. John say Peter draw-ASP himself 'John said that Peter$_j$ drew himself.'	FALSE	TRUE

As explained in Chapter 4, there were two test lists. In each list, each picture was repeated three times, with *ta, ziji*, and *taziji*, respectively; pictures differing only in the local vs. LD readings never appeared in the same list. A potential concern was that each picture was repeated three times (with three anaphors) in each list. This could potentially give rise to two problems: first, priming effects, with participants over-accepting sentences (i.e., choosing TRUE responses more often), because they had seen similar sentences before;[46] and second, that participants would actually

[46] The priming effect could be fixed by having six test lists with 36 verbs, fully counterbalancing token sets across lists. As mentioned in Chapter 4, the token sets were not fully counterbalanced due to limited verbs that could be used in a picture-based task with HSs and L2ers.

under-accept sentences with *ziji* because they had seen other anaphor forms (*ta* and *taziji*) that were (less) unambiguous.

7.2. Prediction

NSs are expected to give a TRUE response to both local and LD readings of *ziji*, but only to local readings of *taziji* and LD readings of *ta*. However, based on previous experimental studies that report much variability, it is predicted the rates of TRUE responses will not be at ceiling. For HSs and L2ers, both groups might choose TRUE only for local readings because they transfer their knowledge of English reflexives to Mandarin reflexives (*ziji* and *taziji*) and/or they did not receive enough input of LD readings. For *ta* and *taziji*, both HSs and L2ers are expected to perform similarly to NSs, since there are no cross-linguistic differences. However, if learners choose LD readings for *taziji* (as in C. Chen, 2019b; Zeng, 2010), one interpretation is that they treat *taziji* as *ta* (plus an intensifying usage of *ziji*); if so, they should perform similarly on both *ta* and *taziji*. See Table 7.2 for the predictions. No HS advantage is predicted on LD reflexives because it is late-acquired in monolingual children.

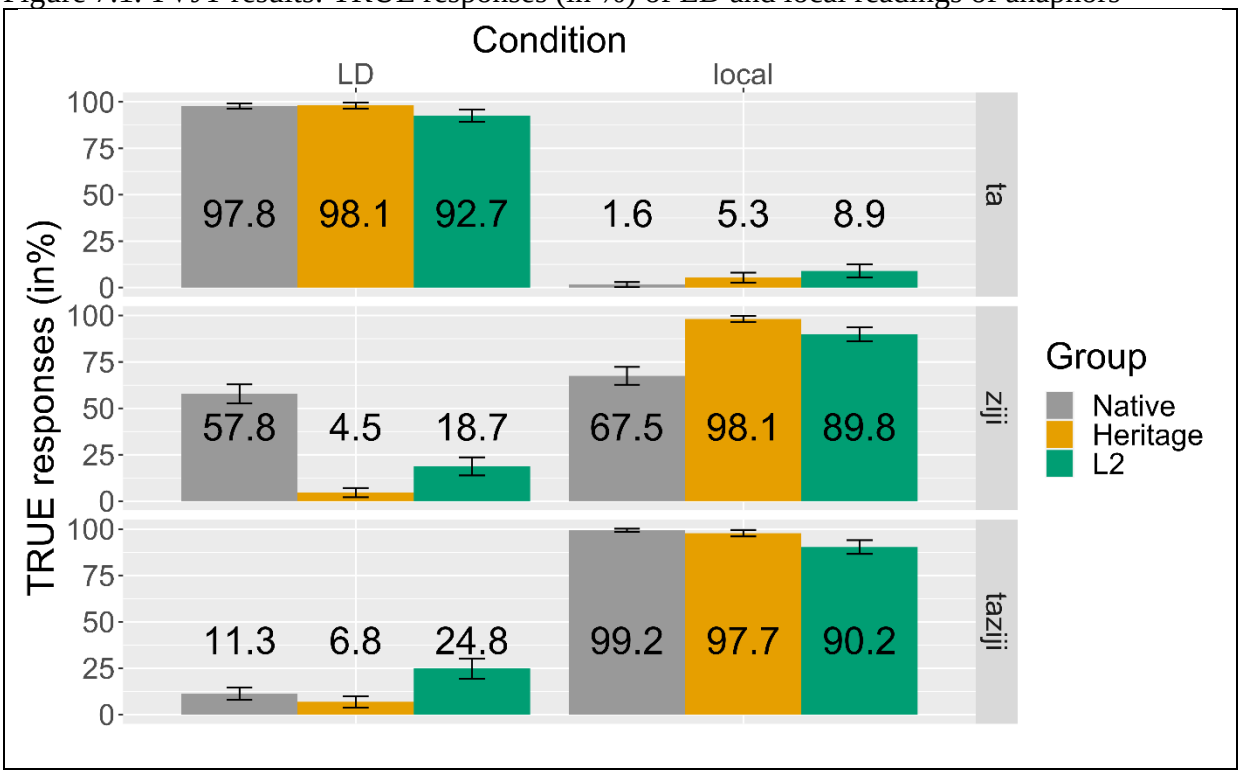
Table 7.2. Predictions for the TVJT conditions on anaphors (non-target-like responses are shaded in gray)

	Native speakers	HSs and L2ers
LD readings of *ta*	TRUE	TRUE
Local readings of *ta*	FALSE	FALSE
LD readings of *ziji*	TRUE	FALSE
Local readings of *ziji*	TRUE	TRUE
LD readings of *taziji*	FALSE	FALSE
Local readings of *taziji*	TRUE	TRUE

7.3. Group analysis

Data from 62 NSs, 44 HSs, and 41 L2ers were included (see Chapter 4 for participant details). Following J.-H. Kim et al. (2009) and C. Chen (2019b), a TRUE response was coded as "1" and a FALSE response was coded as "0", regardless of the target response (since both LD and local readings of *ziji* are correct). Then, the raw scores in each condition (range 0-6 as there were six tokens) were averaged across the participants and converted to percentages. Figure 7.1 shows the group results on the mean acceptance of LD and local readings.

Figure 7.1. TVJT results: TRUE responses (in %) of LD and local readings of anaphors



Note. Error bars show standard error. (Taken from C. Chen (2020a))

The TVJT results on the anaphors were analyzed in a logistic mixed-effects model (Jaeger, 2008) using the *glmer()* function in the *lme4* package in R (R Core Team, 2019). The dependent variable was the participants' response of 1 or 0. The model included group (NSs, HSs, and L2ers),

anaphor (*ta*, *ziji*, and *taziji*), picture (matching local vs. LD readings of the anaphor) and their interactions as fixed effects. Dummy coding was used. The reference level was the NSs for the variable group and *ta* for the variable anaphor. The random effects included a random intercept for subject. Item was initially included as a random effect, but later removed as the data did not warrant the inclusion of two random effects (singularity issues). See Table 7.3 for the model output using the *Anova()* function to assess the overall effect. There is a main effect of anaphor, antecedent, and group; the two-way and three-way interactions are all significant.

Table 7.3. Results from the logistic mixed-effects model on anaphors

	Chisq	Df	Pr(>Chisq)
(Intercept)	115.740	1	< 2.2e-16 ***
Anaphor	283.931	2	< 2.2e-16 ***
Antecedent	217.943	1	< 2.2e-16 ***
Group	12.350	2	0.002081 **
Anaphor × antecedent	355.293	2	< 2.2e-16 ***
Anaphor × group	131.483	4	< 2.2e-16 ***
Antecedent × Group	26.812	2	1.506e-06 ***
Anaphor× Antecedent × Group	144.080	4	< 2.2e-16 ***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Pairwise comparisons were conducted via *emmeans* (Lenth, 2019) following the significant interactions; the p-values are significant at the Tukey-adjusted alpha level of .05. The between-group differences are reported in the order of *ta*, *ziji*, and *taziji*, before the within-group differences between different types of anaphor readings; see Table 7.4. On *ta*, there were no group differences between NSs and HSs, nor between HSs and L2ers. On local readings of *ta*, NSs and L2ers had a small but significant difference (p=0.02) in that L2ers over-accepted local readings. On *ziji*, NSs differed significantly from HSs and L2ers; HSs and L2ers also differed significantly from each other. Compared to NSs, both HSs and L2ers under-accepted LD readings of *ziji*, and over-accepted local readings of *ziji*, with HSs being more categorical than L2ers. On *taziji*, while HSs

patterned with NSs, L2ers differed significantly from NSs in over-accepting LD readings and under-accepting local readings. L2ers also accepted LD readings of *taziji* significantly more than HSs.

For *ta* and *taziji*, all three groups made a reliable distinction between local vs. LD readings. For *ziji*, while HSs and L2ers still made a reliable distinction between the local vs. LD readings, NSs did not.[47] NSs' responses on *ziji* and *taziji* were significantly different on both the LD and local readings while HSs and L2ers treated *ziji* and *taziji* similarly.

Table 7.4. Pairwise comparison results from the logistic mixed-effects model on anaphors

contrast	estimate	SE	df	z.ratio	p.value	
ta,LD,Native - ta,LD,Heritage	-0.133659	0.583	Inf	-0.229	1.0000	
ta,LD,Native - ta,LD,L2	1.286072	0.442	Inf	2.909	0.2475	
ziji,LD,Native - ziji,LD,Heritage	3.451650	0.327	Inf	10.552	<.0001	*
ziji,LD,Native - ziji,LD,L2	1.851766	0.216	Inf	8.590	<.0001	*
taziji,LD,Native - taziji,LD,Heritage	0.562988	0.307	Inf	1.832	0.9353	
taziji,LD,Native - taziji,LD,L2	-0.975318	0.239	Inf	-4.080	0.0058	*
ta,local,Native - ta,local,Heritage	-1.233829	0.503	Inf	-2.454	0.5659	
ta,local,Native - ta,local,L2	-1.799780	0.477	Inf	-3.769	0.0191	*
ziji,local,Native - ziji,local,Heritage	-3.269284	0.475	Inf	-6.887	<.0001	*
ziji,local,Native - ziji,local,L2	-1.485695	0.255	Inf	-5.822	<.0001	*
taziji,local,Native - taziji,local,Heritage	1.049539	0.718	Inf	1.463	0.9930	
taziji,local,Native - taziji,local,L2	2.600443	0.624	Inf	4.164	0.0041	*
ta,LD,Heritage - ta,LD,L2	1.419731	0.524	Inf	2.710	0.3730	
ziji,LD,Heritage - ziji,LD,L2	-1.599885	0.352	Inf	-4.549	0.0008	*
taziji,LD,Heritage - taziji,LD,L2	-1.538306	0.302	Inf	-5.094	0.0001	*
ta,local,Heritage - ta,local,L2	-0.565951	0.368	Inf	-1.540	0.9878	
ziji,local,Heritage - ziji,local,L2	1.783589	0.509	Inf	3.504	0.0476	*
taziji,local,Heritage - taziji,local,L2	1.550904	0.477	Inf	3.253	0.1026	
ta,LD,Native - ta,local,Native	8.088864	0.548	Inf	14.763	<.0001	*
ziji,LD,Native - ziji,local,Native	-0.431138	0.155	Inf	-2.774	0.3299	
taziji,LD,Native - taziji,local,Native	-7.018500	0.605	Inf	-11.608	<.0001	*
ziji,LD,Native - taziji,LD,Native	2.455286	0.199	Inf	12.335	<.0001	*

[47] For more analyses of *ziji* on the NS data reported here, see C. Chen (2020b). When only NS data is analyzed, there is a marginally significant difference between LD and local readings. When divided by dialect, NSs of mainland Mandarin allowed numerically (but not significantly) more LD readings than local readings, while NSs of Taiwanese Mandarin allowed significantly more local readings than LD readings.

Table 7.4. Pairwise comparison results from the logistic mixed-effects model on anaphors (cont'd)

contrast	estimate	SE	df	z.ratio	p.value	
ziji,local,Native - taziji,local,Native	-4.132077	0.591	Inf	-6.997	<.0001	*
ta,LD,Heritage - ta,local,Heritage	6.988695	0.534	Inf	13.092	<.0001	*
ziji,LD,Heritage - ziji,local,Heritage	-7.152072	0.545	Inf	-13.125	<.0001	*
taziji,LD,Heritage - taziji,local,Heritage	-6.531949	0.486	Inf	-13.450	<.0001	*
ziji,LD,Heritage - taziji,LD,Heritage	-0.433376	0.384	Inf	-1.128	0.9997	
ziji,local,Heritage - taziji,local,Heritage	0.186747	0.614	Inf	0.304	1.0000	
ta,LD,L2 - ta,local,L2	5.003013	0.338	Inf	14.823	<.0001	*
ziji,LD,L2 - ziji,local,L2	-3.768599	0.273	Inf	-13.814	<.0001	*
taziji,LD,L2 - taziji,local,L2	-3.442739	0.266	Inf	-12.941	<.0001	*
ziji,LD,L2 - taziji,LD,L2	-0.371798	0.224	Inf	-1.662	0.9735	
ziji,local,L2 - taziji,local,L2	-0.045938	0.303	Inf	-0.151	1.0000	

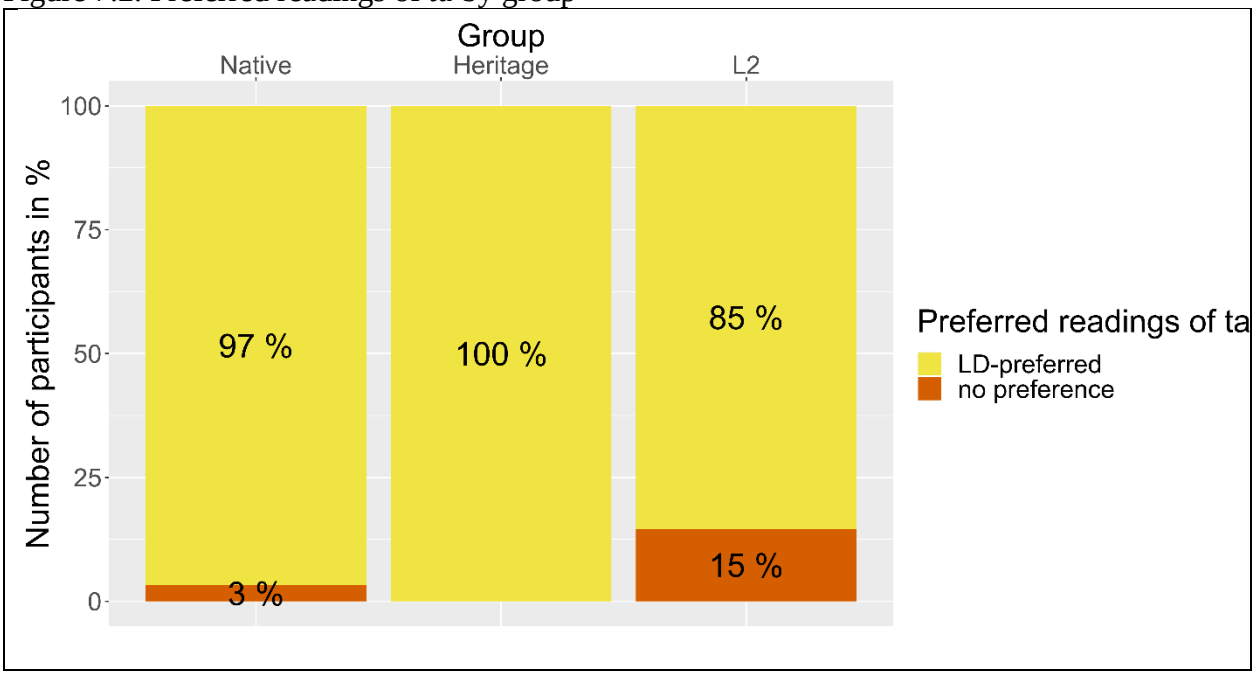
Note: * indicates p<.05

7.4. Individual subjects' analysis

At the group level, HSs show categorical judgment, treating *ziji* and *taziji* as strictly having only local readings. However, L2ers appear to accept some LD readings for *ziji* and *taziji*. To confirm whether such LD readings (about 20%) result from just a few L2ers' strong acceptance or many L2ers' weak acceptance, I conducted an individual subjects' analysis by classifying participants into one of three patterns: those having an LD preference, a local preference, or no preference. A common cutoff for consistent or systematic responses was set at 75% (e.g., three out of four tokens in Thomas, 1995). If a participant accepted at least 75% (i.e., 83% here) LD readings (namely, five or six out of the six tokens), and accepted at most 50% local readings (namely, at most three out of the six tokens), they were classified as having an LD preference. If a participant accepted at most 50% LD readings (at most three out of the six tokens), and accepted at least 75% local readings (five or six out of the six tokens), they were classified as having a local preference. If a participant accepted the same number of tokens from LD and local readings, or differed by just one or two token differences (e.g., five or six LD vs. four local readings), they were classified as having no preference.

Figure 7.2 through Figure 7.4 show the results of the individual subjects' analyses. As seen in Figure 7.2, while all HSs and most NSs and L2ers showed an LD preference for *ta*, two NSs (3%) and six L2ers (15%) showed no preference. Figure 7.3 shows the variable judgment of *ziji* by NSs, with no preference being more common than having an LD or local preference.[48] By contrast, Figure 7.4 shows that all but two NSs had a local preference of *taziji*. Figure 7.3 and Figure 7.4 show that HSs had a local preference for reflexives at the individual level (all HSs showed a local preference for *ziji*, while all but one (2%) showed a local preference for *taziji*). Most L2ers (about 80%) also showed a local preference of *ziji* and *taziji*, though almost 20% had either an LD preference or no preference. Note that none of the NSs nor HSs had an LD preference for *taziji*.

Figure 7.2. Preferred readings of *ta* by group



[48] When grouped by dialects, 36% of the NSs of mainland Mandarin showed an LD preference, 50% showed no preference, and 14% showed a local preference. For NSs of Taiwanese Mandarin, 15% showed an LD preference, 42% showed no preference, and 42% showed a local preference.

Figure 7.3. Preferred readings of *ziji* by group

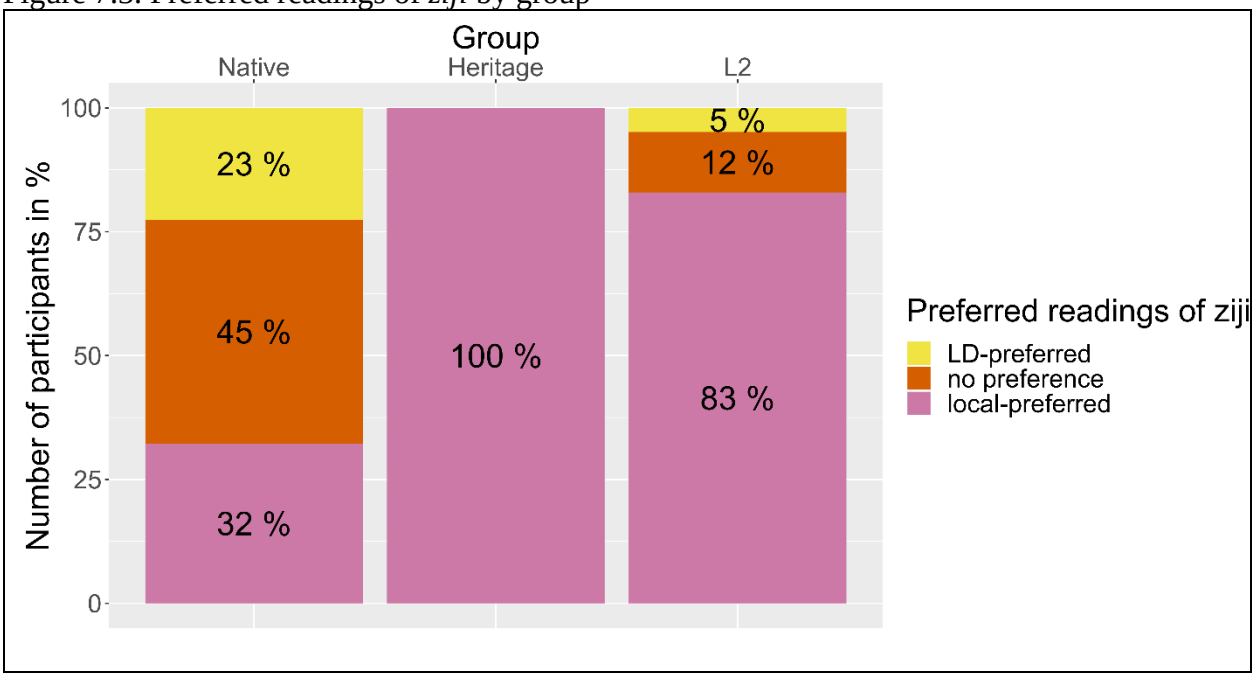
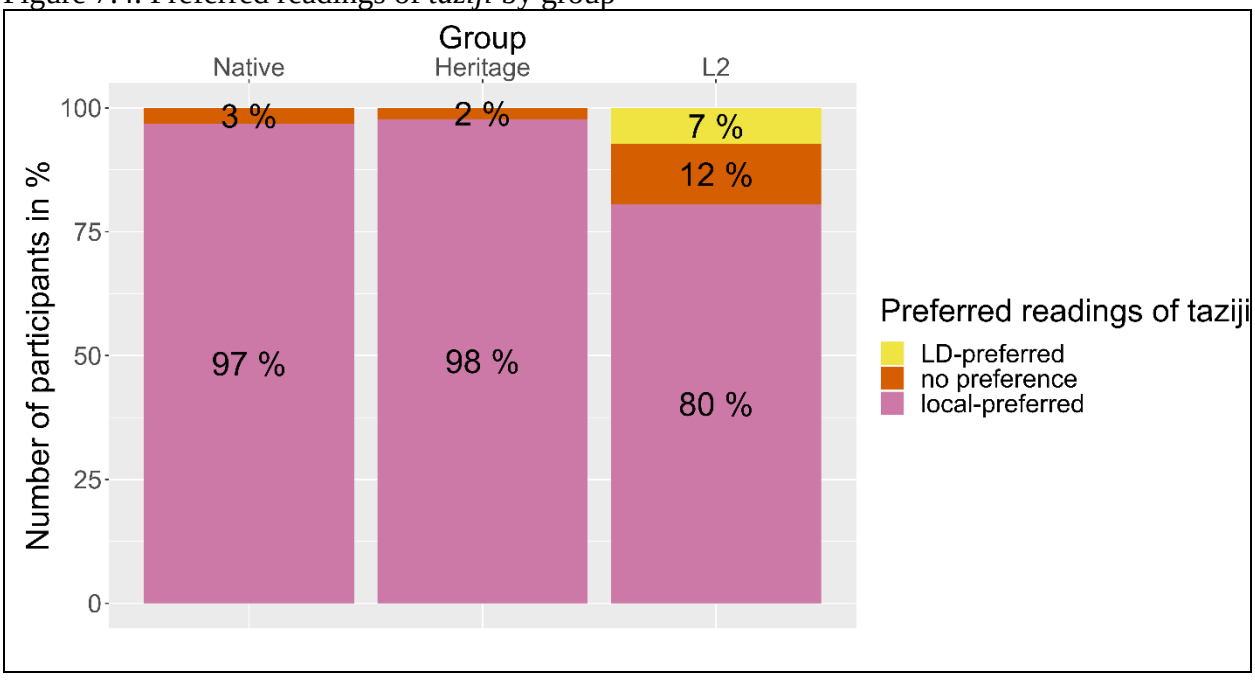


Figure 7.4. Preferred readings of *taziji* by group



To examine the Mandarin proficiency effect on the readings of the anaphors in more detail, I look at the mean scores (max score = 40) for the three pattern groups (LD-preferred, local-preferred, no preference); see Table 7.5. Recall that the average scores were 27.1 for HSs and 29.5

for L2ers (see Table 4.1 in Chapter 4). I begin with those who showed non-target-like responses for *ta* and *taziji*: The L2ers who accepted LD readings of *taziji* and those who showed no preference for *ta* and *taziji* had proficiency scores similar to the average. The only HS who showed no preference of *taziji* scored almost at ceiling on the proficiency test. For *ziji*, while all three patterns are native-like (since NSs exhibited all three patterns), those who showed an LD or no preference had potentially acquired LD readings: the two L2ers who had an LD preference had very high proficiency scores, while the five L2ers who showed no preference of *ziji* had proficiency scores similar to the average.

Table 7.5. Number of participants and the mean scores (max score = 40) for each pattern group of the anaphors

Anaphors	Patterns	Native speakers (*N* = 62)	HSs (*N* = 44)	L1-English L2ers (*N* = 41)
Ta	LD	60	44 (mean: 27)	35 (mean: 29.9)
	No preference	2	0	6 (mean: 27.2)
	Local	0	0	0
Ziji	LD	14	0	2 (mean = 38)
	No preference	28	0	5 (mean: 26.4)
	Local	20	44 (mean: 27)	34 (mean: 29.5)
Taziji	LD	0	0	3 (mean: 26.3)
	No preference	2	1 (score: 39)	5 (mean: 27.4)
	Local	60	43 (mean: 26.8)	33 (mean: 30.2)

7.5. Discussion

I discuss *taziji* before focusing on *ziji* and compare them to previous studies on HSs and L2ers for this phenomenon. Similar to my previous study (C. Chen, 2019b), HSs patterned with NSs on *taziji*, while L2ers over-accepted LD readings and under-accepted local readings. [49] (To be accurate, the between-group difference on the local readings of *taziji* was not significant in C.

[49] Even though there is a significant difference between NSs and L2ers on the local readings of *ta* (p = 0.02), the acceptance rate by L2ers is still very low and therefore does not warrant discussion.

Chen, 2019b, but this is likely due to a smaller number of participants, since the numerical differences between L2ers and other groups are actually larger in C. Chen, 2019b, compared to the differences in this dissertation.) Given that English reflexives have only local readings, the L2ers' pattern cannot be a result of English transfer. Previously, Zeng (2010) and C. Chen (2019b) speculated that it might be due to a misanalysis of reflexives as pronouns. However, since L2ers in this dissertation still accepted local reading of *taziji* at 90%, they did not appear to misanalyse *taziji* as *ta*, at least not at the group level. Since only three L2ers (7%) showed an LD preference and five L2ers (12%) showed no preference, it may be the case that these L2ers misanalysed *taziji* as *ta*, because they interpreted *ziji* as an adnominal intensifier (e.g., Hole, 2008) or an adverbial (Tsai, 2019). When *taziji* appears in the subject position, it indeed means *ta* plus an adverbial or intensifier *ziji*. Given the data, there is no evidence that such misanalyses were made by L2ers with lower Mandarin proficiency. With more participants in the future, the proficiency effect could be re-examined to see if only lower-proficient learners make such misanalysis.

For *ziji*, this dissertation also replicated the results from C. Chen (2019b) in that neither HSs nor L2ers had acquired LD readings for *ziji*. In this dissertation, both HSs and L2ers differed significantly from NSs, and the two groups also differed significantly from each other. While NSs did not accept LD and local readings of *ziji* at ceiling, it is clear that neither HSs nor L2ers had acquired LD reading of *ziji*. While L2ers were more native-like than HSs were in LD readings of *ziji*, the fact that they over-accepted LD readings of *taziji* cautions against a conclusion that they have acquired LD readings better than HSs.

There are different explanations as to why HSs and L2ers might find LD readings of *ziji* difficult to acquire. These include English transfer, (presumably) low frequency, processing difficulty, and the interface property of *ziji*. The difficulty of processing LD readings of *ziji* is well-

documented from monolingual adult NSs (e.g., Jäger et al., 2015; Dillon et al., 2016; but see Lu, 2011, for an opposite finding). While Polinsky and Scontras (2020) argue that difficulty in processing LD dependencies, including LD reflexives and ORCs, explains HSs' non-target-like performance, it is difficult to address the role of processing in the present dissertation because only an offline TVJT was used to test anaphors (see Gračanin-Yuksek, et al., 2020, which used both offline and online tasks and found that Turkish HSs maintained LD readings of reflexives in their heritage language). Low frequency of LD readings of *ziji* (Lu, submitted; but see L. Liu, 2010, for an opposite finding) is a possible candidate for this learning difficulty, though studies that directly examine frequency effect (e.g., Slabakova, 2015a; Hopp et al., 2020) are needed. As will be discussed below, transfer and interface explanations were not well supported by other studies.

To address the role of L1 transfer, using the same TVJT employed here, C. Chen (2019a; C. Chen & Ionin, in preparation) additionally tested L1-Korean L2-Mandarin learners and compared them with L1-English L2ers. (The NSs and a smaller set of L1-English L2ers reported in this dissertation were included in that study.) The Korean group was predicted to outperform the English group since Korean allows LD readings of simplex reflexives *caki* and *casin*. However, results showed that proficiency-matched L2 groups predominantly allowed only local readings of *ziji* (though the Korean group allowed numerically more LD readings than the English group). Thus, English transfer might not be the sole reason for the local preference found in this dissertation.

Under the Interface Hypothesis (e.g., Tsimpli & Sorace, 2006), interface phenomena are more difficult for L2ers to acquire (for extension to HSs, see Montrul & Polinsky, 2011). However, the updated Interface Hypothesis (Sorace, 2011) states that only external interfaces but not internal interfaces present persistent challenges. The typical test case for an external interface is the null

vs. overt subject pronouns (e.g., Margaza & Gavarró, 2020 on L2-Spanish and L2-Greek; Rodríguez-Ordóñez & Sainzmaza-Lecanda, 2018, on L2-Basque). However, it is less clear whether LD reflexives lie at the external or internal interface (but see J.-H. Kim, 2013, who examined the updated Interface Hypothesis by testing Korean HSs on the logophoricity of *caki-casin* and pronoun-*casin*). Following C.-T. J. Huang and C.-S. L. Liu (2001), the LD reading of *ziji* is a logophor at the syntax-discourse interface. Therefore, the updated Interface Hypothesis applies and can explain the difficulty in acquiring LD readings of *ziji*. However, as discussed in Chapter 2, it remains unclear how to differentiate between a syntactic anaphor and a pragmatic logophor. Given that this dissertation tested *ziji* at the direct object position, the LD readings of *ziji* are logophors under C.-T. J. Huang and C.-S. L. Liu (2001), but anaphors under Reinhart and Reuland (1993). More studies are needed to draw a firm conclusion.

Given that neither HSs nor L2ers had acquired LD readings of *ziji*, there is no HS advantage. This lack of HS advantage is predicted given the late age of acquisition (AoA) in L1 acquisition (by at least age eight) for this structure. English transfer, presumably low frequency, processing difficulty, and for HSs, the late AoA, are all possible explanations for why the LD reading of *ziji* is difficult to acquire. Given the data here, it is difficult to tease apart these different reasons. For NS preference, see C. Chen (2020b) for further analysis on the NS data reported here (e.g., dialects and trial order effect). While dialectal differences seem to exist on the readings of *ziji* (see C. Chen, 2020b), given that HSs clearly only allowed local readings of *ziji*, I did not further separate them into mainland Mandarin or Taiwanese Mandarin HS groups and did not further examine for proficiency effects. Another possible source of the NS variation is the influence of English, as S. Zhang (2018) found L1-attrition after long-time contact with English. It is possible that some Mandarin NSs (typically first-generation immigrants) may have already lost LD readings of *ziji*

by the time they had children in an English-speaking country. Thus, HSs might have encountered very few instances of LD readings from their parents while growing up. Studies that examine the language patterns of both parents and children might be able to address this issue.

Lastly, I discuss the different findings in heritage Mandarin, Korean, and Turkish. Recall that the HSs and L2ers of Mandarin reported in C. Chen (2019b) and in this dissertation did not acquire LD readings of reflexives. In contrast, HSs of Turkish (Gračanin-Yuksek, et al., 2020) and HSs of Korean (and L1-English L2-Korean learners) acquired LD readings of reflexives (J.-H. Kim et al., 2009, 2010; S. Y. Lee, 2012). (In J.-H. Kim et al. (2009, 2010), HSs' acceptance rate of LD readings, though significantly lower than Korean NSs, was over 75%). One possible reason for this is that Korean *caki* has a clear preference towards LD readings by Korean NSs (LD: >90% vs. local: <40%; J.-H. Kim et al., 2009, 2010; S. Y. Lee, 2012), while Mandarin *ziji* does not have such a strong preference. With Turkish reflexives *kendi* and *kendisi*, it seems that both LD and local readings are easily available, at least in an antecedent selection task (NSs in Gračanin-Yuksek et al., 2017; HSs in Gračanin-Yuksek et al., 2020).[50] Turkish NSs accepted LD readings at 85% for *kendi* and 96% for *kendisi*, and accepted local readings at 94% for *kendi* and 87% for *kendisi.* For Mandarin *ziji*, whether LD or local readings are preferred in final interpretations is still not certain. While some studies have reported a higher acceptance of local readings by NSs (e.g., Chien, Wexler, & Chang, 1993; the NSs of Taiwanese Mandarin in this dissertation), others have not (e.g., C. Chen 2019b; Zeng, 2010; the NSs of mainland Mandarin in this dissertation). However, based on psycholinguistic evidence, the local reading of *ziji* is the default option in real-time processing (see Dillon, 2014, for an overview of reflexive processing). In any case, while Mandarin has LD

[50] In the antecedent selection task, participants were asked to choose a local, an LD, and/or extra-sentential antecedent after reading isolated sentences containing reflexives. Given that multiple choices per trial were allowed, it is not surprising that the acceptance rate was so high.

reflexives as do Korean and Turkish, NSs of Mandarin seem to accept LD reflexives to a lower degree than NSs of Korean and possibly also NSs of Turkish. This lowered acceptance might have contributed to the difficulty of acquiring LD readings of *ziji* by HSs and L2ers of Mandarin.

Chapter 8. General discussion and conclusion

8.1. Summary and discussion of the findings

This dissertation aimed to answer two broad RQs, repeated below:

- Broad RQ 1: Can HSs and L2ers of Mandarin whose dominant language is English fully acquire the properties of Mandarin that are different from or absent in English?
- Broad RQ 2: Do HSs have selective advantages over proficiency-matched L2ers, and does this vary by linguistic domain?

I address these two broad RQs by examining four linguistic phenomena by conducting three offline tasks: the Tone Identification Task, the Acceptability Judgement Task (AJT), and the picture-based Truth Value Judgement Task (TVJT). The specific RQs for each linguistic phenomenon (T3 sandhi, aspect, RCs, and anaphors) are instantiations of the broad ones. The major findings from each of the four linguistic phenomena are summarized and discussed below.

The Tone Identification Task was designed to examine whether HSs and L2ers acquired T3 sandhi in the domain of phonology, despite lack of tones in English. Recall that the rule of T3 sandhi is that a T3 syllable becomes a T2 syllable if followed by another T3 syllable. Given that T3 sandhi is largely in place in monolingual children by age three and that phonology is known to be difficult for adult L2ers, it was hypothesized that HSs would have a clear advantage over L2ers in this domain. The results show that HSs in this task were indeed more native-like than L2ers, despite some undesired task effects that caused even NSs to not score at ceiling on the critical condition (but scored at ceiling on other conditions). Thus, the hypothesis that HSs would have an advantage over L2ers is supported, though not as strongly as expected. The only condition that might be related to English intonation transfer is T4T3, wherein L2ers did not perform well.

Given that each syllable must have a tone and that T3 sandhi is presumably very frequent, both HSs and L2ers should have plenty of experience hearing T3 sandhi. However, HSs still performed

more native-like than L2ers, which is consistent with previous studies. A widely-accepted explanation is that phonetics and phonology are subject to a biologically-determined sensitive, if not critical, period, which ends earlier than in other domains (e.g., Granena & Long, 2013). Due to decreasing brain plasticity, adult L2ers have difficulty acquiring tones and T3 sandhi, despite frequent occurrences in the input. Whether processing plays a role in T3 sandhi acquisition remains an open question, as previous studies have not actively linked processing to acquisition in this domain, which is the case with ORCs and LD reflexives.

The finding that HSs are more native-like than L2ers in tones and T3 sandhi is consistent with phonetics/phonology studies with other languages (e.g., Korean: Oh et al., 2003; Spanish: J. Y. Jim, 2020) or other phenomena in Mandarin (e.g., C. Chang et al., 2011; C. Chang & Yao, 2016; B. Yang, 2015). This is not surprising given that researchers have found that internationally adopted children who stopped hearing Chinese by age two were still able to retain some unconscious memory of tones (Pierce et al., 2014). As Pierce et al. (2019, p. 475) explains, phonology is "activated more frequently over the course of development than other linguistic elements" and is thus more resistant to language loss. While L2ers struggle with phonology, HSs are able to utilize the tones and T3 sandhi they acquired in their early childhood.

The fact that the HS advantage in tones and T3 sandhi was not as robust as expected may be due to task effects and/or a later acquisition of tone and T3 sandhi than previously thought. While early studies report that children acquire Mandarin tones by age two, a recent study by Wong and Strange (2017) reported that six-year-olds were still not adult-like. Similarly, while early studies report that children acquire T3 sandhi by age three (e.g., H. Zhu, 2002; Y.-H. Huang, 2006), recent acoustic studies have shown that it is fully attained after age five (e.g., P. Tang et al., 2019; Xu Rattanasone et al., 2018).

The second linguistic phenomenon examined in this dissertation was (grammatical) aspect, tested in the AJT. This AJT was designed to examine whether HSs and L2ers had acquired the interaction between grammatical aspect and lexical aspect. Recall that, unlike English *-ed*, Mandarin perfective marker *-le* with states and activities result in incomplete sentences; unlike English *-ing*, Mandarin progressive marker *zai* and durative marker *-zhe* are not compatible with achievements. Given that Mandarin aspect is largely acquired by age five in monolingual children, HSs are predicted to have a slight advantage over L2ers in this domain. Additionally, given that the Mandarin aspect marking tested in the present dissertation involves morphology, it is predicted to pose some challenges to L2ers according to the Bottleneck Hypothesis (Slabakova, 2008, 2014) whereby inflectional morphology is challenging for L2ers. However, aspectual morphology may also be difficult for HSs (Montrul, 2018, who extends the Bottleneck Hypothesis to HSs in general; Mikhaylova, 2018, on Russian aspect). The results showed that HSs in this task were more native-like than L2ers were and seemed to be less influenced by English transfer of *-ing*. Thus, the hypothesis is also supported.

Note that many verbs are not marked with aspect markers (e.g., more than 60% in L. Jin & Hendriks, 2005 in elicited story telling; more than 70% in child-directed speech and more than 95% in adult-to-adult speech in C.-C. Huang, 2003). Additionally, different aspect markers tend to co-occur with different lexical predicates. This lack of consistency in aspectual marking and English transfer might make it difficult for both L2ers and HSs to acquire, but HSs have a slight advantage due to early AoA. After all, other devices such as adverbials or modal verbs are often available to help learners comprehend the temporal meaning. Like T3 sandhi, whether processing in aspect plays a role in acquisition remains an open question since previous literature has not actively linked processing to acquisition in this domain. However, it could potentially explain why

HSs and L2ers gave higher ratings to the more frequent/prototypical combinations (e.g., *zai* with activities; *-le* with accomplishments) than other grammatical combinations. The non-target-like performance on aspect occurs with HSs of other languages such as Spanish (Montrul, 2002) and Russian (Mikhaylova, 2012, 2018), though Spanish and Russian are morphologically more complex than Mandarin and pose different problems to learners.

The third linguistic phenomenon examined in this dissertation is RCs, tested in the AJT and the TVJT. This AJT was designed to examine if HSs and L2ers knew that Mandarin RCs are head-final, unlike English RCs which are head-initial. The TVJT was designed to examine if they had correct interpretations of Mandarin RCs when animacy cues were removed. As RCs are typically acquired by age five in monolingual children, HSs are predicted to have acquired RCs. L2ers are also predicted to have acquired Mandarin RCs successfully since Mandarin RCs fall in the domain of syntax (word order), which is known to be easy for L2ers to acquire. Thus, the hypothesis is that HSs would pattern similarly with L2ers on both tasks and acquire RCs successfully. The results show that both groups largely accepted head-final RCs and rejected head-initial RCs in the AJT, indicating that they overcame English transfer on RC headedness. For interpretations tested in the TVJT, however, the accuracy dropped. While HSs scored numerically higher than L2ers, no significant differences were found between HSs and L2ers. Taken together, the hypothesis is still supported in not finding an HS advantage.

While RCs are presumably frequent, RCs with two animate nouns (which the TVJT used) are very infrequent (for corpus findings, see F. Wu et al., 2012, for a summary). Despite low frequency, most HSs and L2ers were able to interpret RCs with two animate nouns successfully. In terms of processing, Polinsky and Scontras (2020) termed HSs' difficulty of processing ORCs, relative to SRCs, as "the distance problem" because of the larger distance between the head and

the gap/verb in ORCs. However, whether an SRC is easier to process than an ORC remains controversial in Mandarin NSs (SRC advantage: Jäger et al., 2015 vs. ORC advantage: Gibson & H.-H. I. Wu, 2013, among many others) and L2ers (SRC advantage: Xu, 2014c vs. ORC advantage: Y.-T. Sung et al., 2016, among many others). In the present dissertation, only L2ers, but not HSs, showed an SRC advantage in matching (but not mismatching) conditions in the TVJT. Thus, like the inconclusive results in the literature, while a clear SRC advantage exists in languages like English due to multiple reasons (e.g., higher frequency, shorter distance between the head noun and the gap/verb), the SRC/ORC asymmetry in Mandarin is not as clear, because different factors that influence acquisition/processing point to opposing directions and render an SRC or ORC advantage difficult to detect. From a typological perspective, the fact that Mandarin has a rare combination of mixed headedness (SVO word order with head-final RCs) may explain why it does not have a clear SRC advantage like most languages do.

The relatively successful acquisition of RCs differs from other studies focusing on heritage Russian (Polinsky, 2011) and heritage Korean (O'Grady et al., 2001; T. Lee, 2016), both involving case morphology to interpret the RCs. English-dominant HSs and L2ers may have a relatively easier time acquiring Mandarin RCs compared to acquiring Russian and Korean RCs because acquiring new RC headedness in Mandarin is easier than acquiring case marking systems on common nouns in Russian and Korean. (Certainly, to acquire Korean RCs, new RC headedness also needs to be learned by English speakers).

Lastly, the dissertation examined anaphors in the TVJT, with a focus on the simplex reflexive *ziji*. Recall that unlike English, *ziji* allows both LD and local readings. The TVJT was designed to test if HSs and L2ers would allow both LD and local reading of *ziji*, while only allowing LD readings of the pronoun *ta* and local readings of the simplex reflexive *taziji*. Given that

monolingual children acquire LD readings of *ziji* after age eight and that such readings are at the interface of syntax and semantics/discourse, it was hypothesized that HSs would not have an advantage over L2ers because neither group will have acquired LD readings of *ziji.* The results show that both groups indeed only allowed local readings of *ziji*, supporting the hypothesis.

However, English transfer is only one possible reason for why HSs and L2ers only accepted local readings of reflexives. Other possible reasons include the interface properties (the Interface Hypothesis, Sorace, 2011; Tsimpli & Sorace, 2006) or local readings as the default option, which in turn could be due to the processing difficulty and the presumably very low frequency of LD readings. Relative to locally bound reflexives, the distance between the LD antecedent and the anaphor is longer. Difficulty with LD reflexives constitutes another example of "the distance problem" described in Polinsky and Scontras (2020), which cites Korean anaphors as an example. Reducing both LD and locally bound reflexives to just local reflexives is also consistent with one of the common HS outcomes: "shrinking of structure" (Polinsky & Scontras, 2020, again citing Korean anaphors as an example).

Unlike studies in heritage/L2 Korean (J.-H. Kim et al., 2009, 2010; S. Y. Lee, 2012) and heritage Turkish (Gračanin-Yuksek, et al., 2020), which reported successful acquisition of LD readings of reflexives, HSs and L2ers in the present dissertation did not. Compared to Korean, Mandarin *ziji* does not have a strong LD bias as Korean *caki* does. This may explain why it is more difficult for English-dominant HSs and L2ers to acquire LD readings of *ziji* in Mandarin than LD readings of *caki* in Korean.

Table 8.1 below summarizes the results of the four linguistic phenomena, expanding on Table 2.11 from Chapter 2.

Table 8.1. Summary of the results (> means 'significant difference between'; ≥ means 'almost reached significance'; = means 'no significant difference)

Topics	Linguistic domain	Age of acquisition in monolingual children	Tested in …	Selective advantage?
Tone 3 sandhi	Phonology	2-3	Tone Identification Task	Yes, HSs were slightly more native-like than L2ers NS = HS, NS ≥ L2, HS = L2 in the critical T3T3 condition
Aspect	Morpho-semantics interface	most at 3, some by 5	AJT	Yes, HSs were more native-like than L2ers NS = HS = L2 in states with *-le*; NS = HS, NS > L2, HS = L2 in achievements with *zai*; NS = HS, NS ≥ L2, HS = L2 in achievements with *-zhe*
RCs	Syntax	3-4, stabilize at 5	AJT and TVJT	Yes, HSs were more native-like than L2ers in the AJT, and slightly more native-like than L2ers in the TVJT NS = HS = L2 in the head-initial SRC/ORC conditions in the AJT NS = HS, NS > L2, HS = L2 in the head-final SRC condition in the AJT NS > HS, NS > L2, HS = L2 in the head-final ORC condition in the AJT NS = HS = L2 in the SRC-matching condition in the TVJT NS > HS, NS > L2, HS = L2 in the ORC-matching condition in the TVJT NS ≥ HS, NS > L2, HS = L2 in the SRC-mismatching condition in the TVJT NS ≥ HS, NS > L2, HS ≥ L2 in the ORC-mismatching condition in the TVJT
LD reflexives	Syntax-semantics interface	At least 8	TVJT	No NS > HS, NS > L2, HS > L2 in the LD readings of *ziji* condition in the TVJT (L2ers were more native-like than HSs)

To summarize, the results of this dissertation are partially consistent with the original hypotheses, though the advantage on T3 sandhi is not as strong as initially predicted and there seems to be a slight HS advantage on RCs. Within the same group of HSs and L2ers, HSs have a weak advantage over L2ers in T3 sandhi because monolingual children acquire T3 sandhi by age three, and phonology is difficult for L2ers to acquire despite frequent occurrences of T3 sandhi. For aspect, HSs have an advantage over L2ers in T3 sandhi because monolingual children acquire aspect by age five, and morphology-related linguistic phenomena are difficult for L2ers to acquire. Aspect markers are not as frequent nor consistent as tense/aspect markers in English and may present challenges to learners. While RCs (syntax) are also acquired early by age five, I originally hypothesized that they would be acquired by L2ers without much difficulty, hence no HS advantage. However, there seemed to be a slight HS advantage in the AJT, though the evidence was not strong. Despite low frequency of RCs with two animate nouns tested in the TVJT, HSs and L2ers were able to give correct answers most of the time. LD readings of reflexives are challenging to acquire for multiple reasons: late acquisition by monolingual children (by at least age eight), interface properties, local reading as the default option (including presumably low frequency and processing difficulty of LD readings of reflexives).

Related to this, the frequency of LD reflexives that HSs and L2ers were exposed to is likely lower than what was reported in the corpus studies (L. Liu, 2010; Lu, submitted). S. Zhang (2018) found that Mandarin NSs undergo L1 attrition of LD readings of reflexives, but not perfective/durative aspect marking (also see S. Zhang, 2020), after living in an English-speaking environment for an average of 13 years. If this is a common scenario, given that these L1 attriters often constitute the input for HSs, the input HSs receive is qualitatively different from the homeland varieties, making it even harder for them to acquire LD readings of reflexives. Similarly,

beginner L2ers often acquire Mandarin from English-Mandarin bilinguals, thereby also receiving input from bilinguals who may be heavily influenced by English. If so, the input they receive may also contain fewer LD reflexives than the homeland varieties.

8.2. Limitations and future directions

One major limitation is that the proficiency measure adopted in this dissertation puts naturalistic learners (including some HSs) at a disadvantage. If Romanization had been added, the proficiency scores of HSs might have been higher (D. Zhang et al. 2019); thus, the proficiency of HSs in this dissertation might have been underestimated. Even so, they still performed similarly to L2ers in LD reflexives, which provides firm evidence that there is no HS advantage on this linguistic phenomenon. To remedy the possible underestimation, I added another proficiency test with Pinyin after the data collection began, but this was only taken by some of the participants. A re-examination of the data based on this other proficiency test might reveal similar or different patterns.

Another limitation is the unexpected results from the NSs on the T3 sandhi condition and some conditions on aspect. The fact that NSs did not choose T2T3 for the T3 sandhi condition indicates that the test design can be improved. A forced-choice task (between T2T3 and T3T3) is one possibility though the task would become very explicit. Analyzing the data from the production task will provide complementary evidence for the perception task reported here.

Some unexpected NS results also appeared in the AJT which tested aspect. It is especially difficult to judge the acceptability of (seemingly) incomplete sentences, as seen in the individual analyses: even NSs gave variable judgements. Based on informal post-test feedback, some participants, even NSs, would consider “I love-*le* him” acceptable, but when I asked them to

compare the sentence with *-guo* and/or add a time phrase such as “three years”, most of them immediately confirmed that the latter two were much better. Thus, other context-based tasks are needed to complement the sentence-level AJT employed here. Another limitation is that many of the state verbs were used repeatedly, as I intended to use only verbs that were classified as such without controversy.

For future directions, given that the language background questionnaire was very detailed, the data presented here can be further analyzed by adding additional biographical predictors. Continuous variables include the amount of Mandarin input/exposure or Mandarin proficiency, to name just two. For example, the amount of Mandarin instruction can be approximated by summing up the time of instruction HSs and L2ers have received while the amount of naturalistic input/exposure can be approximated by taking into account the time spent in Mandarin-speaking countries. Binary variables, for example, could include whether HSs and L2ers are instructed learners or naturalistic (i.e., uninstructed) learners. In addition, given that RCs are usually explicitly taught in Chinese classrooms, these instructed learners might outperform naturalistic learners in the RCs tested in the form-based AJT, which requires more metalinguistic awareness, but may not in the meaning-based TVJT, which requires less metalinguistic awareness.

For HSs in particular, age of onset of bilingualism (i.e., AoA of English in this case) can be modeled as a continuous variable or as a binary variable (sequential vs. simultaneous bilinguals), given enough HSs who differ in the AoA of English. HSs who are exposed to English later (sequential bilinguals) are expected to outperform those who are exposed to English earlier (simultaneous bilinguals) in Mandarin, following previous findings with Spanish HSs (Montrul, 2002) and Mandarin HSs (L. Jia & Bayley, 2008) on grammatical aspect.

Another issue is the source of the divergent performance between NSs and HSs/L2ers. Some may question whether English transfer is indeed the reason for such non-native-like performance. It is a valid question/critique and is best answered with evidence from another L1 group. Data from an L1-Korean group has been collected using the same TVJT and is reported in C. Chen (2019a, 2019c; C. Chen & Ionin, in preparation). Recall that this dissertation found that the acquisition of Mandarin head-final RCs by English-dominant HSs and L2ers was not hindered by head-initial RC in English, which is likely due to word order being easy to acquire. In C. Chen (2019c), proficiency-matched L1-English L2-Mandarin learners outperformed L1-Korean L2-Mandarin learners in Mandarin RCs, indicating that having to learn a different way of interpreting RCs (from Korean case marking to Mandarin word order) is more difficult than learning new RC headedness (from English head-initial RCs to Mandarin head-final RCs). In C. Chen (2019a) and Chen and Ionin (in preparation), the fact that L1-Korean L2-Mandarin learners predominantly allow only local readings of *ziji* (despite having LD reflexives in Korean) shows that local readings as the default option might outweigh L1 transfer in L2 acquisition. Recruiting another L1 group to take the AJT will provide further evidence to see, for example, if accepting progressive *zai* with achievements result from English transfer or whether it is common in L2-Mandarin regardless of a learner's native language(s).

8.3. Concluding remarks

To conclude, this dissertation contributes to language acquisition theories by comparing English-dominant HSs and L2ers of Mandarin. Unlike Spanish (where most HS/L2 comparisons have been done), Mandarin is typologically distant from English, and generally difficult for English speakers to acquire. Similar to findings in other languages, AoA, linguistic domains, and dominant language transfer all contribute to the acquisition outcomes of Mandarin HSs, compared

to adult L2ers. Similar to other heritage phonology studies that find an HS advantage (e.g., Korean: Oh et al., 2003; Spanish: J. Y. Jim, 2020), the present dissertation finds an HS advantage on T3 sandhi, albeit smaller than expected. The non-target-like performance on aspect occurs with HSs of other languages such as Spanish (Montrul, 2002) and Russian (Mikhaylova, 2018), though Spanish and Russian are morphologically more complex languages and pose different problems to learners. Like HSs of other languages such as Spanish (Montrul, 2002) and Russian (Mikhaylova, 2012, 2018), Mandarin HSs were not completely native-like on aspect, but outperformed L2ers (as in Mikhaylova, 2012). Different from findings in heritage/L2 Korean (e.g., O'Grady et al., 2001; T. Lee, 2016) and heritage Russian (Polinsky, 2011), Mandarin RCs appear to be acquired by both Mandarin HSs and, to a lower degree, L2ers. The different findings on RCs can be attributed to different ways of marking SRCs and ORCs cross-linguistically (e.g., case marking in Koran and Russian). Unlike findings in heritage/L2 Korean (J.-H. Kim et al., 2009, 2010; S. Y. Lee, 2012) and heritage Turkish (Gračanin-Yuksek et al., 2020), neither Mandarin HSs nor L2ers acquired LD readings of reflexives.

These findings will benefit the field on the relative contributions of these factors to the acquisition of several distinct linguistic domains. In addition to this central contribution, this dissertation contributes to the debate on the SRC/ORC advantage in acquisition and provides a full picture of all aspect markers with all lexical predicates. Furthermore, the methodology adopted in this dissertation makes a critical and timely contribution to the study of heritage/L2 acquisition in testing multiple domains with the same group of participants. Given that this dissertation examines a number of phenomena that differ in English and Mandarin, this dissertation has potential implications for language pedagogy (e.g., drawing learners' attention to the different predicates in Mandarin and the word order in RCs), which is particularly important given both the growing

interest in learning Mandarin in the United States and the growing population of Mandarin HSs in college classrooms. While this dissertation is not itself pedagogical in nature, it helps identify problem areas facing learners of Mandarin, and provides information about where HSs do or do not face difficulties relative to traditional L2 classroom learners. These findings in turn may potentially inform pedagogical studies and, eventually, lead to design of HS-specific teaching materials.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249-306.

Andersen, R. W., and Shirai, Y. (1996). The primacy of aspect in first and second language acquisition: The pidgin-creole connection. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 527–570). San Diego: Academic Press.

Au, T. K. F., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological science*, *13*(3), 238-243.

Au, T. K. F., Oh, J. S., Knightly, L. M., Jun, S. A., & Romo, L. F. (2008). Salvaging a childhood language. *Journal of memory and language*, *58*(4), 998-1011.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3-4), 129-181.

Chan, A., Chen, S., Matthews, S., & Yip, V. (2017). Comprehension of subject and object relative clauses in a trilingual acquisition context. *Frontiers in psychology*, *8*, 1641.

Chang, C. B. (to appear). Phonetics and phonology. In S. Montrul & M. Polinsky (Eds.), *The Cambridge Handbook of Heritage Languages and Linguistics*. Cambridge, UK: Cambridge University Press.

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *The Journal of the Acoustical Society of America*, *138*(6), 3703-3716.

Chang, C. B., & Yao, Y. (2016). Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal*, *13*(2), 134-160.

Chang, C. B., & Yao, Y. (2019). Production of neutral tone in Mandarin by heritage, native, and second language speakers. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 2291-2295).

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America*, *129*(6), 3964-3980.

Chang, L.-P. (2012). The study of the vocabulary size at the CEFR levels for CFL/CSL learners (對應於歐洲共同架構的華語詞彙量). *Journal of Chinese Language Teaching (華語文教學研究), 9*(2), 77-96.

Chang, L.-P. (2017). The acquisition of relative clauses in L2 Chinese: a corpus-based study (華語學習者關係子句的習得考察─ 以語料庫為本的研究). *Journal of Chinese Language Teaching (華語文教學研究,)14*(1), 47-80.

Chang, Y. H. S. (2011). Distinction between mandarin tones 2 and 3 for L1 and L2 Listeners. In *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)* (Vol. 1, pp. 84-96).

Chao, Y. R. (1930). A system of tone letters. *Le Maitre Phonetique, 45*, 24-27.

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience*, *31*(6), 751-760.

Chen, C. (2017). *Second language processing of Mandarin relative clauses: A self-paced reading study*. Talk given at the Second Language Research Forum - SLRF 2017. Ohio State University, Columbus, OH.

Chen, C (2019a). *Second language acquisition of Mandarin binding by Korean and English native speakers.* Talk given at the XV Generative Approaches to Second Language Acquisition Conference (GASLA), March 22-24. University of Nevada, Reno, NV, USA.

Chen, C. (2019b). *The acquisition of Mandarin reflexives by heritage speakers and second language learners.* In M. Rispoli & T. Ionin (Eds.), *Three Streams of Generative Language Acquisition Research* (pp. 225-251). Language Acquisition and Linguistic Disorders Series. John Benjamins.

Chen, C. (2019c). *Second language acquisition of Mandarin relative clauses by Korean and English native speakers*. Poster resented at Second Language Research Forum 2019, September 20-22. Michigan State University, East Lansing, MI, Michigan.

Chen, C. (2020a). *The acquisition of Mandarin relative clauses and binding by heritage speakers and second language learners*. In M. M. Brown & A. Kohut (Eds.), *Proceedings of the 44th Boston University Conference on Language Development* (pp. 77-90). Somerville, MA: Cascadilla Press.

Chen, C. (2020b). *An experimental investigation of Mandarin reflexive ziji*. Talk given at the 32th North American Conference on Chinese Linguistics (NACCL-32), September 18-20. University of Connecticut, Storrs, CT, USA.

Chen, C & Ionin, T. (in preparation). Interpretation of Mandarin pronouns and reflexives by Korean and English native speakers. Manuscript in preparation.

Chen, D. (1995). Mandarin reflexive ziji in second language acquisition. *University of Pennsylvania Working Papers in Linguistics*, *2*(2), 5.

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

Chen, J., & Shirai, Y. (2015). The acquisition of relative clauses in spontaneous child speech in Mandarin Chinese. *Journal of child language*, *42*(02), 394-422.

Chen, Q. (1997). Toward a sequential approach for tonal error analysis. *Journal of Chinese language teachers association*, *32*, 21-39.

Chen, S., He, Y., Wayland, R., Yang, Y., Li, B., & Yuen, C. W. (2019). Mechanisms of tone sandhi rule application by tonal and non-tonal non-native speakers. *Speech Communication*, *115*, 67-77.

Chen, Z., Jäger, L., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. *Empirical approaches to linguistic theory: Studies of meaning and structure*, 43-62.

Cherici, A., Chang, Y.-Y., & Tanaka, N. (2019). *The production of relative clauses by L1-English learners of Chinese.* Paper presented at the NINJAL-UHM Linguistics Workshop on Syntax-Semantics Interface, Language Acquisition, and Naturalistic Data Analysis, Honolulu, HI.

Chien, Y. C., & Lust, B. (2006). Mandarin children's knowledge of the Binding Principles. In P. Li, L. H. Tan, E. Bates and O. J. L Tzeng (Eds.), *The handbook of east Asian psycholinguistics* (pp. 23-38). Cambridge University Press.

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

Chomsky, N. (1981). *Lectures on Government and Binding.* Dordrecht: Foris.

Christensen, R. H. B. (2018). Regression Models for Ordinal Data. R package version 2018.4-19. http://www.cran.r-project.org/package=ordinal/.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly journal of experimental psychology*, *69*(5), 817-828.

Christie, K., & Lantolf, J. (1998). Bind me up bind me down: Reflexives in L2. In S. Flynn, G. Martohardjono, & W. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 239 – 260). Mahwah, NJ: Erlbaum.

Cole, P., Hermon, G., & Huang, C.-T. J. (2006). Long-Distance binding in Asian languages. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 21-84). Blackwell Publishing.

Cole, P., Hermon, G., & Lee, C. L. (2001). Grammatical and discourse conditions on long distance reflexives in two Mandarin dialects. In P. Cole, G. Hermon, & C.-T. J. Huang (Eds.), *Long-distance reflexives* (pp.1-46). New York: Academic Press.

Cole, P., Hermon, G., & Sung, L. M. (1990). Principles and parameters of long-distance reflexives. *Linguistic inquiry*, *21*(1), 1-22.

Collart, A., & Chan, S. (2019). Aspect is distinct from time reference: An ERP study of the perfective marker-le in Mandarin Chinese. Poster presented at the 32nd Annual CUNY Sentence Processing Conference, March 29-31, University of Colorado Boulder.

Comrie, B. (2008). Prenominal relative clauses in verb-object languages. *Language and Linguistics*, *9*(4), 723-733.

Cowart, W. (1997). *Experimental syntax*. Thousand Oaks, CA: Sage.

Cui, Y. (2013). L2 processing of relative clauses in Mandarin. *Arizona Working Papers in SLA & Teaching*, *20*, 20-39.

Daskalaki, E., Chondrogianni, V., Blom, E., Argyri, F., & Paradis, J. (2019). Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research*, *35*(3), 421-445.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean?. In J. F. Kroll & A. M. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*,(pp. 88-108). Oxford University Press.

Dillon, B. (2014). Syntactic memory in the comprehension of reflexive dependencies: an overview. *Language and Linguistics Compass*, *8*(5), 171-187.

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*, 1025.

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*, 1966.

Domínguez, L., Hicks, G., & Slabakova, R. (2019). Terminology choice in generative acquisition research: The case of "incomplete acquisition" in heritage language grammars. *Studies in Second Language Acquisition*, *41*(2), 241-255.

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in second language acquisition*, *13*(04), 431-469.

Dryer, M. (2005). Order of subject, object, and verb. In M. Haspelmath, M. Dryer, D. Gil and B. Comrie (Eds.), *The world atlas of language structures (Chap. 96).* Oxford: Oxford University Press.

Dugarova, E. (2007). Acquisition of the Chinese reflexive 'ziji' by Russian and English speakers. In N. Hilton, R. Arscott, K. Barden, A. Krishna, S. Shah, & M. Zellers (Eds.), *Proceedings of CamLing 2007* (pp. 48–55). Cambridge: Cambridge Institute of Language Research.

Duanmu, S. (1999). Stress and the development of disyllabic words in Chinese. *Diachronica*, *16*(1), 1-35.

Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Dubinina, I., and M. Polinsky. 2013. Russian in the USA. In M. Moser & M. Polinsky (Eds.), *Slavic Languages in Migration* (pp. 131-160). Wien: University of Vienna.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143-188.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied linguistics*, *27*(1), 1-24.

Erbaugh, M. (1978). The Acquisition of Temporal and Aspectual Distinctions in Mandarin. In *Papers and Reports on Child Language Development* 15 Aug 30–37.

Erbaugh, M. S. (1992). The acquisition of Mandarin. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 3, pp. 373–455). Hillsdale, NJ: Erlbaum.

Felser, C. (2020). Do processing resource limitations shape heritage language grammars?. *Bilingualism: Language and Cognition*, *23*(1), 23-24.

Flecken, M., Walbert, K., & Dijkstra, T. (2015). 'Right Now, Sophie∗ Swims in the Pool?!': Brain Potentials of Grammatical Aspect Processing. *Frontiers in psychology*, *6*, 1764.

Flege, J. E., Yeni-Komshian, G., & Liu, H. (1999). Age constraints on second language acquisition. Journal of Memory and Language, *41,* 78-104.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Freund, N. (2016). Recent Change in the Use of Stative Verbs in the Progressive Form in British English: I'm loving it. *University of Reading Language Studies Working Papers*, *7*, 50-61.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of phonetics*, *11*(2), 149-175.

Gass, S. (1980). An investigation of syntactic transfer in adult second language learners. In R. Scarchella & S. Krashen (Eds.), Research in second language acquisition research forum (pp. 132-141). Rowley, MA: Newbury House.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76.

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image. language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., & Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, *28*(1-2), 125-155.

Gómez Soler, I. (2014). Beyond interfaces: Pragmatic development vs. syntactic deficiencies in the L2 acquisition of reverse psychological predicates. *Linguistic Approaches to Bilingualism*, *4*(4), 494-525.

Gračanin-Yuksek, M., Lago, S., Şafak, D. F., Demir, O., & Kırkıcı, B. (2017). The interaction of contextual and syntactic information in the processing of Turkish anaphors. *Journal of Psycholinguistic Research*, *46*(6), 1397-1425.

Gračanin-Yuksek, M., Lago, S., Şafak, D. F., Demir, O., & Kırkıcı, B. (2020). The interpretation of syntactically unconstrained anaphors in Turkish heritage speakers. *Second Language Research*, *36*(4), 475-501.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311-343.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem?. *Second Language Research, 28*(2), 191-215.

Guo, Y. (2020). From a simple to a complex aspectual system: Feature reassembly in L2 acquisition of Chinese imperfective markers by English speakers. *Second Language Research*.

Gürel, A. (2020). Towards a comprehensive model of heritage language development. *Bilingualism: Language and Cognition*, *23*(1), 27-28.

Håkansson, G. (1995). Syntax and morphology in language attrition. A study of five bilingual expatriate Swedes. *International Journal of Applied Linguistics 5,* 153–171.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 159–166). Pittsburgh, PA: Association for Computational Linguistics.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*(2), 101-123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643-672.

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The View from Building 20* (pp. 111-176). Cambridge: MIT Press.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of phonetics*, *40*(2), 269-279.

Hao, Y. C. (2018). Contextual effect in second language perception and production of Mandarin tones. *Speech Communication*, *97*, 32-42.

Hawkins, J. A. (1994). *A performance theory of order and constituency* (Vol. 73). Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press.

He, X., & Kaiser, E. (2016). Processing the Chinese Reflexive "ziji": Effects of Featural Constraints on Anaphor Resolution. *Frontiers in psychology*, *7*, 284.

He, W., Xu, N., & Ji, R. (2017). Effects of age and location in Chinese relative clauses processing. *Journal of psycholinguistic research*, *46*(5), 1067-1086.

Hole, D. (2008). Focus on identity–The dark side of zìjǐ. *The Linguistic Review*, *25*(3-4), 267-295.

Hoot, B. (2017). Narrow presentational focus in heritage Spanish and the syntax–discourse interface. *Linguistic Approaches to Bilingualism*, *7*(1), 63-95.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*(1), 3-27.

Hsu, C. C. (2014). The role of age in Mandarin-speaking children. *Concentric: Studies in Linguistics*, *40*(2), 29-54.

Hsu, C. C. N., Hermon, G., & Zukowski, A. (2009). Young children's production of head-final relative clauses: Elicited production data from Chinese children. *Journal of East Asian Linguistics*, *18*(4), 323-360.

Hsu, D. B. (2017). Second language learners' use of good-enough representation in grammatical processing . *Taiwan Journal of Linguistics, 15(2)*, 77-116.

Hsu, J. (1996). *Multiple comparisons: theory and methods*. CRC Press.

Hu, S., Gavarró, A., Vernice, M., & Guasti, M. T. (2016a). The acquisition of Chinese relative clauses: contrasting two theoretical approaches. *Journal of child language*, *43*(01), 1-21.

Hu, S., Gavarró, A., & Guasti, M. T. (2016b). Children's production of head-final relative clauses: The case of Mandarin. *Applied Psycholinguistics*, *37*(2), 323-346.

Hu, X., & Liu, C. (2007). Restrictive relative clauses in English and Korean learners' second language Chinese. *Second Language Research*, *23*(3), 263-287.

Huang, C.-T. J. and Liu, C.-S. L. (2001). Logophoricity, Attitudes, and Ziji at the Interface. In P. Cole, G. Hermon, & C.-T. J. Huang (Eds.), *Long-Distance Reflexives* (pp. 141–195). New York: Academic Press.

Huang, C.-T. J. and Tang, C.-C. J. (1991). The local nature of the long-distance reflexives in Chinese. In J. Koster and E. Reuland, (Eds.), *Long-Distance Anaphor* (pp. 263-282). Cambridge University Press.

Huang, C.-T. J., Li, Y. H. A., & Li, Y. (2009). The syntax of Chinese. *Cambridge, Cambridge*.

Huang, C.-C. (2003). Mandarin temporality inference in child, maternal and adult speech. *First Language*, *23*(2), 147-169.

Huang, Y.-H. (2006). The role of input in acquisition of tone sandhi rules in Mandarin Chinese. In *Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation* (pp. 287-294).

Huang, H. (2009). *Age-related Effects on the Acquisition of Second Language Phonology and Grammar*. Doctoral dissertation, UCLA.

Ionin, T., Goldshtein, M., Luchkina, T., & Styrina, S. (2020). *Word Order and Information Structure in Russian as a Heritage or Second Language*. In M. M. Brown & A. Kohut (Eds.), *Proceedings of the 44th Boston University Conference on Language Development* (pp. 210-223). Somerville, MA: Cascadilla Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446.

Jäger, L., Chen, Z., Li, Q., Lin, C. J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, *79*(80), 97-120.

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.

Jackendoff, R. (2002). *Foundations of language*. Oxford University Press.

Jacob, G., Şafak, D. F., Demir, O., & Kırkıcı, B. (2019). Preserved morphological processing in heritage speakers: A masked priming study on Turkish. *Second Language Research*, *35*(2), 173-194.

Jegerski, J. (2015). The processing of case in near-native Spanish. *Second Language Research*, *31*(3), 281-307.

Jegerski, J. (2018a). Sentence Processing in Spanish as a Heritage Language: A Self-Paced Reading Study of Relative Clause Attachment. *Language Learning*, *68*(3), 598-634.

Jegerski, J. (2018b). The processing of the object marker a by heritage Spanish speakers. *International Journal of Bilingualism*, *22*(6), 585-602.

Jegerski, J., Keating, G. D., & VanPatten, B. (2016). On-line relative clause attachment strategy in heritage speakers of Spanish. *International Journal of Bilingualism*, *20*(3), 254-268.

Jegerski, J., & Sekerina, I. A. (2020). The processing of input with differential object marking by heritage Spanish speakers. Bilingualism: Language and Cognition, *23*(2), 274-282.

Jensen, I. N., Slabakova, R., Westergaard, M., & Lundquist, B. (2020). The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English. *Second Language Research*, *36*(1), 3-29.

Jia, L., & Bayley, R. (2008). The (re) acquisition of perfective aspect marking by Chinese heritage language learners. In A. W. He and Y. Xiao (Eds.), *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 205-224). Honolulu, HI: University of Hawaii Press.

Jia, R. (2016). *Language development in Mandarin heritage language children* (Doctoral dissertation, University of Alberta).

Jia, R., & Paradis, J. (2020). The acquisition of relative clauses by Mandarin heritage language children. *Linguistic Approaches to Bilingualism*, *10*(2), 153-183.

Jiang, L. (2009). A referential/quantified asymmetry in the second language acquisition of English reflexives by Chinese-speaking learners. *Second Language Research*, *25*(4), 469-491.

Jin, L. & Hendriks, H. (2005). The development of aspect marking in L1 and L2 Chinese. *Working Papers in English and Applied Linguistics*, *9*, 69–99.

Jin, W. (2019). Acquisition of tone sandhis by English speaking learners of Chinese. *Journal of National Council of Less Commonly Taught Languages*, *25*(1), 67-107.

Kan, T. Y. R. (2018). *Development of Cantonese as a heritage language in children: Experiments in phonology and morphosyntax* (Doctoral dissertation, University of Essex).

Kearns, K. (2011). Semantics. (Second Edition). Macmillan International Higher Education.

Keating, G. D., Jegerski, J., & Vanpatten, B. (2016). Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition*, *19*(1), 36-49.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8,* 63–99.

Kidd, E., Chan, A., & Chiu, J. (2015). Cross-linguistic influence in simultaneous Cantonese–English bilingual children's comprehension of relative clauses. *Bilingualism: Language and Cognition*, *18*(3), 438-452.

Kim, C. E., & O'Grady, William. (2016). Asymmetries in children's production of relative clauses: data from English and Korean. *Journal of child language*, *43*(05), 1038-1071.

Kim, J.-H. (2007). *Binding interpretations in adult bilingualism: A study of language transfer in L2 learners and heritage speakers of Korean*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Kim, J.-H. (2013). Testing the Interface Vulnerability Hypothesis with Korean Heritage Speakers. *응용언어학*, *29*(2), 49-72.

Kim, J.-H., Montrul, S., & Yoon, J. (2009). Binding interpretations of anaphors by Korean heritage speakers. *Language Acquisition*, *16*(1), 3-35.

Kim, J.-H., Montrul, S., & Yoon, J. (2010). Dominant language influence in acquisition and attrition of binding: Interpretation of the Korean reflexive caki. *Bilingualism: Language and Cognition*, *13*(01), 73-84.

Kim, J. Y. (2016). *The perception and production of prominence in Spanish by heritage speakers and L2 learners* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Kim, J. Y. (2020). Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition*, *23*(2), 233-250.

Kong, S. (2009). The interpretation of Chinese monomorphemic and polymorphemic anaphors by adult English speakers. *Hong Kong Journal of Applied Linguistics*, *12*(1) 45-62.

Knightly, L. M., Jun, S. A., Oh, J. S., & Au, T. K. F. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America*, *114*(1), 465-474.

Kupisch, T., Lein, T., Barton, D., Schröder, D. J., Stangen, I., & Stoehr, A. (2014). Acquisition outcomes across domains in adult simultaneous bilinguals with French as weaker and stronger language. *Journal of French Language Studies*, *24*(3), 347-376.

Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, *22*(5), 564-582.

Kuo, K., & Vasishth, S. (2006). *Processing relative clauses: evidence from Chinese* (Unpublished manuscript). University of Potsdam.

Laleko, O., & Polinsky, M. (2016). Between syntax and discourse: Topic and case marking in heritage speakers and L2 learners of Japanese and Korean. *Linguistic Approaches to Bilingualism*, *6*(4), 396-439.

Leal, T., Rothman, J., & Slabakova, R. (2014). A rare structure at the syntax-discourse interface: Heritage and Spanish-dominant native speakers weigh in. *Language Acquisition*, *21*(4), 411-429.

Leal Méndez, T., Rothman, J., & Slabakova, R. (2015). Discourse-sensitive clitic-doubled dislocations in heritage Spanish. *Lingua*, *155*, 85-97.

Lee, S. Y. (2012). The acquisition of Korean reflexive *caki* by heritage speakers and L2 learners. *언어*, *37*(2), 383-400.

Lee, S. Y., Moon, J., & Long, M. H. (2009). Linguistic correlates of proficiency in Korean as a second language. *어학연구*.

Lee, T. (2016). Dominant language transfer in the comprehension of L2 learners and heritage speakers. *International Journal of Applied Linguistics*, *26*(2), 190-210.

Lee, Y. S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527-542.

Lee-Ellis, S. (2011). The elicited production of Korean relative clauses by heritage speakers. *Studies in Second Language Acquisition*, 57-89.

Lee-Ellis, S. (2012). *Looking into Bilingualism through the heritage speaker's mind* (Doctoral dissertation, University of Maryland).

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means (Version 1.3. 4).

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Li, D., & Duff, P. (2008). Issues in Chinese heritage language education and research at the postsecondary level. In A. W. He and Y. Xiao (Eds.), *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 13-36). Honolulu, HI: University of Hawaii Press.

Li, D., & Duff, P. (2018). Learning Chinese as a heritage language in postsecondary contexts. In C. Ke (Ed.), *The Routledge handbook of Chinese second language acquisition* (p. 318-348). *New York: Routledge*.

Li, J. I., & Hsieh, M. L. (2015). L2 Acquisition of *Zhe* in Mandarin Chinese: A Corpus Study. *Concentric: Studies in Linguistics*, *41*(1), 27-61.

Li, P., & Bowerman, M. (1998). The acquisition of lexical and grammatical aspect in Chinese. *First Language*, *18*(54), 311-350.

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. New York: Mouton deGruyter.

Li, Q., Guo, X. , Yao, Y. , & Müller, N. (2016). Relative clause preference in learners of Chinese as a second language. *Chinese Journal of Applied Linguistics*, *39*(2), 199-214.

Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PloS one*, *10*(12), e0143097.

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

Li, Y. A. (2016). Verbs and verb phrases. In C. R. Huang, & D. Shi (Eds.), *A reference grammar of Chinese* (pp. 81-115). Cambridge University Press.

Lim, J. H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism, 16*(3), 518.

Lin, C.-J.C., & Bever, T. G. (2011). Garden path and the comprehension of head-final relative clauses. In H. Yamashita, Y. Hirose & J. Packard (Eds.), *Processing and producing head-final structures* (pp. 277-297). Dordrecht: Springer.

Lin, J. W. (2003). Temporal reference in mandarin Chinese. *Journal of East Asian Linguistics*, *12*(3), 259-311.

Lin, Y. J., & Hsu, Y. Y. (2018). Whether and how do Mandarin sandhied tone 3 and underlying tone 2 differ?. 32nd Pacific Asia Conference on Language, Information and Computation

Lin, Y., & Garnsey, S. M. (2011). Animacy and the resolution of temporary ambiguity in relative clause comprehension in Mandarin. In H. Yamashita, Y. Hirose & J. Packard (Eds.), *Processing and producing head-final structures* (pp. 241-275). Berlin: Springer.

Liu, F.-H. (2012). L2 acquisition of the progressive marker *zai* in Mandarin Chinese. *Chinese as a Second Language Research, 1*(2),153-192.

Liu, F.-H. (2017). Second language acquisition of aspect in Mandarin Chinese. In I. Kecskes & C. Sun (Eds.), *Key Issues in Chinese as a Second Language Research* (pp. 214-234). New York: Routledge.

Liu, L. (2010). A pragmatic account of anaphora: the cases of the bare reflexive in Chinese. *Journal of Language Teaching and Research, 1*(6), 796-806.

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

Long, M. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International review of applied linguistics in language teaching*, *43*(4), 287-317.

Lu, H. Y. (2011). *The effects of verb bias, context and tasks on mandarin Mandarin reflexives* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Lu, H. Y. (submitted). *'Self in Reality'—A Corpus Study of Mandarin Reflexive 'ziji'.* Manuscript submitted for publication.

Mikhaylova, A. (2012). Aspectual Knowledge of High Proficiency L2 and Heritage Speakers of Russian. *Heritage Language Journal*, *9*(2), 50-69.

Mikhaylova, A. (2018). Morphological bottleneck: The case of Russian heritage speakers. *Journal of Language Contact*, *11*(2), 268-303.

Mikhaylova, A. (2019). What We Know about the Acquisition of Russian Aspect as a First, Second, and Heritage Language: State of the Art. *Heritage Language Journal*, *16*(2), 183-210.

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.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations?. *Memory & Cognition*, *31*(5), 663-672.

Mai, Z. (2013). *Properties of the (shi)... de focus construction in adult L2 acquisition and heritage language acquisition of Mandarin Chinese* (Doctoral dissertation, University of Cambridge).

Mai, Z. (2016). L2 Chinese: Grammatical development and processing. *Second Language Research*, *32*(1), 123-141.

Mai, Z, & Deng, X. (2019). Selective vulnerability and dominant language transfer in the acquisition of the Chinese cleft construction by heritage speakers. *Linguistic Approaches to Bilingualism*, *9*(2), 202-227.

Mansbridge, M. P., Tamaoka, K., Xiong, K., & Verdonschot, R. G. (2017). Ambiguity in the processing of Mandarin Chinese relative clauses: One factor cannot explain it all. *PloS one*, *12*(6), e0178369.

Manzini, R. M. & Wexler, K. (1987). Parameters, Binding Theory, and learnability. *Linguistic Inquiry, 18*(3), 413-444.

Margaza, P., & Gavarró, A. (2020). Null/overt subject alternations in L2 Spanish and L2 Greek. *Glossa: a journal of general linguistics, 5*(1).

Mayberry, R. I. (2010). Early language acquisition and adult language ability: What sign language reveals about the critical. In M. Marschark & P. E. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education* (pp. 281-291).

Mayberry, R. I., & Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition*, *21*(5), 886-905.

Meir, N., Parshina, O., & Sekerina, I. A. (2020). *The Interaction of Morphological Cues in Bilingual Sentence Processing: An Eye-Tracking Study*. In M. M. Brown & A. Kohut (Eds.), *Proceedings of the 44th Boston University Conference on Language Development* (pp. 376-389). Somerville, MA: Cascadilla Press.

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, *5*(01), 39-68.

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor* (Vol. 39). John Benjamins Publishing.

Montrul, S. (2010a). Dominant language transfer in adult second language learners and heritage speakers. *Second Language Research*, *26*(3), 293-327.

Montrul, S. (2010b). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, *30*(1), 3-23.

Montrul, S. (2012). Is the heritage language like a second language?. *Eurosla Yearbook*, *12*(1), 1-29.

Montrul, S. (2013). How "native" are heritage speakers?. *Heritage Language Journal*, *10*(2), 15-39.

Montrul, S. (2014). *Interlanguage, transfer and fossilization: Beyond second language acquisition*. In Han, Z., & Tarone, E. (Eds.) *Interlanguage: Forty years later* (pp. 75-104). John Benjamins Publishing Company.

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

Montrul, S. (2018). The Bottleneck Hypothesis extends to heritage language acquisition. In Choo, J., Iverson, M, Judy, T, Leal, T, & *Shimanskaya*, E. (Eds.) *Meaning and Structure in*

Second Language Acquisition: In honor of Roumyana Slabakova (pp. 149-177). Amsterdam: John Benjamins.

Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2014). Early language experience facilitates gender agreement processing in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 17(1), 118-138.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, *58*(3), 503-553.

Montrul, S., & Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism*, *13*(4), 449.

Montrul, S., & Ionin, T. (2012). Dominant language transfer in Spanish heritage speakers and second language learners in the interpretation of definite articles. *The Modern Language Journal*, *96*(1), 70-94.

Montrul, S., & Polinsky, M. (2011). Why not heritage speakers? A response to Sorace. *Linguistic Approaches to Bilingualism, 1*(1), 58–62.

Montrul, S., & Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition*, *20*(2), 109-132.

Nagy, N. (2015). A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua*, *164*, 309-327.

O'Grady, W. (2005). *How children learn language*. Cambridge University Press.

O'Grady, W., Lee, M., & Choo, M. (2001). The acquisition of relative clauses by heritage and non-heritage learners of Korean as a second language: A comparative study. *Journal of Korean Language Education, 12*(2), 283-294.

O'Grady, W., Lee, M., & Choo, M. (2003). A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Studies in Second Language Acquisition*, *25*(03), 433-448.

O'Grady, W., Lee, M., & Kwak, H.-Y. (2009). Emergentism and second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 69–88). Bingley: Emerald Group.

Oh, J. S., Au, T. K. F., Jun, S. A., & Lee, R. M. (2019). Childhood language memory in adult heritage language (re) learners. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition*. Oxford University Press.

Oh, J. S., Jun, S. A., Knightly, L. M., & Au, T. K. F. (2003). Holding on to childhood language memory. *Cognition*, *86*(3), B53-B64.

Okuma, T. (2015). L2 acquisition of discourse constraints on the use of Japanese pronouns. In *McGill Working Papers in Linguistics*.

Orfitelli, R., & Polinsky, M. (2017). When performance masquerades as comprehension: Grammaticality judgments in experiments with non-native speakers. In M. Kopotev, O. Lyashevskaya, & A. Mustajoki (Eds.), *Quantitative approaches to the Russian language* (pp. 197-214). London: Routledge.

Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 770-802.

Packard, J. L. (2000). The morphology of Chinese: A linguistic and cognitive approach.

Packard, J. L. (2008). Relative clause processing in L2 speakers of Mandarin and English. *Journal of the Chinese Language Teachers Association*, *43*(2), 107-146.

Packard, J., Ye, Z., & Zhou, X. (2011). Filler-gap processing in Mandarin relative clauses: Evidence from event-related potentials. In H. Yamashita, Y. Hirose & J. Packard (Eds.), *Processing and producing head-final structures* (pp.219-240). Dordrecht: Springer.

Pallier, C., Dehaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first?. *Cerebral cortex*, *13*(2), 155-161.

Parshina, O., Laurinavichyute, A. K., & Sekerina, I. A. (2020). Eye-movement benchmarks in Heritage Language reading. *Bilingualism: Language and Cognition*, 1-14.

Pascual y Cabo, D., & Rothman, J. (2012). The (il) logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied linguistics*, *33*(4), 450-455.

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, *41*(1), 59-86.

Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic approaches to bilingualism*, *5*(4), 409-453.

Pierce, L. J., Klein, D., Chen, J. K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences*, *111*(48), 17314-17319.

Pierce, L. J., Genesee, F., & Klein, D. (2019) Language Loss and Language Learning in Internationally Adopted Children. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition*. Oxford University Press.

Polinsky, M. (2011). Reanalysis in adult heritage language. *Studies in Second Language Acquisition*, *33*(2), 305-328.

Polinsky, M. (2018a). Bilingual children and adult heritage speakers: The range of comparison. International Journal of Bilingualism, 22(5), 547-563.

Polinsky, M. (2018b). *Heritage languages and their speakers* (Vol. 159). Cambridge University Press.

Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and linguistics compass*, *1*(5), 368-395.

Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition*, *23*(1), 4-20.

Pu, M. M. (2007). The distribution of relative clauses in Chinese discourse. *Discourse Processes*, *43*(1), 25-53.

Putnam, M. T., & Arnbjörnsdóttir, B. (2015). Minimizing (interface) domains: The loss of long-distance binding in North American Icelandic. In B. R. Page & M. T. Putnam (Eds.),

Moribund Germanic heritage languages in North America: Theoretical perspectives and empirical findings (pp. 203-223). Brill.

Qiu, Y., & Zhou, X. (2012). Processing temporal agreement in a tenseless language: An ERP study of Mandarin Chinese. *Brain research*, *1446*, 91-108.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1-23.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic inquiry*, *24*(4), 657-720.

Rodríguez-Ordóñez, I., & Sainzmaza-Lecanda, L. (2018). Bilingualism effects in Basque Subject Pronoun Expression: Evidence from L2 Basque. *Linguistic Approaches to Bilingualism*, *8*(5), 523-560.

Ross, C. (1991). Coverbs and category distinctions in Mandarin Chinese. *Journal of Chinese Linguistics*, *19*(1), 79-115.

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.

Sandel, T. L., Chao, W. Y., & Liang, C. H. (2006). Language shift and language accommodation across family generations in Taiwan. *Journal of Multilingual and Multicultural Development*, *27*(2), 126-147.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research*, *12*(1), 40-72.

Scontras, G., Fuchs, Z., & Polinsky, M. (2015). Heritage language and linguistic theory. *Frontiers in psychology*, *6*.

Scontras, G., & Putnam, M. T. (2020). Lesser-studied heritage languages: An appeal to the dyad. *Heritage Language Journal*, *17*(2), 152-155.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House.

Sekerina, I. A., & Sauermann, A. (2015). Visual attention and quantifier-spreading in heritage Russian bilinguals. *Second language research*, *31*(1), 75-104.

Sekerina, I. A., & Trueswell, J. C. (2011). Processing of contrastiveness by heritage Russian bilinguals. *Bilingualism: Language and Cognition*, *14*(3), 280-300.

Shi, M. (2011). Incomplete knowledge of aspect in heritage speakers of Mandarin Chinese in the Netherlands (Unpublished Master Thesis, Radboud University, Nijmegen, Netherlands).

Shi, Y. (2013). Analysis of L2 learners' knowledge of perfective LE in Mandarin. *Arizona Working Papers in SLA & Teaching*, *19*, 40-54.

Shih, C. L. (1986). *The prosodic domain of tone sandhi in Chinese* (Doctoral dissertation, University of California, San Diego).

Shih, C., & Sproat, R. (1992). Variations of the Mandarin rising tone. In *Proceedings of the IRCS workshop on prosody in natural speech* (Vol. 92, No. 37, pp. 193-200).

Shirai, Y., & Andersen, R. W. (1995). The acquisition of tense-aspect morphology: A prototype account. *Language*, 743-762.

Silva-Corvalán, C. (2018). Simultaneous bilingualism: Early developments, incomplete later outcomes?. *International Journal of Bilingualism*, *22*(5), 497-512.

Slabakova, R. (2008). *Meaning in the Second Language*. Berlin: de Gruyter.

Slabakova, R. (2011). Which features are at the syntax–pragmatics interface?. *Linguistic Approaches to Bilingualism*, *1*(1), 89-93.

Slabakova, R. (2014). The bottleneck of second language acquisition. *Foreign Language Teaching and Research, 46*(4), 543-559.

Slabakova, R. (2015a). The effect of construction frequency and native transfer on second language knowledge of the syntax–discourse interface. *Applied Psycholinguistics*, *36*(3), 671-699.

Slabakova, R. (2015b). Acquiring temporal meanings without tense morphology: The case of L2 Mandarin Chinese. *The Modern Language Journal*, *99*(2), 283-307.

Slabakova, R. (2019) *The Bottleneck Hypothesis updated*. In M. Rispoli & T. Ionin (Eds.), *Three Streams of Generative Language Acquisition Research* (pp. 319-315). Language Acquisition and Linguistic Disorders Series. John Benjamins.

Smiecinska, J. (2003). Stative verbs and the progressive aspect in English. *Poznan studies in contemporary linguistics*, *38*(2002), 187-195.

Smith, C. S. (1991). *The Parameter of Aspect*. Boston: Kluwer Academic Press.

Smith, C. S. (1994). Aspectual viewpoint and situation type in Mandarin Chinese. *Journal of East Asian Linguistics*, *3*(2), 107-146.

Smith, C. S. (1997). *The Parameter of Aspect*. (2nd ed.)

Smith, C. S., & Erbaugh, M. S. (2005). Temporal interpretation in mandarin chinese. *Linguistics*, *43*(4), 713-756.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and speech*, *53*(2), 273-293.

Soh, H., & Kuo, J. (2005). Perfective aspect and accomplishment situations in Mandarin Chinese. In H. J. Verkuyl, H. E. de Swart, & A. Van Hout (Eds.), *Perspectives on aspect* (pp. 199–216). Berlin: Springer.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to bilingualism*, *1*(1), 1-33.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339-368.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*(2), 195-210.

Speer, S. R., Shih, C. L., & Slowiaczek, M. L. (1989). Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Language and Speech*, *32*(4), 337-354.

Sperlich, D. (2013). *The acquisition of long-distance reflexives in Chinese as an interlanguage: An experimental study* (Doctoral dissertation, University of Auckland).

Sperlich, D. (2016). Pragmatic or Syntactic Ziji? Evidence from Language Transfer. *Journal of Cognitive Science*, *17*(4), 607-652.

Sperlich, D. (2017). The instantiation of binding through pragmatic and syntactic processes. In I. Kecskés & C. Sun (Eds.), *Key Issues in Chinese as a Second Language Research* (pp. 188-213).

Su, Y.-C. (2004). The development of long-distance binding for Mandarin ziji revisited. *Chinese Journal of Psychology*, *46*(3), 249-257.

Su, Y.-C. (2017a) Rethinking parameter setting from locality and orientation constraints for *ziji* in Mandarin. Paper presented in the 1st International Conference on Theoretical East Asian Psycholinguistics, March 10-12, Chinese University of Hong Kong.

Su, Y.-C. (2017b). Logophoric *ziji* in Mandarin child language. In M. Nakayama, Y.-C. Su, & A. Huang (Eds.), *Studies in Chinese and Japanese Language Acquisition: In honor of Stephen Crain* (pp. 65-83).

Sun, C. (2015). The uses of DE 的 as a noun phrase marker. *The Oxford Handbook of Chinese Linguistics*,. In W. S. Y. Wang, & C. Sun (Eds.), *The Oxford Handbook of Chinese Linguistics* (pp. 362-376). Oxford University Press, USA.

Sung, Y.-T., Cha, J.-H., Tu, J.-Y., Wu, M.-D., & Lin, W.-C. (2016a). Investigating the processing of relative clauses in Mandarin Chinese: evidence from eye-movement data. *Journal of psycholinguistic research*, *45*(5) 1089–1113.

Sung, Y.-T., Tu, J.-Y., Cha, J.-H., & Wu, M.-D. (2016b). Processing Preference Toward Object-Extracted Relative Clauses in Mandarin Chinese by L1 and L2 Speakers: An Eye-Tracking Study. *Frontiers in Psychology*, *7*, 4.

Tai, J. (1978). Phonological Changes in Modern Standard Chinese in the People's Republic of China Since 1949 (Washington, DC: Office of Research, United States Information Agency).

Tai, J. H.-Y. (1982). Relevant Categorical Distinctions in Chinese. In K. Tuite, R. Schneider, & R. Chametsky. (Eds), *Papers from the Eighteenth Regional Meeting of the Chicago Linguistic Society* (pp. 495-506). Chicago: Chicago Linguistic Society.

Tai, J. H.-Y. (1984). Verbs and times in Chinese: Vendler's four categories. In D. Testen, V. Mishra, & J. Drogo. (Eds), *Papers from the Parasession on Lexical Semantics* (pp. 289-296). Chicago: Chicago Linguistic Society.

Tang, C.-C. J. (1989). Chinese reflexives. *Natural Language & Linguistic Theory*, *7*(1), 93-121.

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., & Demuth, K. (2019). The acquisition of phonological alternations: The case of the Mandarin tone sandhi process. *Applied Psycholinguistics*, *40*(6), 1495-1526.

Tang, S.-W. (2016). Aspectual system. In C. R. Huang, & D. Shi (Eds.), *A reference grammar of Chinese* (pp. 116-142). Cambridge University Press.

Thomas, M. (1995). Acquisition of the Japanese reflexive zibun and movement of anaphors in Logical Form. *Second Language Research*, *11*(3), 206-234.

Tong, X., & Shirai, Y. (2016). L2 acquisition of Mandarin *zai* and *-le*. *Chinese as a Second Language Research*, *5*(1), 1-25.

Tsai, W. T. D. (2008). Tense anchoring in Chinese. *Lingua*, *118*(5), 675-686.

Tsai, W. T. D. (2019). *Causality, Commitativity, Contrastivity, and Selfhood*. In J. Hu & H. Pan (Eds.), *Interfaces* in Grammar (pp. 101-132). John Benjamins.

Tsao, F.-M. (2016) Speech perception in normally developing infants and children. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, and J. Myers (Eds.), Encyclopedia of Chinese Language and Linguistics. Retrieved February 17, 2017, from http://dx.doi.org/10.1163/2210-7363_ecll_COM_000208

Tse, H. (2016). Variation and change in Toronto heritage Cantonese: An analysis of two monophthongs across two generations. *Asia-Pacific Language Variation*, *2*(2), 124-156.

Tsay, J. (2016) Acquisition of Tone, L1. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, and J. Myers (Eds.), Encyclopedia of Chinese Language and Linguistics. Retrieved February 17, 2017, from http://dx.doi.org/10.1163/2210-7363_ecll_COM_000106

Tsimpli, I., & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In *Proceedings of the Annual Boston University Conference on Language Development 30 (pp. 653-664).*

Tsoi, E. Y. L., Yang, W., Chan, A., & Kidd, E. (2019). Mandarin–English speaking bilingual and Mandarin speaking monolingual children's comprehension of relative clauses. *Applied Psycholinguistics*, *40*(4), 933-964.

Tsukada, K., Xu, H. L., & Rattanasone, N. X. (2015). The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a Second Language Research*, *4*(2), 141-161.

van Osch, B. (2019). Vulnerability and cross-linguistic influence in heritage Spanish: Comparing different majority languages. *Heritage Language Journal*, *16*(3), 340-366.

van Osch, B., & Sleeman, P. (2018). Spanish heritage speakers in the Netherlands: Linguistic patterns in the judgment and production of mood. *International Journal of Bilingualism*, *22*(5), 513-529.

Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PloS one*, *8*(10), e77006.

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

Ventureyra, V. A., Pallier, C., & Yoo, H. Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, *17*(1), 79-91.

Wang, C.-Y. (2011). *Children's Acquisition of Tone 3 Sandhi in Mandarin*. (Unpublished doctoral dissertation). Michigan State University.

Wang, F., & Wu, F. (2020). Postnominal relative clauses in Chinese. *Linguistics*, *1*(ahead-of-print).

Wang, J.-H. (2015). A Study of the Formation of Time Adverbs *Zai* and *Zhengzai*. (時間副詞「在」與「正在」的形成探究) *Language and Linguistics*, *16*(2), 187.

Wang, X. (2006). Perception of L2 tones: L1 lexical tone experience may not help. In *Proceedings of Speech Prosody* (pp. 85-88).

Wang, Y., & Feng, L. (2014). Research on acquisition difficulty of relative clauses in Chinese as a second language. In N. Jiang (Ed.), *Advances in Chinese as a Second Language: Acquisition and Processing* (pp. 254-270). Cambridge Scholars Publishing.

Wang, Y., Sereno, J. A., & Jongman, A. (2006). L2 acquisition and processing of Mandarin tones. In P. Li, L. H. Tan, E. Bates and O. J. L Tzeng (Eds.), *The handbook of east Asian psycholinguistics* (pp. 250-256). Cambridge University Press.

Wang, Y., Jongman, A., & Sereno, J. A. (2016) Acquisition of Tone, L2. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, and J. Myers (Eds.), Encyclopedia of Chinese Language and Linguistics. Retrieved February 17, 2017, from http://dx.doi.org/10.1163/2210-7363_ecll_COM_000105

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681-712.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Wong, P. (2012). Monosyllabic Mandarin tone productions by 3-year-olds growing up in Taiwan and in the United States: Interjudge reliability and perceptual results. *Journal of Speech, Language, and Hearing Research*.

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, *133*(1), 434-443.

Wong, P., & Strange, W. (2017). Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PloS one*, *12*(8), e0182337.

Woo, I. H. (2015). Aspectual coercion and the aspectual marker–*zhe* in Mandarin Chinese. *International Journal*, *3*(2), 119-134.

Wu, F. (2009). *Factors affecting relative clause processing in Mandarin.* Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes*, *27*(10), 1489-1524.

Wu, Y., Adda-Decker, M., & Lamel, L. (2020). Mandarin lexical tones: a corpus-based study of word length, syllable position and prosodic position on duration. *Proc. Interspeech 2020*, 1908-1912.

Xiao, Z., & McEnery, A. (2004a). A corpus-based two-level model of situation aspect. *Journal of linguistics*, *40*(02), 325-363.

Xiao, R., & McEnery, T. (2004b). *Aspect in Mandarin Chinese: A corpus-based study* (Vol. 73). John Benjamins Publishing.

Xiao, Y. (2008). Home literacy environment in CHL development. In A.W. He & Y. Xiao (Eds.), *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 151-166). Honolulu, HI: University of Hawaii Press.

Xiao, Y. (2014). Heritage Language Education in the United States. In M. Bigelow & J. Ennser-Kananen (Eds.). *The Routledge handbook of educational linguistics* (pp. 370-382). Routledge.

Xu, H. (2015). The Chinese aspectual system. Ph.D. dissertation. The Hong Kong Polytechnic University.

Xu, K., Duann, J. R., Hung, D. L., & Wu, D. H. (2019). Preference for object relative clauses in Chinese sentence comprehension: Evidence from online self-paced reading time. *Frontiers in psychology*, *10*, 2210.

Xu, Y. (2014a). Evidence of the Accessibility Hierarchy in Relative Clauses in Chinese as a Second Language. *Language and Linguistics*, *15*(3), 435-464.

Xu, Y. (2014b). Processing relative clauses in Chinese as a second language. *Second Language Research*, *30*(4), 439–61.

Xu, Y. (2014c). Processing relative clauses in Chinese as a second language. *Second Language Research*, *30*(4), 439–61.

Xu, Y. (2020). Perfective-le Use and Consciousness-raising among Beginner-level Chinese Learners. *Languages*, *5*(2), 16.

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in psychology*, *9*, 817.

Yan, J., & Matthews, S. (2017). Relative clauses in English-Mandarin bilingual children: Language transfer and development in Singapore. *Chinese Language and Discourse*, *8*(1), 1-17.

Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer Berlin Heidelberg.

Yang, C. (2016). *The Acquisition of L2 Mandarin Prosody: From experimental studies to pedagogical practice* (Vol. 1). John Benjamins Publishing Company.

Yang, C., & Jin, W. (2018). Chinese as a Second Language Pronunciation Teaching Survey. *Journal of the National Council of Less Commonly Taught Languages*, *23*, 153-189.

Yang, S. (1995). The Aspectual System of Chinese. PhD thesis. University of Victoria.

Yang, S., & Huang, Y. (2013). A survey of the four major Chinese aspect markers in different modes of discourse (体标记在不同语体中的分布情况考察). *Contemporary Linguistics (当代语言学), 15*(3), 268-283.

Yao, Q., & Renaud, C. (2016). Processing Chinese relative clauses: An investigation of second-language learners from different learning contexts. *Chinese as a Second Language Research, 5*(2), 155-186.

Yao, Y. (2018). Processing relative clauses in Chinese as a second language. In X. Wen, & Jiang, X. (Eds.), *Studies on Learning and Teaching Chinese as a Second Language* (pp. 105–128). London: Routledge.

Yap, F. H., Chan, Y. H., Shirai, Y., Tan, L. H., Matthews, S., & Li, P. (2004). A perfective–imperfective asymmetry in the human mind: Evidence from Cantonese and Mandarin. Paper presented at the 28th International Congress of Psychology, Beijing.

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.

Ying, H. G. (1999). Access to UG and language transfer: A study of L2 learners' interpretation of reconstruction in Chinese. *Second Language Research, 15*(1), 41-72.

Yip, V., & Matthews, S. (2007). Relative clauses in Cantonese-English bilingual children: Typological challenges and processing motivations. *Studies in Second Language Acquisition*, *29*(02), 277-300.

Yuan, B. (1993). Directionality of difficulty in second language acquisition of English and Chinese. PhD thesis, Department of Linguistics, University of Edinburgh.

Yuan, B. (1994). Long-distance and short-distance reflexives in second language acquisition. In C. S. Rhys, D. Adger and A. von Klopp (Eds.), *Functional Categories, Argument Structure and Parametric Variation* (pp. 115-147). Edinburgh University Press.

Yuan, B. (1998). The interpretation of binding and orientation of the Mandarin reflexive ziji by English and Japanese speakers. *Second Language Research*, *14*(4), 325-340.

Yuan, J., & Chen, Y. (2014). 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics*, *42*(1), 218-237.

Zeng, L. (2010). A cross language study of L2 acquisition of Chinese and English reflexives by English and Chinese speakers (Unpublished PhD dissertation). City University of Hong Kong.

Zhang, C., & Peng, G. (2013). Productivity of Mandarin third tone sandhi: a wug test. In *Eastward flows the great river: Festschrift in honor of Prof. William SY. Wang on his 80th birthday*. City University of Hong Kong Press.

Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics*, *33*, 149-162.

Zhang, D., Lin, C. H., Zhang, Y., & Choi, Y. (2019). Pinyin or no pinyin: does access to word pronunciation matter in the assessment of Chinese learners' vocabulary knowledge?. *The Language Learning Journal*, *47*(3), 344-353.

Zhang, H. (2013). The Second Language Acquisition of Mandarin Chinese Tones by English, Japanese and Korean Speakers. Unpublished Ph.D. dissertation. The University of North Carolina at Chapel Hill.

Zhang, H. (2017). The effect of theoretical assumptions on pedagogical methods: a case study of second language Chinese tones. *International Journal of Applied Linguistics, 27*(2), 363-382.

Zhang, H. (2018a). Current Trends in Research of Chinese Sound Acquisition. In Ke, C-R. (ed.) *The Routledge Handbook of Chinese Second Language Acquisition* (p. 217-233). New York: Routledge.

Zhang, H. (2018b). *Second language acquisition of Mandarin Chinese tones: Beyond first-language transfer*. Brill.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 153-201.

Zhang, Q. (2016). Investigating the teaching of Chinese aspect markers *zai, zhe, le* and *guo* -- Grammar-Translation approach and a communicative approach: Grammar-Translation approach and a communicative approach. *Researching and Teaching Chinese as a Foreign Language*, *2*(1), 1-41.

Zhang, S. (2018). *The effect of L2 exposure on processing and perceiving aspect marking and reflexive binding in Mandarin-English bilinguals* (Doctoral dissertation, University of Reading).

Zhang, S. (2020). Does extensive L2 exposure trigger L1 attrition of perfective and durative aspect marking in Mandarin Chinese?. *Lost in Transmission: The role of attrition and input in heritage language development*, *59*, 255.

Zhang, Y., & Zhang, J. (2008). Brain responses to agreement violations of Chinese grammatical aspect. *Neuroreport*, *19*(10), 1039-1043.

Zhao, L. X. (2012). Interpretation of Chinese overt and null embedded arguments by English-speaking learners. *Second Language Research*, *28*(2), 169-190.

Zhao, Y. (2011). A tree in the wood: A review of research on L2 Chinese acquisition. *Second Language Research*, *27*(4), 559-572.

Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children* (Vol. 3). Multilingual Matters.

Zhu, H. (2016) Acquisition of Phonology, L1. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, and J. Myers (Eds.), Encyclopedia of Chinese Language and Linguistics. Retrieved February 17, 2017, from http://dx.doi.org/10.1163/2210-7363_ecll_COM_000105

Zhu, C. (2019). *Chinese Aspectual Particle le: A Comprehensive Guide*. Hong Kong University Press.

Appendix A: Blocks and items in the Tone Identification Task

	block 1	block 2	block 3	block 4	block 5
T1-T3	yi1ma3	da1yi3	ke1da3	yi1ke3	da1you3
T2-T3	da2you3	yi2ma3	da2yi3	ke2da3	yi2ke3
T3-T3	yi3ke3	da3you3	yi3ma3	da3yi3	ke3da3
T4-T3	ke4da3	yi4ke3	da4you3	yi4ma3	da4yi3
T3-T1	da3yi1	ke3da1	yi3ke1	da3you1	yi3ma1
T3-T2	yi3ma2	da3yi2	ke3da2	yi3ke2	da3you2
T3-T4	da3you4	yi3ma4	da3yi4	ke3da4	yi3ke4

Appendix B. Items in the Acceptability Judgement Task testing aspect

Lexical aspect	Token set 1-5	Token set 6-10	Token set 11-15	Token set 16-20
State (individual-level states)	漂亮 be beautiful 依賴父母 rely on parents 像媽媽 resemble mother 姓陳 take the surname of Chen 像奶奶 resemble grandmother	姓林 take the surname of Lin 愛那個男人 love that man 美麗 be gorgeous 姓李 have the last name of Li 愛一個人 love a man	喜歡中文課 like Chinese class 像爸爸 resemble father 帥 be handsome 聰明 be smart 姓王 have the last name of Wang	愛一個好朋友 love a good friend 愛那個女人 love that woman 像爺爺 resemble grandfather 英俊 be good-looking 誠實 be honest
Activities	吵架 fight (orally) 打球 play balls 喝酒 drink wine 寫功課/作業 do homework 看海 see the sea	準備早餐 prepare breakfast 看書 read books 跑步 run 游泳 swim 彈鋼琴 play piano	跳舞 dance 畫圖 draw 看報紙 read newspapers 唱歌 sing songs 睡午覺 take a nap	洗澡 take a bath 踢足球 play soccer 開飛機 fly planes 吃飯 eat (rice) 聽 K-POP listen to K-POP
Accomplish-ments (all with quantified numerals, so that they are unambiguous-ly bounded events)	畫三幅畫 draw three pictures 寫三封信 write three letters 喝一杯酒 drink a glass of wine 做三個箱子 make three boxes 設計三個新產品 design three new products	買三條裙子 buy three dresses 寫三份報告 write three reports 燒三份文件 burn three documents 修三個馬桶 repair three toilets 蓋三座橋 build three bridges	挖三個洞 dig three holes 烤三個蛋糕 bake three cakes 蓋三間房子 build three houses 吃三顆蘋果 eat three apples 出版一本小說 publish a novel	賣一輛車子 sell a car 看三本雜誌 read three magazines 唱三首英文歌 sing three English songs 縫一件衣服 sew a dress/shirt 看三本書 read three books
Achievements (all with objects, so that *-le* is not sentence-final *LE*)	注意到交通問題 notice traffic problems 贏游泳比賽 win swimming race 到臺灣 arrive Taiwan 弄丟錢 lose money 弄丟錢包 lose wallet	到美國 arrive the United States 找到幸福 find happiness 看見警察 see police 贏賽跑 win race 走進阿姨的房間 walk in aunt's room	弄丟手機 lose cellphone 到英國 arrive England 到中國 arrive China 發現一個秘密 find a secret 輸足球比賽 lose soccer game	取消比賽 cancel game 發現垃圾 find garbage 弄丟電腦 lose computer 遇見明星 meet star 輸籃球比賽 lose basketball game