

CHINESE SENTENCE PROCESSING BY FIRST AND SECOND
LANGUAGE SPEAKERS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF
THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

SECOND LANGUAGE ACQUISITION

November 2015

By

Zhijun Wen

Dissertation Committee:

Bonnie D. Schwartz, Chairperson
Robert Bley-Vroman
Song Jiang
William O’Grady
Amy Schafer
Kamil Ud Deen

Keywords: acquisition, language processing, Mandarin, L2 learners, second language

© Copyright Zhijun Wen 2015
All Rights Reserved

ACKNOWLEDGMENTS

My Ph.D. journey has been an incredible venture. It is difficult to imagine how I could ever have completed my doctoral study without the help and support of so many extraordinary people. Because of this, I am full of thanks. God has so greatly blessed me through these wonderful people.

First of all, I owe a huge debt of gratitude to my Ph.D. supervisor and dissertation committee chair, Bonnie D. Schwartz, who has had an enormous impact on my academic development. This dissertation project actually evolved from her classes. My dissertation was greatly improved through countless hours of discussions and talks with her. It would have looked very different without her step-by-step guidance. Throughout these years, I have learned immensely from her. I am always amazed at her logical thinking, intellectual rigor, and sharp insights every time she gives me feedback on my work. Besides my dissertation, she has also helped me improve the quality of my conference abstracts, conference presentations, grant proposals, job applications, published papers, etc. She always treats my work with the utmost care, in terms of the great effort and time she devoted to helping me develop and improve it. I feel so fortunate to have her as my doctoral advisor and also as my mentor and friend.

I am also deeply indebted to those who have played important roles in my academic pursuits. My other dissertation committee members, Robert Bley-Vroman, Kamil Deen, Song Jiang, William O’Grady, and Amy Schafer, as well as my former committee member Ying-Che Li (who I affectionately call “Li Laoshi”), have always been so supportive, and their input has without a doubt enhanced the quality of my dissertation. Their passion for linguistic research has greatly influenced the growth of my academic life. They are all my role models as great linguists and great educators. Gregory C. Richter (my M.A. supervisor, Professor of Linguistics at Truman State University) and Chuming Wang (Professor of Linguistics and Applied Linguistics at Guangdong University of Foreign Studies) led me onto the road of language research in my formative years as an M.A. student of linguistics. I took my first linguistics and applied linguistics courses from them. They have been my mentors and dear friends ever since. They helped me to become interested in language research, and they are people

whom I can always count on whenever I need help. Sue Shufen Fujitani (my dear colleague and friend at Kapi‘olani Community College) has given me so much support in my teaching and job searching. Jiuqing Ge encouraged me to study English while I was a senior high school student, and was the first person who saw my potential in scholarship.

I thank the following colleagues and friends in China for helping me recruit participants and conduct the experiments: Haiyan Li and Wenlong Wang (Peking University); Nannan Lu (Tsinghua University); Lijin Liu, Qigui Peng, and Lixin Xia (Guangdong University of Foreign Studies); and Ruifeng Wang (Beijing Language and Culture University). I am grateful to the following people who also helped with my research: in Beijing, Sonya Hallett, Xiaoying Lu, Yajing Wen; in Boston, Yu Wu; and in Honolulu, Jung Ying Lu-Chen, Daniel Tom, Daniel Tschudi, Stephen Tschudi, and Qiaona Yu.

I am also very thankful to the many people who participated in my experiments.

My sincere thanks go to those who have shared insights with me via engaging discussions: Harald Clahsen, Laurent Dekydtspotter, Claudia Felser, Theres Grüter, Holger Hopp, Alan Juffs, Edith Kaan, Gregory D. Keating, Chien-Jer Charles Lin, A. Kate Miller, Claire Renaud, Leah Roberts, Roumyana Slabakova, Annie Tremblay, and Lydia White.

I appreciate the feedback I received from the audiences of various conferences, especially the following: the Boston University Conference on Language Development (2009, 2011, 2015); the Chuo-UHM Student Conference on Linguistics, Psycholinguistics and Second Language Acquisition (2015); Generative Approaches to Language Acquisition—North America (2012); Generative Approaches to Second Language Acquisition (2015), and Second Language Research Forum (2015). I am also thankful for the feedback from the people in the Language Acquisition Reading Group at the University of Hawai‘i.

I want to thank my friends and colleagues at Guangdong University of Foreign Studies: Chengru Dong, Yaping Gan, Biyan Li, Dongyun Li, Zirong Li, Darong Liu, Jianfu Liu, Baoliang Peng, Meining Wang, Xiaoxin Wu, Jianglin Zhang, and Li Zhang.

My sincere thanks go to those who, over the years, have given me encouragement, friendship, help, and support in various ways. They include the following faculty

members and staff at the University of Hawai‘i and the East West Center: Ray Allen, Richard R. Day, Mendl Djunaidy, Kenton Harsch, Thom Hudson, Gabriele Kasper, John David Kawahara, Teri Kim, Emily Lee, Karen Matsumoto, Cynthia Ning, John Norris, Lourdes Ortega, Richard Schmidt, Jim Yoshioka, and Shuqiang Zhang. They also include my dear colleagues and friends I met in Hawai‘i: Yunchuan Chen, Wei Chu, Erik Hill, Yao Z. Hill, Hyunwoo Kim, Kitaek Kim, Elaine Lau, Houxiang Li, Nian Liu, Mari Miyao, Jun Nomura, Akira Omaki, Kyaee-Sung Park, Carl Polley, Matthew Prior, Barbara Schulz, Yijun Shiung, Aya Takeda, Juan Wang, Shu-Ling Wu, and Jing Zhang.

Thanks must also go to Elizabeth Bates, Jeffrey L. Elman, Ronald Langacker, and Marta Kutas at the University of California at San Diego, through whom I developed an interest in psycholinguistic research during the time I was a visiting scholar there.

My special appreciation goes to my parents, my brother and three sisters, my wonderful wife, Jing, and my dear sons, Michael and James. My parents and siblings have always taken great pride in what I have achieved, and their love always encourages me to strive for the best. Jing not only provided spiritual and emotional support, but also helped with collecting and analyzing the experimental data. Michael and James can always make me laugh—even during very stressful times—and that was more than occasionally sorely needed!

The dissertation research received generous funding from the National Science Foundation under a Doctoral Dissertation Research Improvement Grant (BCS 1023660) and from the Research Corporation of the University of Hawai‘i under a Dissertation Completion Award, which I gratefully acknowledge. I am also thankful for the support provided from the 2010–2011 University of Hawai‘i–Peking University Exchange Program, a GALANA Travel Fellowship (2012), and Paula Menyuk Travel Awards (2009, 2011, 2015). I acknowledge the various forms of support from the University of Hawai‘i via an Arts & Sciences Student Research Award, the Chung Fong and Grace Ning Chinese Studies Fund, the Elizabeth Carr-Holmes Scholarship Fund (Department of Second Language Studies), a Graduate Student Organization Grant, and a Ruth Crymes Memorial Grant (Department of Second Language Studies).

There are so many people who deserve to be credited for their help and support, but I need to stop here. I want to thank Almighty God, my Creator and the Creator of the universe, for His bringing me here to Hawai‘i, a paradise on earth and a place of *aloha*. I have enjoyed living here. I thank God for giving me this life and for being my constant source of strength, inspiration, and perseverance. May Your name be honored in my life and research.

ABSTRACT

This dissertation research investigates real-time second language (L2) sentence processing, with a focus on L2 Mandarin Chinese. It seeks to reveal how adult L2 learners (“L2ers”) exploit different sources of information (morphosyntactic, lexical-semantic, and discourse-context) and hence illuminates the issue of whether L2ers are capable of fully specified processing.

The research employs two different processing tasks: comprehension-focused self-paced reading (CFSPR) and acceptability-judgment self-paced reading (AJSPR); there are four main studies: one in English comparing native English speakers with L1-Chinese L2ers of English, and three in Chinese comparing native Chinese speakers with learners of Chinese whose native language (L1) is English (or Japanese). The first study assesses whether AJSPR is more sensitive than CFSPR in detecting deep-level processing by examining L1 and L2 processing of English subject-verb number agreement in an AJSPR task in comparison to Wen’s (2007) study employing a CFSPR task. The second study compares L1 and L2 processing of (grammatical vs. ungrammatical) negation–aspect interactions in Chinese to test whether L2ers whose L1 lacks such morphosyntax can ultimately acquire it and use it in online sentence processing. The third study investigates use of lexical-semantic information in Chinese to test whether L2ers, in comparison to Chinese natives, over-depend on lexical-semantic information in sentence processing; this is achieved by comparing L1 and L2 participants’ sensitivity to temporary violations of the selectional restrictions of Chinese transitive verbs. The fourth study examines whether L2ers rely more on discourse-context information in processing Chinese sentences that contain null objects than Chinese natives do.

The results of these studies reveal that (a) L2ers are able to access and use the different sources of information (morphosyntactic, lexical-semantic, discourse-context) in online sentence processing as L2 proficiency rises; (b) they do not always over-rely on lexical-semantic or contextual information in L2 sentence processing; (c) they are capable of fully specified processing. The research suggests that L2 processing difficulties are not necessarily indicative of deficient grammatical representations; rather,

L2 sentence processing is subject to the influences of task demands, L2 proficiency, and L1 transfer. In addition, the research establishes AJSPR as an appropriate tool for gauging deep-level L2 processing.

TABLE OF CONTENTS

Acknowledgments.....	iii
Abstract.....	vii
List of Tables.....	xv
List of Figures.....	xvi
Chapter 1: Introduction.....	1
Chapter 2: Methodology for Probing Fully Specified L2 Processing.....	14
2.1 Introduction.....	14
2.2 L2 Processing of Subject-Verb Number Agreement in English.....	16
2.3 Detailed Description of Wen (2007).....	17
2.4 Observations and Thoughts.....	20
2.5 The Present Study: Acceptability-Judgment Self-Paced Reading.....	22
2.5.1 Method.....	23
2.5.1.1 Participants.....	23
2.5.1.2 Materials.....	24
2.5.1.3 Task and Procedure.....	24
2.5.2 Results.....	24
2.5.3 Discussion.....	29
2.6 Conclusion.....	33
Chapter 3: General Methods.....	34
3.1 Participants.....	34
3.2 Proficiency Measure—A Quasi-C-Test in Chinese.....	35
3.3 Material Development.....	37
3.4 Experimental Tasks and Procedures.....	38
3.5 Data Analysis.....	40
Chapter 4: Processing Morphosyntax In Chinese—Negation—Aspect Interactions.....	42
4.1 Negation in Chinese.....	42
4.2 Aspect Markers <i>-le</i> and <i>-guo</i>	45
4.3 Negation—Aspect Interactions.....	47

4.4 L2 Theoretical Background	51
4.5 Experiment 4.1: Comprehension-Focused Self-Paced Reading	55
4.5.1 Method	55
4.5.1.1 Participants.....	55
4.5.1.2 Materials	56
4.5.1.3 Procedure	58
4.5.2 Results.....	59
4.5.2.1 Comprehension Accuracy	59
4.5.2.2 Reading Times	61
4.5.3 Discussion.....	68
4.6 Experiment 4.2: Acceptability-Judgment Self-Paced Reading.....	69
4.6.1 Method	69
4.6.1.1 Participants.....	69
4.6.1.2 Materials	70
4.6.1.3 Procedure	71
4.6.2 Results.....	71
4.6.2.1 Acceptability-Judgment Accuracy	71
4.6.2.2 Reading Times	73
4.6.3 Discussion.....	80
4.7 General Discussion	81
4.8 Conclusion	86
Chapter 5: Processing Semantic Information in Chinese	88
5.1 Native and Nonnative Processing of Lexical-Semantic Information	89
5.2 Experiment 5.1: Comprehension-Focused Self-Paced Reading	97
5.2.1 Method	97
5.2.1.1 Participants.....	97
5.2.1.2 Materials	99
5.2.1.3 Procedure	105
5.2.2 Results.....	105
5.2.2.1 Comprehension Accuracy	106
5.2.2.2 Reading Times	113

5.2.3 Discussion.....	117
5.3 Experiment 5.2: Acceptability-Judgment Self-Paced Reading.....	119
5.3.1 Method.....	120
5.3.1.1 Participants.....	120
5.3.1.2 Materials.....	121
5.3.1.3 Procedure.....	121
5.3.2 Results.....	122
5.3.2.1 Acceptability Judgments.....	122
5.3.2.2 Reading Times.....	124
5.4 General Discussion.....	129
5.5 Conclusion.....	131
Chapter 6: Processing Contextual Information in Chinese.....	132
6.1 L1 Processing of Contextual Information.....	132
6.2 Processing of Contextual Information in Chinese.....	135
6.3 L2 Processing of Discourse-context information.....	138
6.4 Experiment 6.1: Comprehension-Focused Self-Paced Reading.....	147
6.4.1 Method.....	148
6.4.1.1 Participants.....	148
6.4.1.2 Materials.....	149
6.4.1.3 Procedure.....	156
6.4.2 Results.....	156
6.4.2.1 Offline Task Results.....	156
6.4.2.2 Online Task Results.....	158
6.5 Experiment 6.2: Acceptability-Judgment Self-Paced Reading.....	172
6.5.1 Method.....	173
6.5.1.1 Participants.....	173
6.5.1.2 Materials.....	174
6.5.1.3 Procedure.....	174
6.5.2 Results.....	175
6.5.2.1 Offline Task Results.....	175
6.5.2.2 Online Task Results.....	176

6.5.3 Discussion.....	188
6.6 General Discussion	193
6.7 Conclusion	195
Chapter 7: Conclusion.....	198
7.1 Summary of the Major Findings.....	199
7.2 Implications.....	202
7.4 Concluding Remarks.....	204
Appendix A: Instructions for the Experiments in Wen (2007) and Chapter 2	206
Appendix B: Experimental Items, Fillers, and Practice Sentences Used in the Experiments in Wen (2007) and Chapter 2	207
Appendix C: Mean Acceptability-Judgment Accuracy Rates for the Experimental Sentences and Fillers in the Experiment of Chapter 2	211
Appendix D: Trimmed Raw RT Means (in Milliseconds) for the Experimental Sentences Used in the Experiment of Chapter 2	212
Appendix E: Language Background Questionnaire	213
Appendix F: The Chinese Proficiency Test	214
Appendix G: Scoring Method and Answer Key for the Chinese Proficiency Test.....	215
Appendix H: The Special Type of Filler Used in the Online Tasks.....	216
Appendix I: Instructions Used in the Online Tasks	217
Appendix J: Practice Items for the Self-Paced Reading Tasks	219
Appendix K: Experimental Items for Experiments 4.1 and 4.2.....	221
Appendix L: Grammatical <i>bu</i> Fillers for Experiments 4.1 and 4.2	223
Appendix M: Mean Comprehension Accuracy Rates for the Experimental Sentences in Experiment 4.1	224
Appendix N: Trimmed Raw RT Means (in Milliseconds) for the Experimental Sentences in Experiment 4.1	225
Appendix O: Mean Acceptability-Judgment Accuracy Rates for the Experimental Sentences in Experiment 4.2	226
Appendix P: Trimmed Raw RT Means (in Milliseconds) for the Experimental Sentences in Experiment 4.2	227

Appendix Q: Analyses Of Variance for the Effects of Condition Before and After the Exclusion of the Incorrect Trials in Experiment 4.2	228
Appendix R: Analyses of Variance for the Effects of Condition Before and After the Removal of the Achievement-Predicate Item	229
Appendix S: Experimental Sentences for Experiments 5.1 and 5.2 and the Corresponding Comprehension Questions in Experiment 5.1	230
Appendix T: Sentences in Norming Study Part 1 for Experiments 5.1 and 5.2	232
Appendix U: Sentences in Norming Study Part 2 for Experiments 5.1 and 5.2.....	233
Appendix V: Mean Comprehension Accuracy Rates for the Experimental Sentences in Experiment 5.1	234
Appendix W: Trimmed Raw RT Means (in Milliseconds) for the Experimental Sentences in Experiment 5.1	235
Appendix X: Paired-Sample <i>T</i> Tests Based on all Trials vs. on Correct Trials Only In Experiment 5.1	236
Appendix Y: Mean Acceptance Rates for the Experimental Sentences In Experiment 5.2	237
Appendix Z: Trimmed Raw RT Means (in Milliseconds) for the Experimental Sentences in Experiment 5.2	238
Appendix AA: Experimental Items for the Online Reading Tasks in Experiments 6.1 and 6.2	239
Appendix AB: Mean Comprehension Accuracy Rates for the Experimental Sentences in the Offline Task of Experiment 6.1	241
Appendix AC: Mean Comprehension Accuracy Rates for the Experimental Sentences in the Online Task of Experiment 6.1	242
Appendix AD: Trimmed Raw RT Means (in Milliseconds) for the Experimental Sentences in Experiment 6.1	243
Appendix AE: Paired-Sample <i>T</i> Tests Based on All Trials vs. on Correct Trials Only in Experiment 6.1	245
Appendix AF: Mean Comprehension Accuracy Rates for the Experimental Sentences in the Offline Task of Experiment 6.2	246

Appendix AG: Mean Acceptance Rates for the Experimental Sentences in the Online Task of Experiment 6.2.....	247
Appendix AH: Trimmed Raw RT Means (in Milliseconds) for the Experimental Sentences in Experiment 6.2	248
References.....	249

LIST OF TABLES

Table	Page
2.1. Sample Set of Experimental Stimuli in Wen (2007)	18
2.2. Chinese Participants' English Background in Wen (2007).....	18
2.3. Chinese Participants' English Background in the AJSPR Task.....	23
4.1. Participants' Chinese Background Information in Experiment 4.1	55
4.2. Sample Set of Experimental Stimuli in Experiment 4.1	56
4.3. Mean Comprehension Accuracy Rates for the Fillers in Experiment 4.1	59
4.4. Participants' Chinese Background Information in Experiment 4.2.....	70
4.5. Mean Acceptability-Judgment Accuracy Rates for the Fillers in Experiment 4.2	71
4.6. Situation Types of the Predicates Used in the Experimental Sentences.....	85
5.1. Chinese Background Information of the Participants in Experiment 5.1.....	98
5.2. Sample Pair of Experimental Stimuli in Experiment 5.1	101
5.3. Participants' ($n = 10$) Acceptance Rates of the Norming Sentences ($k = 32$) Used for Assessing Whether the Verbs Can Take Human Objects.....	103
5.4. Participants' ($n = 20$) Verb-Selection Accuracy Rates for the Norming Sentences ($k = 16$) Used for Assessing Whether the Verbs in the Locally Plausible Condition Can Take Inanimate Objects	104
5.5. Mean Comprehension Accuracy Rates for the Fillers in Experiment 5.1	106
5.6. Chinese Background Information of the Participants in Experiment 5.2.....	120
5.7. Mean Acceptability-Judgment Accuracy Rates for the Fillers in Experiment 5.2	122
6.1. Chinese Background Information of the Participants in Experiment 6.1.....	149
6.2. Sample Set of Experimental Stimuli in Experiment 6.1	153
6.3. Participants' ($N = 20$) Acceptance Rates of the Norming Sentences ($k = 16$) in Experiment 6.1	155
6.4. Mean Comprehension Accuracy Rates of the Fillers in Experiment 6.1	159
6.5. Summary of the Main Online RT Results of Experiment 6.1	170
6.6. Chinese Background Information of the Participants in Experiment 6.2.....	173

6.7.	Mean Acceptability-Judgment Accuracy Rates of the Fillers in Experiment 6.2	177
6.8.	Summary of the Main Online RT Results of Experiment 6.2	188

LIST OF FIGURES

Figure	Page
2.1. Natives' Mean Reading Time (msec) Profile in Wen (2007).	19
2.2. L2ers' Mean Reading Time (msec) Profile in Wen (2007).	20
2.3. Comparison of L2 Proficiency in Wen (2007) and the Current Study.	23
2.4. Natives' and L2ers' Mean Accuracy Rates for the Experimental Sentences in the AJSPR Task.	25
2.5. Natives' and L2ers' Mean Acceptance Rates for the Experimental Sentences in the AJSPR Task.	26
2.6. Natives' Mean Reading Time (msec) Profile in the AJSPR Task.....	27
2.7. L2ers' Mean Reading Time (msec) Profile in the AJSPR Task.....	28
4.1. Feature-Based Framework of Negation–Aspect Interactions in Chinese.....	50
4.2. Mean Comprehension Accuracy Rates for the Experimental Sentences of Experiment 4.1.	60
4.3. Natives' Mean RT (msec) Profile for the <i>-guo</i> Sentences of Experiment 4.1.	63
4.4. Natives' Mean RT (msec) Profile for the <i>-le</i> Sentences of Experiment 4.1.	64
4.5. Advanced L2ers' Mean RT (msec) Profile for the <i>-guo</i> Sentences of Experiment 4.1.	65
4.6. Advanced L2ers' Mean RT (msec) Profile for the <i>-le</i> Sentences of Experiment 4.1.	65
4.7. Intermediate L2ers' Mean RT (msec) Profile for the <i>-guo</i> Sentences of Experiment 4.1.	67
4.8. Intermediate L2ers' Mean RT (msec) Profile for the <i>-le</i> Sentences of Experiment 4.1.	67
5.1. Mean Comprehension Accuracy Rates for the Experimental Sentences of Experiment 5.1.	106
5.2. Mean Accuracy Rates for the Shallow-Targeted Questions vs. the Non-Shallow-Targeted Questions.	109
5.3. Mean Accuracy Rates for the Common Verification Questions.....	111
5.4. Natives' Mean RT (msec) Profile of Experiment 5.1.	115

5.5.	Advanced L2ers' Mean RT (msec) Profile of Experiment 5.1	115
5.6.	Intermediate L2ers' Mean RT (msec) Profile of Experiment 5.1	116
5.7.	Mean Acceptance Rates for the Experimental Sentences of Experiment 5.2.....	123
5.8.	Natives' Mean RT (msec) Profile of Experiment 5.2.	125
5.9.	Advanced L2ers' Mean RT (msec) Profile of Experiment 5.2.	126
5.10.	Intermediate L2ers' Mean RT (msec) Profile of Experiment 5.2.	126
6.1.	Mean Comprehension Accuracy Rates for the Experimental Sentences in the Offline Task of Experiment 6.1.	157
6.2.	Mean Comprehension Accuracy Rates for the Experimental Sentences in the Online Task of Experiment 6.1.	159
6.3.	Natives' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.	163
6.4.	Natives' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.	163
6.5.	English Advanced L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.	164
6.6.	English Advanced L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.	165
6.7.	English Intermediate L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.	166
6.8.	English Intermediate L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.	166
6.9.	Japanese Advanced L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.	167

CHAPTER 1

INTRODUCTION

Second language (L2) sentence processing deals with how (adult) L2 learners (henceforth “L2ers”) analyze and interpret (i.e., *parse*) sentences in their nonnative language in real time, seeking to discover the underlying mechanisms responsible for L2 use (and, ultimately, L2 acquisition). It employs online methodologies such as self-paced reading or listening, cross-modal priming, and eye tracking (for reviews, see Frenck-Mestre, 2005; Marinis, 2003; Roberts, 2012b; Roberts & Siyanova-Chanturia, 2013) as well as event-related brain potential (ERP) and functional magnetic resonance imaging (fMRI) technologies (for reviews, see Kotz, 2009; Mueller, 2005; van Hell & Tokowicz, 2010).

Research on real-time L2 sentence processing started about two decades ago with the publication of Juffs and Harrington’s (1995) pioneering study on how adult L2ers process *wh*-movement in English. Since then, the field began to take off. The development of the field, to some extent, has responded to Gregg’s (1996, 2003) call for a theory about L2 development (i.e., what he calls an *L2 transition theory*) to account for how the L2 learning mechanism works on the target language (TL) input to bring about changes in the Interlanguage grammar. The field has seen rapid growth in recent years. New hypotheses, models, and proposals have sprung up to tackle L2 processing from different perspectives, including, for example, Acquisition by Processing Theory (Sharwood Smith & Truscott, 2014; Truscott & Sharwood Smith, 2004), the Amelioration Hypothesis (O’Grady, 2012, 2013a, 2013b), the Full Transfer/Full Access/Full Parse theory (Dekydtspotter, Schwartz, & Sprouse, 2006; Schwartz & Sprouse, 1996), the Fundamental Identity Hypothesis (Hopp, 2007), the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b, 2006c), and the UG (Universal Grammar)-Based Universal Parser proposal (Dekydtspotter, 2001; Dekydtspotter & Renaud, 2014). A large number of publications have also appeared, addressing the nature of L2 sentence processing and how it unfolds in real time (for conceptual discussions and reviews of L2 sentence processing, see, e.g., Clahsen & Felser, 2006a, 2006b, 2006c; Dekydtspotter et al., 2006; Felser & Clahsen, 2009; Harrington, 2001; Juffs, 2001;

Roberts, 2010, 2012a, 2013). L2 sentence processing has now become a burgeoning area of L2 research (Keating & Jegerski, 2015). It is perhaps not an exaggeration to state that L2 sentence processing is gradually transforming the entire field of second language studies.

Questions often raised by L2 sentence processing researchers include, for example, whether L2ers transfer their first language (L1) processing strategies to L2 processing, whether L2 sentence parsing routines change over the course of L2 development, and whether (very) high-level proficiency in the TL can lead to the development of native-like parsing mechanisms in the Interlanguage system (e.g., Marinis, 2003; Papadopoulou, 2005; Roberts, 2010, 2012a). Most of the current studies on L2 sentence processing revolve around L1 and L2 similarities vs. differences in the use of structural information in online sentence comprehension (for a recent review, see Roberts, 2013).

Some major issues that L2 sentence processing studies have explored so far include (a) whether L2ers are (able to become) similar to mature L1 speakers (henceforth “natives”) in online attachment preferences both for relative clauses (e.g., Dekydtspotter, Donaldson, Edmonds, Fultz, & Petrush, 2008; Dussias, 2003; Felser, Roberts, Marinis, & Gross, 2003; Papadopoulou & Clahsen, 2003; Pan, Schimke, & Felser, 2015; Witzel, Witzel, & Nicol, 2012) and for prepositional phrases (e.g., Frenck-Mestre & Pynte, 1997; Pan & Felser, 2011); (b) whether L2ers (can) exhibit native-like processing routines in resolving temporary structural ambiguities of garden-path (GP) sentences (e.g., Juffs, 1998a, 2004; Rah & Adone, 2010; Roberts & Felser, 2011); (c) whether L2ers are (eventually) able to process filler-gap dependencies involved in *wh*-movement in the same way as natives do (e.g., Dallas, DeDe, & Nicol, 2013; Dekydtspotter & Miller, 2013; Dussias & Piñar, 2010; Felser, Cunnings, Batterham, & Clahsen, 2012; Felser & Roberts, 2007; Marinis, Roberts, Felser, & Clahsen, 2005; Miller, 2011, 2014, 2015; Omaki & Schulz, 2011; Pliatsikas & Marinis, 2013; Williams, 2006; Williams, Mobius, & Kim, 2001); and (d) whether L2ers (can) demonstrate native-like sensitivity to agreement violations (e.g., Alemán Bañón, Fiorentino, & Gabriele, 2014; Chen, Shu, Liu, Zhao, & Li, 2007; Coughlin & Tremblay, 2013; Foucart & Frenck-Mestre, 2011, 2012; Gillon Dowens, Guo, Guo, Barber, & Carreiras, 2011; Gillon Dowens, Vergara, Barber,

& Carreiras, 2010; Jiang, 2004, 2007; Keating, 2009; Ojima, Nakata, & Kakigi, 2005; Sagarra & Herschensohn, 2010, 2011; Song, 2015; VanPatten, Keating, & Leiser, 2012; Wen, 2007; Wen, Miyao, Takeda, Chu, & Schwartz, 2010; Wen & Schwartz, 2012). The main goal of L2 sentence processing research is to find out what sources of information L2ers can and cannot make use of in online sentence comprehension in the TL. This issue is of crucial importance in that it can shed light on whether (and if so, when in the course of L2 development) L2ers can ultimately develop native-like processing routines.

In real-time (L1) sentence processing, the language processor draws on multiple sources of information (e.g., lexical, morphosyntactic, semantic, prosodic, and contextual) and applies them in rapid fashion (e.g., Altmann & Steedman, 1988; Gibson & Pearlmutter, 1998; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; MacDonald, Pearlmutter, & Seidenberg, 1994; Marslen-Wilson, Brown, & Tyler, 1988; Tanenhaus & Trueswell, 1995; Trueswell, Tanenhaus, & Garnsey, 1994). The time pressure of language processing requires highly efficient use of the information. Ideally, L2ers should be able to employ and integrate all the sources of information as efficiently and effectively as natives do; in reality, it is often difficult for L2ers to achieve this. Differences between L1 and L2 online processing behavior have led some L2 researchers (e.g., Clahsen & Felser, 2006a, 2006b, 2006c; Felser & Clahsen, 2009) to suggest that L2 processing is fundamentally different from L1 processing in that L2 processing is based primarily on meaning-based information (e.g., lexical-semantic, discourse-context, and real world information), but not on morphosyntactic information (especially inflectional morphology or nonlocal structural, i.e., syntactic, relations). They argue that this is due, ultimately, to the L2 parser (i.e., L2ers' morphosyntactic processing mechanisms) being fed by a severely deficient L2 grammar, which cannot provide the information the parser needs to construct morphosyntactic representations for the TL input during online sentence processing. As a result, L2ers are left with no choice but to

primarily rely on the “shallow processing” route to interpretation, which involves computing representations of the input that lack grammatical detail and are based primarily on semantics-driven processing heuristics. (Felser & Clahsen, 2009, p. 312)

Inspired by Townsend and Bever's (2001) integrated processing model, the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, 2006b, 2006c; Felser & Clahsen, 2009) presupposes that the human sentence comprehension system provides two processing routes for the interpretation of the input: *shallow processing* and *full parsing*.¹ The two processing routes differ in terms of both the information they access and the representations they create. The shallow processing route relies on nonsyntactic information, such as lexical semantics, pragmatics, and surface cues, and, consequently, it produces less detailed, good-enough representations for the input, whereas the full parsing route is fed by the grammar and outputs fully specified representations. The SSH claims that adult L2 sentence processing is, in essence, (necessarily) restricted to the shallow processing route. The full parsing route, which constitutes the backbone of L1 sentence processing, is largely unavailable to adult L2ers. As Clahsen and Felser (2006a, p. 118) put it, "with the full parsing route being of limited use in L2 processing, learners' interpretations will typically be derived via the shallow processing route only." If what the SSH contends is correct, it would have serious repercussions for adult L2 acquisition: On the assumption that the way the TL input is processed determines the nature of the knowledge thereby acquired, L2 shallow processing will (inevitably) hinder the Interlanguage grammar from converging on the TL grammar, making attainment of native-like competence impossible for adult L2ers (for discussion, see Dekydtspotter et al., 2006).

The major sources of evidence that Clahsen and Felser originally cited to support the SSH come from behavioral online studies on L2 processing of sentences involving relative-clause (RC)-attachment ambiguity (e.g., Felser et al., 2003; Papadopoulou & Clahsen, 2003) and long-distance filler-gap dependencies (e.g., Felser & Roberts, 2007; Marinis et al., 2005).

¹ The two processing routes to which the SSH makes reference are derived from Townsend and Bever's (2001) (L1) model called *Late Assignment of Syntax Theory* (LAST). In this model, the processing system computes two independent levels of meaning representation for the input: an initial "pseudosyntactic" representation based on lexical and statistical information and a subsequent full representation based on the grammar. The pseudosyntactic representation makes use of superficial cues such as associative connections and templates and provides an initial hypothesis about the meaning, whereas the full syntactic parse serves to verify the initial analysis. The basic tenet of this model, namely, that natives "understand every sentence twice" (Townsend & Bever, 2001, p. 6) inspires Clahsen and Felser to frame human sentence processing in terms of two processing routes (i.e., shallow processing and full parsing).

In resolving RC-attachment ambiguities in sentences such as (1), English-speaking natives prefer low attachment (i.e., attaching the RC *who was standing on the balcony* to the second NP *the actress*), while natives of Greek favor high attachment (i.e., attaching the RC to the first NP *the servant*).

(1) Someone shot the servant of the actress who was standing on the balcony.

Felser et al. (2003) and Papadopoulou and Clahsen (2003) found that L2ers fail to display native-like attachment preferences for such structurally ambiguous sentences. The findings were taken as evidence that L2ers cannot use structure-based parsing principles in online sentence processing.

Comparing L1 and L2 processing of filler-gap dependencies that result from *wh*-movement, Felser and Roberts (2007) and Marinis et al. (2005) observed that L2ers did not show native-like *trace reactivation* patterns. The Felser and Roberts study employed a cross-modal picture priming task. While listening to sentences such as (2), (L1-Greek) L2ers of English were asked to classify picture targets as *alive* or *not alive* at the pre-gap control position and the gap position.

(2) Fred chased the squirrel to which the nice monkey explained the game's [Pre-Gap Control Position] difficult rules [Gap Position] in the class last Wednesday.

The L2ers exhibited priming effects at both positions for picture probes identical to the antecedents (e.g., *squirrel* in [2]) relative to unrelated picture probes (e.g., a picture of a toothbrush for [2]). The finding contrasts with what was found in Roberts, Marinis, Felser, and Clahsen (2007), in which natives (with high working memory span) displayed antecedent priming effects only at the gap positions for the same experimental sentences. Felser and Roberts interpreted the results as indicating L2ers' maintained activation of the antecedent during the processing of the experimental sentences, as opposed to natives' structurally based antecedent reactivation occurring only at the syntactic gap position.

The study by Marinis et al. employed a self-paced reading task. The primary finding is that in reading sentences like (3), natives postulate an intermediate trace

(marked as e'_i) at the clause boundary (before *that*), but L2ers from different L1 backgrounds (Chinese, Japanese, German, and Greek) do not.

- (3) The nurse who _{i} the doctor argued e'_i that the rude patient had angered e_i is refusing to work late.

In the processing of long-distance filler-gap dependencies, reactivation of the intermediate trace left behind by the moved element facilitates the integration of the *wh*-filler with its licensing verb. On the basis of the L2ers' failure to evince trace reactivation effects (in contrast to the natives), it was concluded that no syntactic structure was used in their processing.

Another source of evidence cited for supporting the SSH comes from some ERP studies (e.g., Chen et al., 2007; Hahne, 2001; Hahne & Friederici, 2001; Mueller, Hahne, Fujii, & Friederici, 2005), in which L2ers, when processing morphosyntactic violations, often fail to show the LAN (left anterior negativity) component, which is thought to reflect early automatic first-pass structure-building processes (e.g., Hahne, 2001; Hahne & Friederici, 2001). As Clahsen and Felser (2006a, p. 118) noted,

The consistent absence of early LAN effects in ERP studies on L2 sentence processing might be taken to suggest that the stage at which initial structures are built automatically on the basis of word category information is skipped altogether in nonnative comprehension.

To summarize, studies that were originally taken as evidence for the SSH have demonstrated some differences between L1 and L2 processing, particularly in structural processing.

However, as emphasized by several L2 researchers (e.g., Dekydtspotter et al., 2006; Miller, 2011, 2014, 2015; Omaki & Schulz, 2011), L2ers' nonnative-like online behavior is not necessarily indicative of shallow processing. The factors that lead to L2ers' nontarget-like patterns may not be directly related to *the nature* of L2 processing. For example, previous research has argued that such factors as automaticity in lexical

access (e.g., Dekydtspotter et al., 2006; Miller, 2014) and task demands (e.g., Miller, 2015; Wen & Schwartz, 2012) can also influence L2 online behavior, and these factors have nothing to do with the nature of the L2 grammar/processor.

In addition, some recent studies have obtained evidence that L2ers are indeed able to engage in detailed and fully specified morphosyntactic processing, thereby challenging the basic arguments of the SSH. For example, the studies by Dussias (2003) and Dussias and Sagarra (2007) provide evidence that (L1-Spanish) L2ers of English (like natives) prefer a low attachment preference in processing RCs online. In this regard, Omaki and Schulz (2011) stressed that structural information is just one of the many competing sources of information the parser deploys in resolving structural ambiguities. L2ers' lack of the native-like structural preference, they rightly pointed out, can only demonstrate that structural information does not receive the highest priority in L2 ambiguity resolution processes, compared to other sources of information; but this does not imply that L2ers do not use structural information.

With regard to L2 processing of long-distance filler-gap dependencies, Miller (2014) demonstrated trace reactivation effects for L1-English L2ers of French when lexical access is made easier by using English-French cognates (e.g., *gorille* 'gorilla') (compared to noncognates [e.g., *renard* 'fox']) as the antecedents of *wh*-fillers. Her study suggests that the (supposed) L2 shallow processing found in previous studies comparing L1 and L2 processing of filler-gap dependencies may simply reflect L2ers' under-routinized lexical retrieval rather than lack of structurally defined intermediate traces. Evidence of L2ers' use of structural information in processing filler-gap dependencies also comes from Omaki and Schulz (2011), in which (L1-Spanish) adult L2ers of English, like natives, observe island constraints during online processing of *wh*-moment, as well as from Pliatsikas and Marinis (2013), in which (L1-Greek) adult L2ers with 9 years of naturalistic exposure to English demonstrate evidence of using intermediate traces online to facilitate *wh*-filler integration.²

As for the absence of L2ers' LAN effects in some ERP studies, counterevidence also exists. LAN effects have been observed among advanced L2ers in several studies (e.g., Gillon Dowens et al., 2010; Ojima et al., 2005; Rossi, Gugler, Friederici, & Hahne,

² They used the same materials as in the original Marinis et al. (2005) study.

2006). Crucially, the SSH argument for L2ers' nonnative-like grammatical processes on the basis of their lack of LAN effects, as McLaughlin et al. (2010) pointed out, rests on the assumption that "the LAN is always present during native language processing; however, there are numerous studies of grammatical processing that have failed to show LAN effects in native speakers" (p. 143). In addition, it is not entirely clear what LAN effects really reflect. A systematic inspection of the existing ERP studies by Steinhauer and Drury (2012) suggests that "there seems to be little evidence for a first phase exclusively dedicated to phrase structure processing" (p. 154).

The SSH has sparked a lot of research, but it has also given rise to questions. In fact, the very nature of L2 shallow processing is still unclear. Does L2 shallow processing always suggest shallow representations? How much does it resemble L1 shallow processing? Shallow processing (also often called *partial*, *incomplete*, or *good-enough processing*) has been discussed a lot in the recent L1 literature on sentence comprehension (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Engelhardt, & Jones, 2009; Ferreira & Patson, 2007; Sanford & Graesser, 2006; Sanford & Sturt, 2002; Swets, Desmet, Clifton, & Ferreira, 2008). According to Ferreira and colleagues, a complete, detailed, and accurate interpretation of the linguistic input for natives is not always feasible and in fact is not always necessary for the purpose of comprehension. Utterances in actual language communications often contain errors (e.g., slips of the tongue), which make full parses almost impossible. Moreover, utterances come and go quickly, and comprehenders might not always have the working memory resources or the processing time to consult all the sources of information so as to build detailed and fully specified representations for those fleeting utterances. In these cases, a good-enough representation could often suffice. The famous *Moses illusion* (Erickson & Mattson, 1981) can illustrate how shallow processing happens. The question *How many animals of each sort did Moses put on the ark?* contains a false presupposition because *Moses* is used in place of *Noah*, the right person. Interestingly, when asked this question, most people answer *Two* rather than *None* (i.e., rejecting the presupposition). It seems that people do not always compute accurate semantic representations for the lexical items they encounter in actual sentence comprehension. In another example, Ferreira and colleagues (Ferreira et al., 2002;

Ferreira & Patson, 2007) mentioned that their experimental studies (e.g., Christianson et al., 2001) indicate that the meaning of initial misanalysis of GP sentences like (4) often persists even after comprehenders have revised their incorrect initial analysis and adopted the correct analysis.

(4) While Anna dressed the baby played in the crib.

Suppose that people can accurately answer the question *Did the baby play in the crib?* after they read (4). This would indicate that they have correctly analyzed *the baby* as the subject of the verb *played* rather than the direct object of the verb *dressed*. Ferreira and colleagues observed that most people who can accurately answer this question also surprisingly reply *Yes* to the question *Did Anna dress the baby?* This suggests that the incorrectly constructed initial representation lingers and interferes with computing the accurate meaning. The misinterpretation of a sentence's meaning like this is not uncommon in actual language use and may go undetected and uncorrected, reflecting the fact that the parser may not always engage in precisely accurate and fully specified processing of the input (Christianson et al., 2001). As Ferreira et al. (2009) pointed out, the human language comprehension system has a tendency to generate shallow and superficial representations (e.g., when faced with complex and difficult sentences), and this is consistent with the principle of minimizing the demands on cognitive resources in human language processing. Sanford and colleagues similarly suggest that underspecification characterizes normal language understanding, and “in fact, fully specified interpretations of language can often seem both undesirable and unnecessary” (Sanford & Sturt, 2002, p. 382). They suggested that the depth of processing is responsive to the task demands of language processing (also see Ferreira et al., 2009) and called for “a better understanding of the conditions under which degrees of shallow processing occur and of the nature of the resulting representations” (Sanford & Graesser, 2006, p. 106). With respect to the fact that initial misinterpretations may persist, some recent proposals (e.g., Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013) suggest that the persistence may not be due to the parser's failure in building fully specified structural representations, but rather to its' failure in completely wiping out initially

constructed representations. Regardless of different theoretical accounts for L1 shallow processing phenomena, one thing is clear: It is not a problem with the grammar. If L1 shallow processing is not indicative of a deficient grammar, then, logically, why should we say that L2ers' shallow processing *is* necessarily indicative of a deficient or deviant grammar? Moreover, under what conditions will L2 processing exhibit the characteristics of shallow processing? This dissertation research addresses these issues, especially in Chapter 2 (by comparing the use of tasks with different demands) and Chapter 5 (by the use of shallow-targeted comprehension questions).

As a recent movement, proponents of the SSH contend that L2 processing relies more on meaning-based information such as lexical semantics (e.g., Roberts & Felser, 2011) and discourse contexts (e.g., Pan & Felser, 2011; Pan et al., 2015) than L1 processing does and that L2ers have more difficulty in reanalysis than natives do in processing structural ambiguities (e.g., Jacob & Felser, in press; Roberts & Felser, 2011). However, these claims do not seem to be supported by the results of the present research (see Chapters 5 and 6).

In addition, the SSH stresses that L1 and L2 processing differences persist in the domain of complex syntax, and that L2ers are restricted to only shallow processing for structurally complex phenomena, because, for example, “the L2 grammar does not provide the type of syntactic information required to process nonlocal grammatical phenomena in native-like ways” (Clahsen & Felser, 2006c, p. 565). Clahsen and Felser recommended using complex syntax to study whether L2 processing is ultimately restricted to shallow processing: “To determine the upper limits of L2 grammatical processing ability, more research is needed, focusing on complex structural phenomena” (Clahsen & Felser, 2006c, pp. 568–569). However, using complex syntax to study L2 processing is inherently problematic. L1 research suggests that complex structures are difficult to compute because they are cognitively demanding (Gibson, 1998) and tend to induce shallow processing even among natives (Ferreira et al., 2009). If we use complex syntax to study L2 processing, one question is inevitable: How can we tease apart the confounding of cognitive demands from shallow syntactic representations? Because of this, complex structures may not be the most suitable for investigating L2 online processing: Even if L2ers fail to show online sensitivity to manipulations of complex

syntax, this is open to alternative interpretations based on cognitive limitations. It could certainly mean that L2ers' limited cognitive resources rather than deficient grammar have prevented them from building target-like syntactic representations online. It seems that to study the nature of L2 processing, we should, in fact, try to ease L2ers' processing burden so that their true ability will not be concealed by the cognitive strain arising from dealing with complex syntax (for how this issue is dealt with in this dissertation project, see Chapters 2 and 4).

Besides avoiding the use of overly complex linguistic phenomena, which may be cognitively taxing, we should consider the influence of task demands on the depth of processing. In this respect, standard L1 sentence processing research methodologies may not always be suitable for studying L2 sentence processing. For example, the traditional self-paced reading task used in L1 research usually focuses participants on comprehending overall sentence meaning. However, comprehension-focused self-paced reading (CFSPR) tasks may tend to induce shallow processing among L2ers. This dissertation research addresses this issue by proposing that with the goal of studying L2ers' full parsing, we should use detail-focused tasks such as acceptability-judgment self-paced reading task (AJSPR; see Chapter 2 for details).

To date, most studies that appear to support the SSH are based on the observations of L2 processing in languages that make use of inflectional morphology, such as English, German, and Greek. Can findings from those studies (e.g., Chen et al., 2007; Hahne, 2001; Hahne & Friederici, 2001; Jiang, 2004, 2007) be generalized to languages, such as Chinese, that have virtually no inflectional morphology? Chinese is a less syntactically constrained language (than, say, English), and sentence comprehension in Chinese relies heavily on semantic properties such as animacy (e.g., P. Li, Bates, & MacWhinney, 1993; Miao, 1999). Can L2 processing theories such as the SSH hold for L2 Chinese sentence processing? For example, if L2ers are good at using meaning-based information such as lexical semantics and discourse context, will this give them an edge in dealing with Chinese? If comprehension in Chinese is relatively less reliant on morphosyntactic information, will sentence processing in L2 Chinese still be difficult? If not, what processing difficulties do L2ers face in Chinese sentence processing? The present dissertation research illuminates these issues by exploring L1 and L2 sentence

processing in Chinese (Chapters 4–6), a language typologically different from English, German, Greek, etc.

Briefly, the major research questions addressed in this dissertation project include:

- (a) To what extent are L2ers able to (come to) use different sources of information (morphosyntactic, lexical-semantic, and discourse-context) in online sentence processing?
- (b) What are the underlying causes for L2 shallow processing (e.g., lack of the TL knowledge, costly task demands, insufficient proficiency in the TL, and/or L1 transfer)?
- (c) Can L2 processing become “deep,” in the sense that it does not (inevitably) lack grammatical detail? To answer these questions, three main studies were conducted in Chinese, focusing on L2ers’ use of morphosyntactic, lexical-semantic, and discourse-context information in processing Chinese sentences. These studies are reported, respectively, in Chapters 4, 5, and 6. The specific theoretical and empirical background, research rationales and designs, and predictions are also given in these chapters.

The rest of the dissertation is structured as follows:

Chapter 2 deals with important methodological issues crucial to the investigation of fully specified L2 processing. By comparing L1 and L2 processing of subject-verb number agreement in English in an AJSPR task with a CFSPR task in Wen (2007), the study establishes AJSPR as a useful tool for assessing L2ers’ fully specified processing.

Chapter 3 gives a description of the general methods used in the three studies of Chinese sentence processing carried out in this dissertation research.

Chapter 4 compares L1 and L2 processing of negation–aspect interactions in Chinese to test whether adult L2ers whose L1 lacks such morphosyntax can ultimately acquire it and become able to use it in online sentence processing.

Chapter 5 examines whether L2ers, relative to natives, are necessarily over-reliant on lexical-semantic information in processing Chinese sentences. This is achieved by comparing L1 and L2 participants’ sensitivity to local semantic (im)plausibilities that were created by exploiting the selectional restrictions of Chinese transitive verbs.

Chapter 6 investigates how extrasentential discourse context influences L1 and L2 processing of null objects in Chinese in order to answer the question of whether, in comparison to natives, L2ers are necessarily over-dependent on discourse-context

information in sentence processing or whether they can (come to) make online use this information in the same way as natives do.

Chapter 7 presents the conclusion of the dissertation research.

This dissertation project represents, we would like to suggest, a pioneering study in (L2) Chinese sentence processing. Most current theories on L1 and L2 sentence processing are formulated in the context of (primarily Indo-European) languages which have inflectional morphology. Their generalizability is seldom addressed in languages, such as Chinese, that have virtually no inflectional morphology or in languages, again such as Chinese, that are considered especially susceptible to (pragmatic and discourse) context considerations. In the current geopolitical context, furthermore, research on Chinese sentence processing is badly needed. Studying how natives and L2ers process sentences in Chinese can help researchers obtain a much richer picture of human sentence processing in general and L2 sentence processing in particular.

CHAPTER 2

METHODOLOGY FOR PROBING FULLY SPECIFIED L2 PROCESSING³

This chapter deals with the methodology for assessing full parsing abilities in a second language (L2). We propose an acceptability-judgment self-paced reading (AJSPR) task as a feasible tool for achieving this goal.

2.1 Introduction

As mentioned in Chapter 1, standard online sentence processing tasks such as comprehension-focused self-paced reading (CFSPR), which are often used in first language (L1) sentence processing research, may not be suitable for investigating the full parsing abilities in L2 learners (“L2ers”). Comparing L1 and L2 processing in CFSPR-type tasks may conceal L2ers’ ability to process detailed morphosyntactic information that is (often) nonessential to understanding overall sentence meaning. This issue is of crucial importance for research on L2 processing and L2 acquisition in that whether L2ers can engage in full parsing determines the nature of the representations L2ers construct for the analyzed input during online processing. If L2 processing is restricted to shallow processing only, as the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, 2006b, 2006c) asserts, L2ers can only construct shallow representations online. The consequence is that the L2 grammar will never converge on the target language (TL) grammar, even as L2ers’ proficiency in the TL increases.

However, online processing tasks such as CFSPR may not be able to reveal the true nature of L2 processing. For L2ers whose access to (Interlanguage) grammar is not so automatic, comprehension in the L2 can be stressful. It is sometimes an insurmountable task for L2ers to attend to both structural details and overall meaning at the same time. There is evidence that even for mature adult L1 speakers (“natives”), their syntactic processing can suffer under stress or increased memory cost (e.g., Blackwell & Bates, 1995; McDonald, 2006). If L2ers are compelled to exercise shallow processing as a strategy in coping with challenging comprehension tasks, this implies that those tasks cannot tell us whether L2ers have the relevant morphosyntactic knowledge or not. In this

³ An earlier version of portions of this chapter was published in Wen and Schwartz (2012).

sense, comparing L1 and L2 morphosyntactic processing on holistic comprehension tasks can prejudice against L2ers if such tasks are used to evaluate L2ers' full parsing abilities: It is not a comparison on equal footing. Suppose a person who carries a heavy load runs a little more slowly than others who carry less heavy loads. Can we say for sure, in this case, that this person is a slow runner? Likewise, if L2ers show divergent processing behavior compared to natives in CFSPR tasks, the divergence cannot lead to the logical conclusion that L2ers are unable to use full parsing routines.

To put L2ers back on equal ground with natives, we should seek online tasks that can stimulate fully specified processing and see whether L2ers still stick to shallow processing routines under such task conditions. Such a task should not focus on overall sentence meaning; instead, it should be details-oriented. In this way, the task encourages the use of full parsing routines. In this chapter, we will demonstrate that AJSPR is an appropriate candidate to this end. AJSPR encourages fully specified processing, because the online pressure of making an acceptability judgment pushes the parser to build the sentence structure immediately by making use of all available information. The experiment in this chapter will provide evidence that AJSPR can tap into deep-level L2 processing and can elicit L2 full parsing that CFSPR may not always be able to reveal. Given that AJSPR is sensitive in detecting deep-level parsing processes, we argue that if our primary objective is to evaluate L2ers' full parsing abilities, L2 sentence processing research should seek online methodologies, such as AJSPR, that focus both natives and L2ers on more fully specified processing (i.e., 'detailed' or 'deep' processing) so as not to conceal the ability L2ers (may) have in this type of processing in the TL.

Although AJSPR-type tasks have been employed in sentence processing research on both L1 (e.g., Frazier & Clifton, 1989; Katsika, 2009) and L2 (Jackson & Dussias, 2009; Juffs, 1998a, 1998b, 2004, 2005; Juffs & Harrington, 1995, 1996), they have so far not been used specifically for the purpose of assessing deep-level L2 processing. To validate AJSPR as a useful tool for investigating fully specified L2 processing, we will need to look at a morphosyntactic phenomenon that is particularly difficult for L2ers to process, a phenomenon that often leads to L2 shallow processing in CFSPR tasks. What we have chosen is subject-verb number agreement in English.

2.2 L2 Processing of Subject-Verb Number Agreement in English

Subject-verb number agreement in English is notoriously troublesome for L2ers. Some L2 researchers (e.g., Hawkins & Liszka, 2003; Jiang, 2004; Tsimpli & Dimitrakopoulou, 2007) believe this morphosyntactic phenomenon is not acquirable for L2ers whose L1 does not instantiate it. Available online evidence indeed seems to corroborate this view: L2ers whose L1 lacks subject-verb number agreement, even at advanced proficiency, appear unable to show online sensitivity to violations of such agreement (e.g., Jiang, 2004; Wen, 2007).

In the study by Jiang (2004), “advanced” L1-Chinese⁴ L2ers of English (henceforth “Chinese-English L2ers”) and native controls participated in a self-paced reading task, using experimental sentences, such as (1), that involved subject-verb number (dis)agreement.⁵

- (1) a. The bridges to the island *were* about ten miles away.
b. * The bridge to the island *were* about ten miles away.

Participants were asked to read as quickly as possible for comprehending sentence meaning. Half the experimental sentences were followed with *yes/no* questions to check comprehension. To complete the reading task, participants thus needed to focus on comprehending meaning. The reading time (RT) data showed that L1 participants spent significantly longer at the region (*about*) immediately following the critical (dis)agreeing verb region (*were*) in (1b) than in (1a), whereas L2 participants did not show any significant differences in their RTs at either the critical (dis)agreeing region or the post-critical region.

The study by Wen (2007) also compared L1 and L2 processing of subject-verb number (dis)agreement in a CFSPR task,⁶ namely, a truth-value-judgment self-paced reading task (see Section 2.3). The experimental stimuli are like in (2).

⁴ Chinese has no subject-verb number agreement.

⁵ The verbs were either *was* or *were*.

⁶ The verbs were all lexical verbs.

- (2) a. Beautiful mountain scenery often *attracts* people from big industrial cities.
b.* Beautiful tourist sites often *attracts* people from big industrial cities.
c. Beautiful tourist sites often *attract* people from big industrial cities.
d.* Beautiful mountain scenery often *attract* people from big industrial cities.

The results showed that English natives were sensitive to agreement violations only when the lexical verbs are overtly marked for number, as in (2b) (viz., *not* in [2d]), and that intermediate-to-advanced Chinese-English L2ers were not sensitive to subject-verb number disagreement in either disagreement condition.

The present study employs the exact same materials as in Wen (2007), but in an AJSPP task. This thus enables a direct comparison as regards how task demands modulate the depth of L2 morphosyntactic processing. Since the present study builds upon the Wen (2007) study, we therefore describe it in more detail in the following section.

2.3 Detailed Description of Wen (2007)

The CFSPR task in Wen (2007) required participants to decide for each trial whether the proposition expressed in the sentence they read is “true” or “false.” The participants were instructed to try to concentrate on the meaning of each sentence and ignore grammatical errors. They were also instructed to read each sentence as quickly as possible. These instructions were intended to maximize the possibility that participants focus on overall sentence meaning rather than form, because attending to form would not help with true-false judgments (for the instructions, see Appendix A).

The study used 16 sets of experimental sentences like (2) distributed across 4 presentation lists in a Latin square design. Each sentence contains 10 regions (i.e., words). A sample set of experimental stimuli is provided in Table 2.1.

Table 2.1. *Sample Set of Experimental Stimuli in Wen (2007)*

Condition	Region									
	1	2	3	4	5	6	7	8	9	10
Agree-Vsg	Beautiful	mountain	scenery	often	attracts	people	from	big	industrial	cities.
Disagree-Vsg	* Beautiful	tourist	sites	often	attracts	people	from	big	industrial	cities.
Agree-Vpl	Beautiful	tourist	sites	often	attract	people	from	big	industrial	cities.
Disagree -Vpl	* Beautiful	mountain	scenery	often	attract	people	from	big	industrial	cities.

The *Agree-Vsg* and *Agree-Vpl* conditions are grammatical, as the subject and the verb agree in number; the *Disagree-Vsg* and *Disagree-Vpl* conditions are ungrammatical because the subject and the verb disagree in number. The regions of interest are 5 and 6, where participants are likely to show disagreement effects. There were 64 fillers, of which 16 were ungrammatical, with the errors appearing in various regions. Half the filler items were true. There were also 6 practice sentences at the beginning of the experiment. (Appendix B lists all the experimental items, fillers, and practice sentences.) All the sentences in the experiment were presented using a noncumulative word-by-word self-paced moving-window paradigm (Just, Carpenter, & Woolley, 1982).

The prediction was that if participants process detailed morphosyntactic information of subject-verb number agreement, they should spend more time reading regions at and/or following a violation of agreement such as *attract(s) people* in (2b) relative to (2a), and in (2d) relative to (2c).

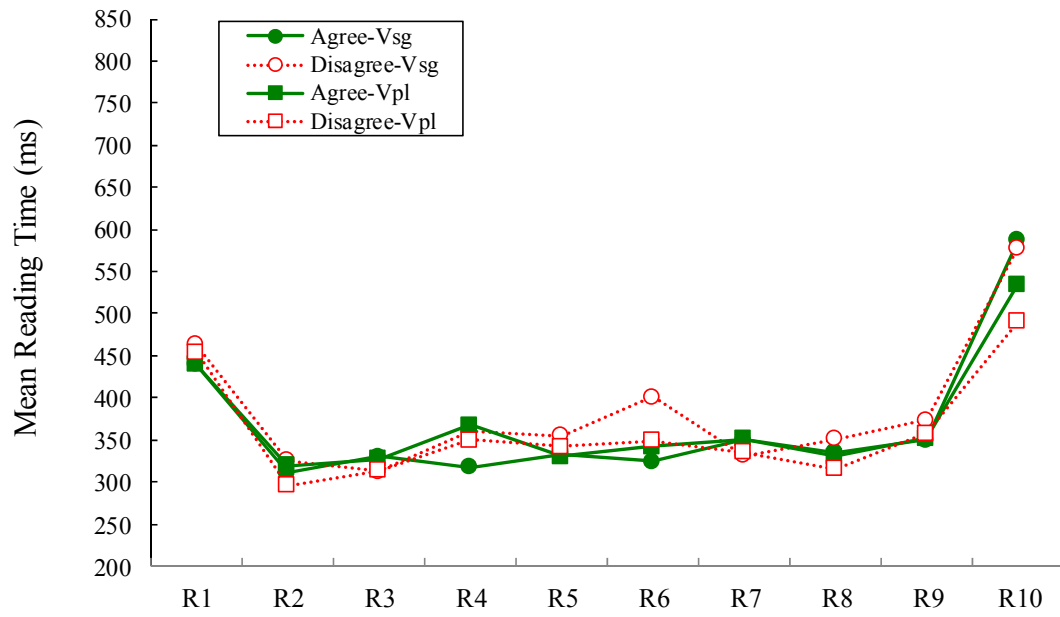
The participants were 28 natives and 28 intermediate-to-advanced Chinese-English L2ers. Table 2.2 gives the English background information of the L2ers.

Table 2.2. *Chinese Participants' English Background in Wen (2007)*

Proficiency Estimate	Mean	SD
Years of learning English	16.18	4.51
Years of residence in the U.S.	2.16	1.63
Self-rating of overall proficiency ^a	4.18	0.86

^a The rating scale was 1 (*beginning*) to 5 (*advanced*)

The natives' and L2ers' online results are presented in Figures 2.1 and 2.2, respectively. As can be seen (at Region 6), the natives were sensitive to agreement violations only when the verbs are overtly marked for number, as in (2b). The L2ers were not sensitive to number disagreement in either type (2b) or (2d).



- a. Beautiful mountain scenery often attracts people from big industrial cities.
- b. * Beautiful tourist sites often attracts people from big industrial cities.
- c. Beautiful tourist sites often attract people from big industrial cities.
- d. * Beautiful mountain scenery often attract people from big industrial cities.

Figure 2.1. Natives' Mean Reading Time (msec) Profile in Wen (2007).

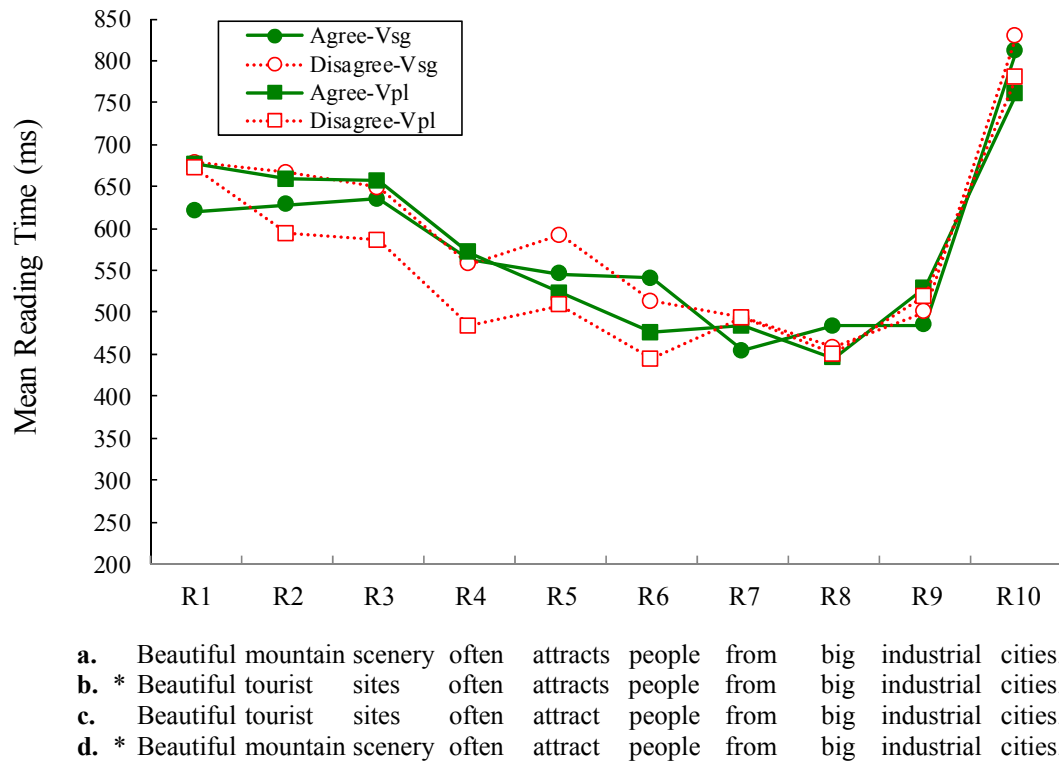


Figure 2.2. L2ers’ Mean Reading Time (msec) Profile in Wen (2007).

The results indicate that even natives ignore detailed morphosyntactic information (i.e., in type [2d]), when the processing task is extremely comprehension-focused. This finding leads us to hypothesize that the parser makes less use of detailed morphosyntactic information in certain CFSPR tasks.

2.4 Observations and Thoughts

As mentioned above, participants in the Jiang (2004) and Wen (2007) studies were required to focus on comprehending overall sentence meaning. This focus may tempt L2ers (and even natives) to strategically overlook agreement information, because heeding subject-verb agreement in English (as in [2]) typically does not aid meaning comprehension (e.g., VanPatten, 2002). If L2ers, when under online comprehension pressure, ignore agreement information as a strategy and thus exhibit “shallow processing” of such nonessential morphosyntactic information, this implies that L2 insensitivity to disagreement found in earlier studies may simply reflect a performance strategy that results from task demands.

Some empirical evidence does seem to suggest that task demands affect the depth of L2 morphosyntactic processing. For instance, in two self-paced reading studies, Jackson and colleagues (Jackson & Bobb, 2009; Jackson & Dussias, 2009) investigated native and nonnative processing of German case marking involved in subject vs. object extractions across clause boundaries. The L2ers, whose L1 was English, were highly proficient in German. The two studies contained similar materials, such as in (3), manipulating both extraction type (subject extraction vs. object extraction) and matrix-clause type (present tense vs. present perfect).

(3) a. Embedded subject extraction, present-tense matrix clause

Wer denkst du, bewunderte den Sportler nach dem Spiel?
who_{NOM} think you admired the_{ACC} athlete after the game
'Who do you think admired the athlete after the game?'

b. Embedded object extraction, present-tense matrix clause

Wen denkst du, bewunderte der Sportler nach dem Spiel?
who_{ACC} think you admired the_{NOM} athlete after the game
'Who(m) do you think the athlete admired after the game?'

c. Embedded subject extraction, present-perfect matrix clause

Wer hast du gedacht, bewunderte den Sportler nach dem Spiel?
who_{NOM} have you thought admired the_{ACC} athlete after the game
'Who did you think admired the athlete after the game?'

d. Embedded object extraction, present-perfect matrix clause

Wen hast du gedacht, bewunderte der Sportler nach dem Spiel?
who_{ACC} have you thought admired the_{NOM} athlete after the game
'Who(m) did you think the athlete admired after the game?'

The two studies differed in task demands: In Jackson & Dussias (2009), participants were asked to make an acceptability judgment after reading each sentence (i.e., an AJSPR-type task); in Jackson & Bobb (2009), participants were instead asked to

verify a statement constructed to focus them on comprehending holistic sentence meaning (i.e., a CFSPR-type task). The results are very interesting: The L2ers in both studies showed a subject preference when processing the matrix clauses, just like natives; however, when processing the complement clauses, which involved a long-distance *wh*-extraction (i.e., across a major clause boundary), the L2ers showed a subject preference only in the first study (which required participants to make end-of-sentence acceptability judgments). Jackson and Bobb (2009, pp. 630-631) suggested that the different results indicate the effects of task demands in L2 morphosyntactic processing. They explained that the online acceptability-judgment task seemed to have drawn participants' attention to the case-marking information and led the L2ers to prioritize some case-matching strategy. An alternative explanation they offered is that L2ers in the comprehension-focused task might have relied on partial, good-enough processing.

The two studies above thus illustrate that AJSPR-type tasks can elicit some deep processes that CFSPR-type tasks might not be able to.⁷

2.5 The Present Study: Acceptability-Judgment Self-Paced Reading

The present study tests whether there are task effects in L2 processing of subject-verb number agreement in English by employing AJSPR, a structure-focused online task, in comparison to Wen (2007), which used CFSPR. Superficially, AJSPR differs from CFSPR only in end-of-sentence task demands: judging acceptability vs. answering a comprehension question. Yet, the different task demands of AJSPR and CFSPR may provoke two different processes: deep vs. shallow processing. In the deep processing, the language processing mechanism (the parser) computes structural details in an immediate and incremental fashion and builds (more) fully specified structural representations. In the shallow processing, the parser is more likely to underspecify structural details and build only partial, sketchy representations, simply because detailed morphosyntax is not always necessary for comprehending meaning. L2 shallow processing exhibited in (difficult) online comprehension tasks has led some researchers (e.g., Clahsen & Felser, 2006a, 2006b, 2006c) to propose that L2 processing operates on

⁷ For discussions relevant to task effects in L2 sentence processing, see Miller (2015) and Roberts (2012a, 2013).

a deficient representation. Therefore, characterizing these two types of processing is vital for making inferences about the potential nature of underlying (L2) knowledge.

2.5.1 Method

2.5.1.1 Participants

This study included 16 English natives and 16 adult Chinese-English L2ers (3 natives and 4 L2ers were excluded for high error rates, $M \geq 31.3\%$). The Chinese participants' English background is given in Table 2.3.

Table 2.3. *Chinese Participants' English Background in the AJSPR Task*

Proficiency Estimate	Mean	SD
Years of learning English	16.38	4.46
Years of residence in the U.S.	1.08	1.94
Self-rating of overall proficiency ^a	4.63	0.50

^aThe rating scale was 1 (*beginning*) to 6 (*near native*)

A comparison of the L2 English proficiency in Wen (2007) and in the present study is presented in Figure 2.3.

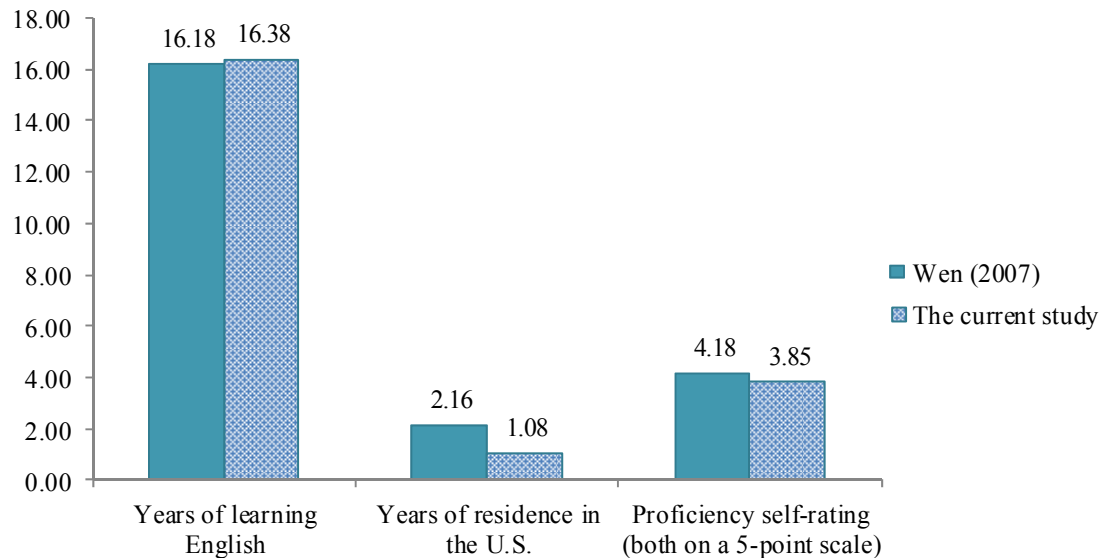


Figure 2.3. Comparison of L2 Proficiency in Wen (2007) and the Current Study.

There was no significant difference in length of learning English, $t(42) = .14, p = .890$. The L2ers in Wen (2007) spent more years in the U.S. than those in the present study,

$t(42) = 1.98, p = .054$. The L2ers' self-ratings in this study, after converted to a 5-point scale, seemed a little lower than in Wen (2007), but the difference was not significant, $t(41) = 1.68, p = .101$. Taken together, we maintain that the L2ers' English proficiency in these two studies is comparable.

2.5.1.2 Materials

The materials in this study are exactly the same as in Wen (2007). This way we can directly compare the results from the two different tasks. (For a complete list of the experimental items, fillers, and practice sentences, see Appendix B.)

2.5.1.3 Task and Procedure

The online task is a word-by-word noncumulative moving-window AJSPR task. After reading each sentence, participants were prompted to judge whether the sentence is acceptable in terms of its well-formedness (see Appendix A for the exact instructions used in the experiment). Other than this, the procedure is exactly the same as in Wen (2007).

2.5.2 Results

The participants' performance is assessed by both judgment accuracy and RTs. Figure 2.4 gives the mean acceptability-judgment accuracy rates of the 2 participant groups. (The descriptive statistics of the experimental sentences as well as the fillers are given in Appendix C.)

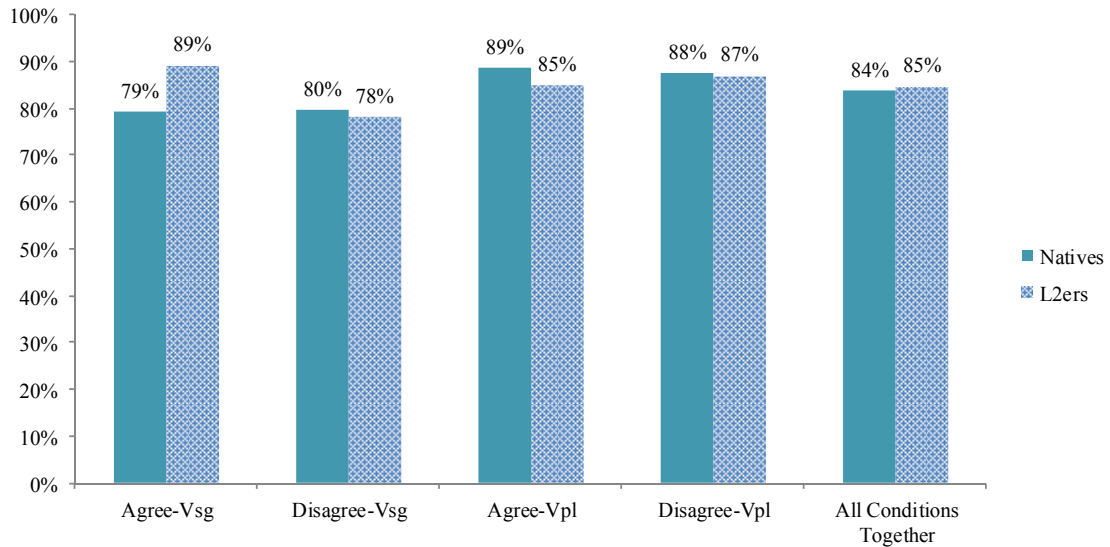


Figure 2.4. Natives' and L2ers' Mean Accuracy Rates for the Experimental Sentences in the AJSPR Task.

A 2 (Group [Natives, L2ers]) \times 4 (Condition [Agree-Vsg, Disagree-Vsg, Agree-Vpl, Disagree -Vpl]) repeated measures ANOVA was performed on the arcsine-transformed accuracy data of the experimental sentences. No significant main effect or interaction emerged ($F_s < 1.33$, $p_s > .27$). On average, the accuracy rates for the experimental sentences in both the L1 group and the L2 group were high: equal to or greater than 84%. Further, the participant groups' mean acceptance rates by condition, as presented in Figure 2.5, indicate that both groups accepted the two grammatical conditions as grammatical to a much higher degree than they did the two ungrammatical conditions.

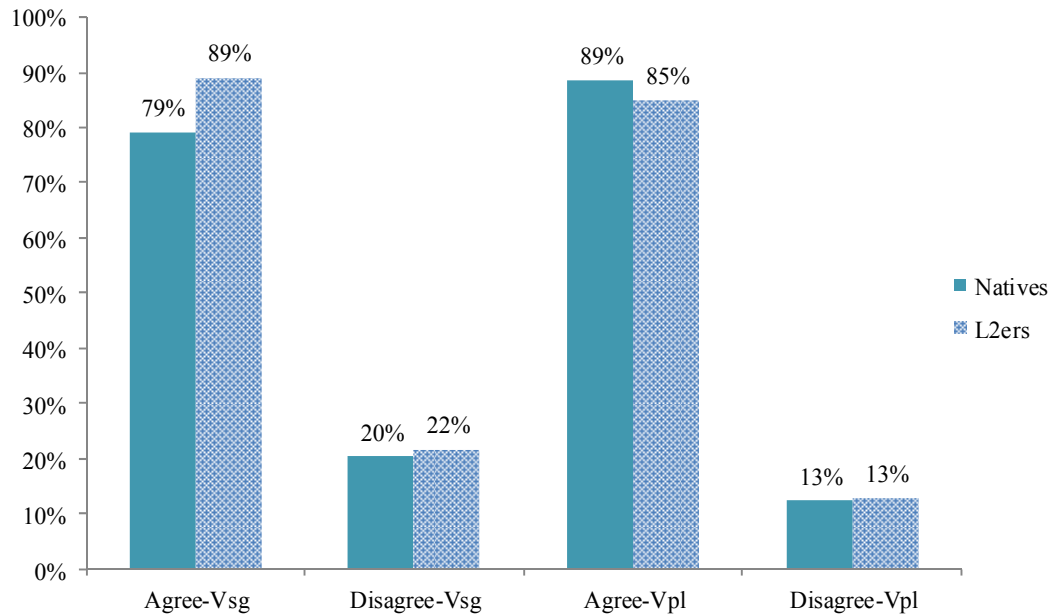
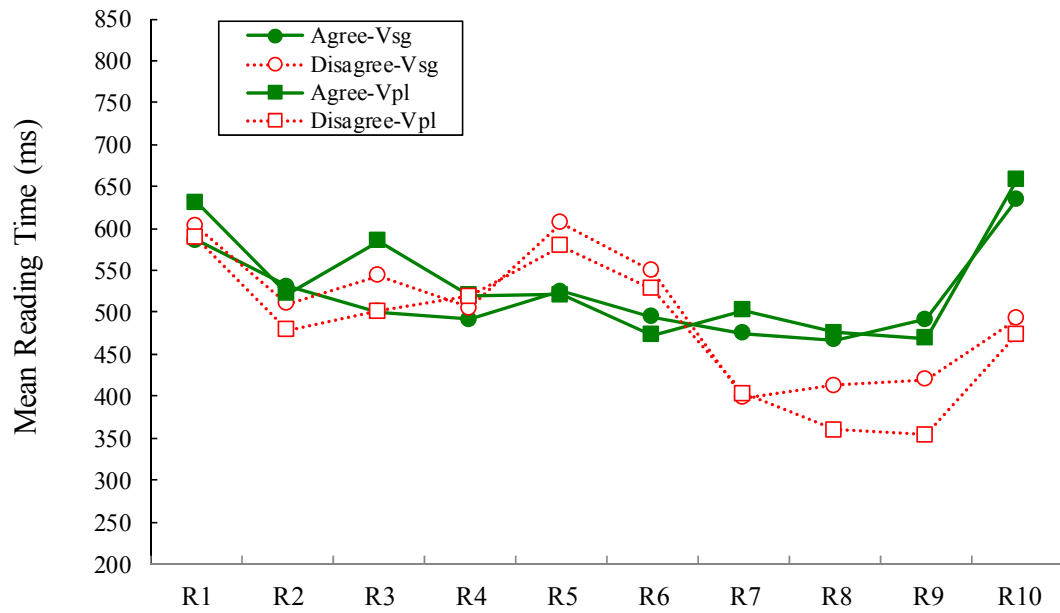


Figure 2.5. Natives' and L2ers' Mean Acceptance Rates for the Experimental Sentences in the AJSPR Task.

A 2 (Group [Natives, L2ers]) \times 2 (Disagreement [agree, disagree]) \times 2 (Verb Number [Vsg, Vpl]) repeated measures ANOVA, performed on the arcsine-transformed acceptance data, showed that this effect was highly significant, $F(1, 30) = 159.44$, $p < .001$. There was no main effect for group or interaction effects ($F_s < 2.70$, $p_s > .11$). As for the two groups' accuracy rates for the fillers, they were also high ($\geq 80\%$). These results indicate that the participants were attentive in performing the online reading task and that the L2ers were similar to natives in their knowledge of subject-verb number agreement in English.

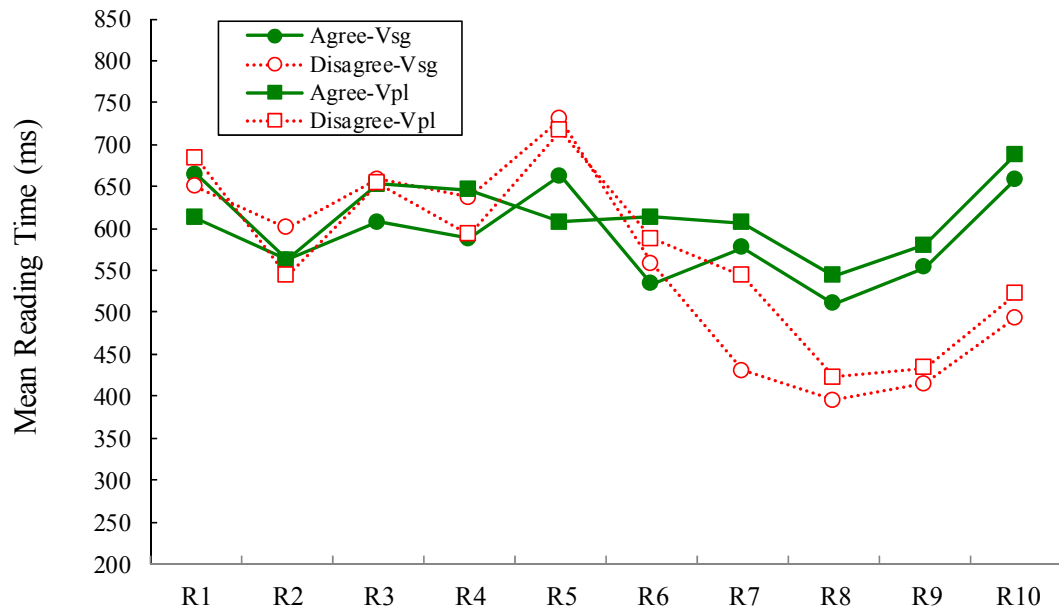
For the raw RT data, a two-step data trimming procedure preceded statistical analysis: (a) RTs of a region (a word) longer than 1,500 ms were replaced with this cutoff; (b) for each participant, RTs equal to or larger than 2 *SDs* from his/her mean RT were replaced with his/her cutoff. This affected 4.4% of the L1 data and 7.1% of the L2 data. We then ran separate 2 (Disagreement [agree, disagree]) \times 2 (Verb Number [Vsg, Vpl]) repeated measures ANOVAs on each group's RTs at Regions 5 and 6 (the critical and post-critical regions).

Figures 2.6 and 2.7 show mean RTs for each region by experimental condition. The trimmed raw RT means are provided in Appendix D.



- a. Beautiful mountain scenery often attracts people from big industrial cities.
- b. * Beautiful tourist sites often attracts people from big industrial cities.
- c. Beautiful tourist sites often attract people from big industrial cities.
- d. * Beautiful mountain scenery often attract people from big industrial cities.

Figure 2.6. Natives' Mean Reading Time (msec) Profile in the AJSPR Task.



- a. Beautiful mountain scenery often attracts people from big industrial cities.
- b. * Beautiful tourist sites often attracts people from big industrial cities.
- c. Beautiful tourist sites often attract people from big industrial cities.
- d. * Beautiful mountain scenery often attract people from big industrial cities.

Figure 2.7. L2ers' Mean Reading Time (msec) Profile in the AJSPR Task.

From Figures 2.6 and 2.7, we can see that both natives and L2ers spent more time at Region 5 (and also Region 6 for natives) in ungrammatical conditions than grammatical conditions.

For the natives, the main effect of disagreement at Region 5 was marginally significant in the participant analysis, $F_1(1, 15) = 4.08, p = .062, \eta_p^2 = .21$, but not significant in the item analysis, $F_2(1, 15) = 2.96, p = .106, \eta_p^2 = .17$. The disagreement effect became significant at Region 6 in the participant analysis, $F_1(1, 15) = 5.24, p = .037, \eta_p^2 = .26$, and marginal in the item analysis, $F_2(1, 15) = 4.27, p = .056, \eta_p^2 = .22$. There was neither a main effect of verb number nor an interaction between disagreement and verb number at Region 5 or Region 6, all $F_s \leq 0.89, p_s \geq .36$. Pairwise comparisons (two-tailed) indicated that in the participant analysis, when the verb was plural, the difference between grammatical and ungrammatical conditions reached marginal significance at Region 5, $t_1(15) = 1.92, p = .075; t_2(15) = 1.65, p = .119$, and became significant at Region 6, $t_1(15) = 2.22, p = .043; t_2(15) = 1.70, p = .109$. When the verb was singular, however, the difference between grammatical and ungrammatical

conditions did not reach statistical significance, either at Region 5, $t_1(15) = 1.62, p = .126$; $t_2(15) = 0.70, p = .492$, or at Region 6, $t_1(15) = 1.43, p = .172$; $t_2(15) = 1.45, p = .168$, although numerically very large.

The L2ers' results showed a significant main effect of disagreement at Region 5, $F_1(1, 15) = 5.89, p = .028, \eta_p^2 = .28$; $F_2(1, 15) = 8.13, p = .012, \eta_p^2 = .35$. Pairwise comparisons (two-tailed) indicated that when the verb was plural, the difference between grammatical and ungrammatical conditions was significant in the participant analysis, $t_1(15) = 2.26, p = .04$, and marginal in the item analysis, $t_2(15) = 2.06, p = .057$. When the verb was singular, the difference between the grammatical and ungrammatical conditions was marginally significant in the participant analysis $t_1(15) = 1.96, p = .068$; $t_2(15) = 1.69, p = .112$. No other main effects or interactions were found at Regions 5 or 6 ($F_s < 2.30, p_s > .10$).

2.5.3 Discussion

This study tested whether task demands can affect depth of L2 processing. We employed an AJSPR task in our experiment and found that both the natives and the L2ers are sensitive to number disagreement in the structure-focused online task. Recall that advanced Chinese-English L2ers were not found to be sensitive to subject-verb number disagreement in comprehension-focused online tasks (in, e.g., Jiang, 2004; Wen, 2007). Our study suggests that task demands indeed make a difference in L2 (and L1) processing of subject-verb number agreement in English.

Our study tackles a critical issue in L2 sentence processing research, that is, whether L2ers are able to engage in fully specified morphosyntactic processing after all. We approach this issue from a methodological perspective and establish AJSPR as a useful tool for assessing L2 deep structural processing.

Most L2 sentence processing research has so far used CFSPR (see Section 2.4). However, CFSPR tasks may be unsuitable for assessing L2 deep processing. First, CFSPR may encourage L2ers to use shallow processing as an effort-saving strategy. For the purpose of comprehending or communicating meaning, there is no need to process detailed structural information all the time. For example, processing subject-verb agreement in *John often drives to work* offers no help in extracting the propositional

content. If L2ers, under the pressure of real-time comprehension or communication, strategically ignore the 3sg present inflectional morpheme *-s* on the verb, it costs them nothing to arrive at the correct meaning. This is perhaps why comprehension-focused online studies (e.g., Jiang, 2004; Wen, 2007) often fail to find L2 sensitivity to violations of subject-verb number agreement. Typically, comprehension in L2 is more effortful than in L1. This is evidenced in the general slowness of L2 processing relative to L1 processing. If L2ers in CFSPR tasks can comprehend meaning without needing to engage in detailed structural processing and can thus save memory resources for other processes, why should they bother to do effortful deep processing? In effect, CFSPR tasks serve as a temptation for L2ers to practice shallow processing. And if L2ers are provoked to use shallow processing as a strategy in CFSPR tasks, this implies that CFSPR tasks may not tell us whether L2ers have the ability to do deep processing or not.

Second, CFSPR tasks may limit L2ers' engagement in deep processing. As discussed in Section 2.1, for L2ers whose access to (Interlanguage) grammar is not so automatic, attending simultaneously to both structural details and overall meaning in online comprehension may exceed their cognitive capabilities. In this respect, using CFSPR tasks to assess L2ers' deep processing biases against them. Accessing grammatical knowledge, whether target-like or not, can overtax L2ers' processing resources and as a result compromise overall performance. In other words, overtaxing memory resources may prevent the parser from accessing knowledge in online comprehension and seduce L2ers to practice shallow processing.

Because L2 comprehension is notably harder than L1 comprehension, we need tasks that are not so focused on holistic comprehension if we continue to use L1 processing as the L2 benchmark; and because we want to examine deep-level L2 processes, we need tasks that require detailed structural processing so as to put comparisons between L1 and L2 processing on relatively equal terms.

For these reasons, we contend that AJSPR is appropriate for assessing L2 deep processing. In an AJSPR task, participants have to build detailed structures immediately and incrementally. This is what the task demands. If they do not build detailed structures, they cannot make a (considered, nonrandom) acceptability judgment at the end of each sentence. If they do not build them immediately and instead wait till the end of the

sentence, they may forget the beginning part by the time they reach the end, because in the self-paced task they cannot go back to look at what they already read. Participants must also use all available information at every word to build the current structure, including morphosyntactic information (as well as lexical-semantic information, pragmatic information, etc.). In other words, AJSPR encourages detailed incremental processing. In contrast, in a CFSPR task, the parser just needs to care about the overall general interpretation and can ignore unessential details of grammar which do not affect extracting sentence meaning. So the task demands of AJSPR and CFSPR (may) promote different types of processing.

Researchers may worry that participants in AJSPR tasks may be induced to resort to metalinguistic knowledge when making acceptability judgments (e.g., Keating & Jegerski, 2015; Roberts, 2013). However, this is not really worrisome in this design. If participants make use of metalinguistic information in making the acceptability judgment, they are likely to turn to it only after they have finished reading (the relevant parts of) the sentence. That is to say, they must first construct the structure from the input, after which they can then consult metalinguistic knowledge to see whether what they built does or does not conform to the relevant metalinguistic rules/knowledge. This is usually unlikely to happen before they finish reading the sentence, at least not before they detect a violation of grammaticality.

Furthermore, our primary concern in AJSPR tasks is *the length of time participants spend on each word before they reach the end of the sentence*. We do not care so much about whether participants can make target-like end-of-sentence acceptability judgments that in principle are more susceptible to the influence of metalinguistic knowledge. In the midst of reading a sentence, participants are rather unlikely to access metalinguistic knowledge, even if they want to. There are several reasons why this is unlikely. First, participants in (our) AJSPR tasks typically cannot easily predict where the error or anomaly occurs. Acceptability-judgment tasks can usually be designed to include different kinds of ungrammatical fillers to make the location of ungrammaticality unpredictable. If participants do not know where the ungrammaticality will occur, how can they anticipate the morphosyntactic rule(s) being tested in the process of reading the sentence? In this case, there is no way for them to

apply relevant metalinguistic knowledge. Second, self-paced reading tasks generally use Latin square designs, and in such a design participants encounter only very few similar cases (because they get only one condition of each experimental sentence). This makes it even harder for participants to predict what is being tested. Third, participants in an AJSPR task have to hold in memory each word they read; and in performing such a cognitively demanding task, there is little motivation for participants to consult conscious knowledge at each word region. In addition, self-paced reading tasks typically involve a large number of sentences, and the time pressure to finish reading all the items may itself deter consulting metalinguistic knowledge at every point of a sentence. (In this AJSPR study as well as in Wen, 2007, participants each read 80 sentences: 16 experimental items and 64 fillers.) Finally, different kinds of fillers in online reading tasks can also serve to discourage participants from using metalinguistic knowledge because they may not even have explicit knowledge of some phenomena targeted in the fillers.

In addition to the metalinguistic issue, researchers may worry that AJSPR tasks are not as natural or ecologically valid as CFSPR tasks. Admittedly, AJSPR does not represent typical real-life reading practice. People in more natural settings usually read for comprehension, not for judging well-formedness of sentences. However, people do often need to practice form-focused reading, such as proofreading, which, arguably, is to some extent similar to the type of reading involved in AJSPR tasks. It is perhaps also worth noting that most experimental studies fall short of true ecological validity. In theory, ecological validity is inherently called into question in virtually all psycholinguistic experiments in the sense that they are controlled, unnatural, and usually conducted in lab(-like) settings; and for these reasons ecological validity becomes less critical an issue for psycholinguistic experiments. For AJSPR, what really matters is whether it can achieve its goal of engaging participants in deep processing.

In summary, we contest the potential objections to adopting AJSPR (in comparison to CFSPR) to measure L2 deep processing. We maintain that AJSPR tasks can be a very useful and appropriate tool to help us determine whether L2ers are indeed *unable* to compute detailed structural information online, as claimed in L2 theories such as the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b, 2006c), or whether they just use shallow processing out of capacity limitations or simply as a

transitory performance strategy (e.g., as Interlanguage develops) to cope with processing load.

2.6 Conclusion

The study we reported in this chapter employed a structure-focused online sentence-processing task (i.e., AJSPR) to investigate whether (intermediate-to-advanced) L2ers are capable of engaging in (more) fully specified morphosyntactic processing. The results indicate that L2ers whose L1 lacks subject-verb number agreement can indeed show online sensitivity to such morphosyntactic violations in L2 English. We reason that L2 insensitivity to English subject-verb number agreement found in previous (self-paced reading) psycholinguistic studies is likely to have resulted from their use of comprehension-focused online tasks. In light of the findings from our study, we argue that to assess L2ers' deep processing, we need research paradigms, such as AJSPR, that encourage fully specified morphosyntactic processing.

The take-home message of the present study is that if L2ers in CFSPR tasks show nontarget-like and/or shallow processing for certain phenomena, we should not be hasty to conclude that they are unable to acquire the knowledge and/or unable to (come to) process these phenomena in a native-like way. Note that we are certainly not contending that CFSPR tasks are useless for studying L2 sentence processing. Our point, rather, is that holistic online comprehension tasks may obscure L2 deep processing, and we may benefit from using detail-focused tasks when our primary objective is to evaluate L2ers' deep processing. We need to try detail-focused tasks such as AJSPR, because such tasks have the potential not only to clarify whether L2ers can perform deep-level processing but also to reveal whether there is hope for L2ers to develop native-like processing abilities.

CHAPTER 3

GENERAL METHODS

This dissertation research includes 3 major studies on Mandarin Chinese (henceforth “Chinese”) sentence processing, covering native (“L1”) and nonnative (“L2”) online use of morphosyntactic, lexical-semantic, and discourse-context information. The studies are reported respectively in Chapters 4–6. In each study there were 2 main experiments: comprehension-focused self-paced reading (CFSPR) and acceptability-judgment self-paced reading (AJSPR). This chapter provides an overview of the general research methods used in these experiments.

3.1 Participants

The participants recruited for this dissertation research are all healthy adults. They received a fee for taking part in the experiments.

The participants in each experiment included both L1 speakers (“natives”) and L2 speakers (“L2ers”). The natives served as controls. They were recruited in Beijing and the surrounding communities. To avoid recruiting L1 participants whose grammar deviates from that of standard monolingual Chinese speakers, the study ensured that the L1 participants did not have a southern dialect background and that their English proficiency was minimal. This was checked in two ways: (a) The researcher or the researchers’ colleagues who helped with recruitment must personally know the participants and their language background; otherwise, (b) the researcher conducted a brief informal interview to ask the participants about their dialectal and English backgrounds.

As for the L2ers, they were recruited from four universities in Beijing (the Beijing Foreign Studies University; the Beijing Language and Culture University; Peking University; Tsinghua University) and one university in Guangzhou (the Guangdong University of Foreign Studies). These universities have well-established programs for teaching Chinese as an L2; they have advanced-level L2ers of Chinese from the United States, Japan, and other countries. The L2 participants’ language backgrounds were checked through a language background questionnaire (see Appendix E).

The L2 participants were always divided into 2 proficiency groups (Advanced and Intermediate), based on their scores from a Chinese proficiency test (maximum score = 50) developed for the present dissertation research (see the next section). The cutoff for grouping the participants in each study was almost always the median score of the L1-English L2ers of Chinese in the CFSPR experiment of that study. The only exception is in the first study, in which the cutoff was slightly adjusted from the median (for the specific rationale, see Chapter 4). The resulting cutoffs were, respectively, 20 for the first study (see Chapter 4), 22 for the second study (see Chapter 5), and 21 for the third study (see Chapter 6). For the 2 online experiments of each study, the same cutoff was used; it is important to do so because this enables us to compare the results of the 2 online experiments in each study. We did not use exactly the same cutoff for all the studies, because we did not intend to make “cross-study” comparisons.

3.2 Proficiency Measure—A Quasi-C-Test in Chinese

The importance of using an independent measure to assess language proficiency is increasingly recognized in the field of L2 research.⁸ This is because the L2 population is heterogeneous, characterized by substantial variability in target language (TL) proficiency. If a sample in an L2 experiment consists of participants at various TL proficiency levels, the results from that experiment are very likely to be a miscellany of diverse behavior, which may render the results unreliable or reduce chances of obtaining real effects. Existing standardized proficiency tests in Chinese such as Hanyu Shuiping Kaoshi (known as HSK or the Chinese Proficiency Test) are often too long and unsuitable for L2 research. For this reason, a short proficiency test for L2 research in Chinese is much needed.

The proficiency test developed for this research is a fill-in-the-blank quasi-C-test (see Appendix F). It contains two short, approximately equal-length texts, in which the Chinese character in the second half of some words has been left out. As illustrated in (1), the characters in parentheses are taken out.

⁸ See such efforts in English (Brown, 1980), French (Tremblay, 2011), and Korean (Lee-Ellis, 2009).

- (1) 英语中的 情况 恰好 相反。
Yingyu-zhong-de qing(kuang) qiahao xiang(fan).
English-in-DE case exactly opposite
'The case in English is exactly the opposite.'

Participants need to complete the words so that the sentences and texts make sense, filling in each blank with just one character. (If they did not know how to write a character, they could use *pinyin* instead.) The format of the test is similar to a C-test in alphabetic languages. C-tests are considered an economic and valid way of testing general language proficiency (e.g., Eckes & Grotjahn, 2006; Jafarpur, 1999). The standard practice for C-test development (Klein-Braley, 1997) requires deletion of the second half of every second word beginning from the second word in the second sentence. This method is obviously not applicable to Chinese. Elements that can be deleted in Chinese are characters, not alphabetic strings, and it is impossible to keep the rate and the starting point of deletions rigid, since a Chinese word can be just one character. Further, if a single-character word is deleted, there will be no clue left for it to be recovered. This means that only those words with two or more characters can be selected for character deletion. For this reason, the present test is called *a quasi-C-test*.

Other things that were considered in developing the Chinese proficiency test include the following: (a) The content of the texts used in the test does not favor participants with specialized knowledge; (b) the words selected for the partial deletion include both content words and function words; (c) each blank should be recovered with one best possible target answer, whenever possible; and (d) consistent scoring is used (see Appendix G for the scoring method and answer key).

The proficiency test was piloted and revised 3 times (i.e., given 4 times). In the final version, each text in the test contains 25 blanks. Participants were given 5 minutes to complete each text. The results for the final version, based on a pilot study of 3 Chinese natives and 9 L2ers of Chinese, show that the test has achieved excellent internal consistency, $KR20 = 0.97$, $SEM = 2.61$. The Pearson correlation between the participants' scores on the 2 parts of the test was also excellent, $r = 0.99$.

3.3 Material Development

For all the experiments in Chapters 4–6, the stimuli used simplified Chinese characters. To ensure that the Chinese characters and the Chinese words used in the stimuli are common and easy (enough) for L2ers, the researcher selected the vocabulary for the materials from beginning- or intermediate-level L2-Chinese textbooks. In addition, an informal norming study was carried out, in which 3 native Chinese speakers who were teaching Chinese as an L2 (at 3 different universities in Beijing) were asked to assess whether the experimental materials were appropriate for intermediate-level learners. To make sure that the sentences contained no difficult words, the original sentences were revised no less than 3 times—each time by consulting the 3 teachers as to whether the words were likely to be familiar to intermediate learners. Experimental sentences were distributed across presentation lists in a counterbalanced (i.e., Latin square) design; each participant received one list and thus saw only one version of each experimental sentence.

Each experiment used 3–5 times more fillers than experimental sentences to conceal what was being tested. The fillers included sentences of various constructions. A large proportion of fillers were actual experimental sentences in other studies. Sometimes, the fillers included sentences that are structurally more complex than the experimental sentences and thus may be slightly more difficult for participants to comprehend. However, by using fillers of various constructions, including slightly more difficult ones, the experiments can better hide the intentions of those studies and also better mirror natural reading situations. This way, the fillers should make the reading tasks more natural (i.e., more ecologically valid) than otherwise. Following the practice in Wen et al. (2010), one type of filler ($k = 8$) in each experiment was specifically devoted to assessing whether participants paid enough attention to performing the task. Such fillers are all simple sentences (see Appendix H). In the CFSPR tasks, the special fillers are all grammatical and easy to comprehend, because they are plain, true factual statements such as (2).

- (2) Yige xingqi zonggong you qi tian.
one week altogether have seven days
‘There are altogether seven days in a week.’

In the AJSPR tasks, they were all obviously ungrammatical sentences, such as (3), and thus they were easy to judge in terms of their (un)acceptability.

- (3) * Ta yibian gancai kan-le shu.
3sg one-CL just-now read-LE book

Note that the problem with (3) is word order: The quantifier expression *yibian* ‘one-CL’ should not appear before the verb (i.e., *kan* ‘read’). The correct word order should be as in (4).

- (4) Ta gancai kan-le yibian shu.
3sg just-now read-LE one-CL book
‘He/She read the book once just now.’

All items in the CFSPR experiments were always followed with a comprehension question. The questions are usually simple verification questions. For example, the comprehension question for (2) is *Yige xingqi zonggong you qi tian, dui ma?* ‘There are altogether seven days in a week, right?’ (The correct answer is *dui* ‘right’.) The overall *yes* (*dui* ‘right’) vs. *no* (*bu dui* ‘not right’) answers to the comprehension questions in each experiment were generally kept in balance. In the AJSPR tasks, all items were followed with a prompt for initiating an acceptability judgment. The prompt always used this question: *Juzi tongshun ma?* ‘Is this sentence acceptable?’ Note that the Chinese word *tongshun* literally means ‘smooth’ or ‘fluent’. Previous studies (e.g., Lu, Bates, Li, Tzeng, Hung, Tsai et al., 2000) suggest that asking whether a sentence is *tongshun* is appropriate for the purpose of eliciting acceptability judgments. In each AJSPR experiment, the overall *tongshun* (‘well-formed, acceptable’) vs. *bu tongshun* (‘not well-formed, not acceptable’) judgments were kept in balance whenever possible.

3.4 Experimental Tasks and Procedures

This dissertation research used online reading tasks in each study: CFSPR and AJSPR. Comparing results from these two tasks (see Chapters 4, 5, & 6) can help

determine whether L2ers show different processing strategies under different task conditions (for a detailed discussion of these two tasks, see Chapter 2). In a CFSPR task, participants are required to answer a comprehension question after reading each sentence, while an AJSPR task requires participants to make an acceptability judgment instead. The comprehension questions in the CFSPR tasks are usually simple *dui* ‘right’/*bu dui* ‘not right’ (i.e., *yes/no*) verification question. To answer such a question, participants need to press one of the two designated keys for *Dui* (Yes) and *Bu dui* (No) on the computer keyboard. In the AJSPR tasks, participants make an acceptability judgment by pressing the designated *Tongshun* (‘well-formed, acceptable’) key or the designated *Bu tongshun* (‘not well-formed, not acceptable’) key on the computer keyboard.

The online reading tasks were designed and administered in a similar way: All the sentences were presented using a noncumulative moving-window self-paced reading technique (Just et al., 1982). All participants were tested individually. They were briefed about the purpose of the research in a very general and vague way like this: “The primary objective of the research is to investigate how speakers of Chinese comprehend Chinese sentences.” They were not told about the specific purposes of the research so as not to give them clues about what was tested in the experiments. At the beginning of each experiment, participants were given detailed instructions in Chinese as well as a specific example. (The exact instructions for the CFSPR and AJSPR experiments can be found in Appendix I.) In the CFSPR experiments, the example requires a *bu dui* ‘not right’ answer. The example in the AJSPR experiments is ungrammatical and requires a *bu tongshun* ‘not well-formed, not acceptable’ answer. Each online task started with 8 practice items. (Appendix J lists all the practice items used in the CFSPR and AJSPR experiments.) Out of the 8 practice items in the CFSPR experiments, 4 require a *bu dui* ‘not right’ answer, 3 require a *dui* ‘right’ answer, and 1 requires a B answer (a choice from A or B). One practice item in the CFSPR tasks is ungrammatical; the purpose of this item was to give participants an impression that the sentences they are going to read are not all perfectly well-formed sentences. In the AJSPR experiments, 4 out of the 8 practice items are grammatical and 4 are ungrammatical. The ungrammatical practice items (together with the example in the instructions) can help participants understand what the Chinese terms *tongshun* (‘well-formed, acceptable’) and *bu tongshun* (‘not well-formed, not acceptable’)

mean. Participants were allowed to ask the experimenter questions when doing the practice items, and the experimenter also gave participants some feedback on their responses to the practice items. Once the experimental session started, no further feedback would be provided.

The DMDX software (Forster & Forster, 2003) was used to present materials (in a pseudo-randomized order) and to collect response data. In doing the self-paced reading tasks, participants first saw a blank screen, and when they made the first key-press, the first segment (a phrase, word, or character) would appear. Subsequently, each key-press would trigger the appearance of a new segment and the disappearance of the preceding segment. The reading time (RT) that a participant spent on each segment was therefore recorded as the amount of time between the two key-presses (corresponding to the onset and the offset of the segment display). Note that in our self-reading tasks, each key-press causes the previous segment to be covered by spaces rather than dashes while the current segment is being shown. We think that using spaces for replacing segments is better than using dashes in that spaces can prevent participants from making predictions about the sentence length (and hence possibly sentence patterns). The Chinese characters (font: STSong; font size: 14) were displayed in black color on a grey background. Each trial was presented in a single line on the computer screen.

3.5 Data Analysis

As mentioned before, participants' attentiveness in performing an online task was assessed by 8 special fillers (simple factual statements in the CFSPR tasks and obvious ungrammatical sentences in the AJSPR tasks). Participants' mean accuracy rates for those fillers were the main consideration in participant exclusions.

Participants' performance in online tasks was assessed by looking at both the accuracy data (i.e., comprehension accuracy rates in CFSPR; acceptability-judgment accuracy and also acceptance rates in AJSPR) and RTs. Main effects and interactions (within and between groups) were analyzed by using (mixed repeated-measures) ANOVAs. Pairwise comparisons within groups were conducted by using paired-samples *t* tests. Whenever possible, ANOVAs and *t* tests were performed for both participant (F_1 and t_1) and item (F_2 and t_2) means. All statistical tests used an alpha level of .05, with

Bonferroni correction when necessary. In interpreting statistical results, marginal effects ($.05 < \alpha < .10$) are also considered, in addition to significant effects.

For all the experiments, we conducted separate statistical analyses for the RT data according to proficiency groups (i.e., Natives, Advanced L2ers, and Intermediate L2ers), regardless of whether global main effects or interactions (for all the groups together) existed. This is because what we were concerned with is the processing patterns of individual proficiency groups rather than the groups as a whole (i.e., the groups' aggregate effects).

Before statistical analyses were carried out, all the RT data were subjected to a two-step data-trimming procedure. This was done in order to eliminate outlier responses. In the first step, participants' RTs longer than a cutoff point were replaced with that cutoff point. In the first study (Chapter 4), the cutoff was 2,000 ms; the other 2 studies (Chapters 5 and 6) used 2,500 ms as the cutoff. The cutoff was smaller in the first study because it used shorter reading units for self-paced reading than the other two studies did: In the first study (see Chapter 4), most of the segments into which an experimental sentence was divided for the purpose of self-paced reading contained just one Chinese character, while the segments in the other 2 studies were mostly two-or-more-character words (see Chapters 5 and 6). The cutoffs were derived in this way: For each study, we first plotted all participants' raw RTs per condition from the shortest to the longest and located the point where all the data lines began to become very sparsely dotted, and then we rounded up that point to the closest multiple of 500 ms; this multiple thus became the cutoff point for that study. Since natives' data were not as diverse as L2ers' data, the cutoff point was in effect based on L2ers' RTs. Since there was little need to set a separate cutoff for natives' data, we applied the same cutoff to the data of both natives and L2ers in each study. In the second step of the data-trimming procedure, we first calculated each participant's mean RT and standard deviation (*SD*) and then replaced the RTs that were at least 2 *SDs* away from the mean with the corresponding cutoff.

Briefly, this chapter has outlined the general research methods followed in the experiments of this dissertation research. For specific and detailed information, please refer to the Method section of each experiment in the chapters that follow.

CHAPTER 4

PROCESSING MORPHOSYNTAX IN CHINESE—NEGATION—ASPECT INTERACTIONS⁹

This chapter examines whether adult learners of a nonnative language (“L2”) can acquire and process morphosyntax absent in the native language (“L1”). The specific morphosyntactic phenomenon at issue here involves the interactions of negation and aspect in Chinese. Employing 2 self-paced reading tasks, this study tests whether as proficiency increases, adult L2 learners (“L2ers”) of Chinese whose L1 (i.e., English) lacks negation–aspect interactions can acquire the knowledge in the target language (TL) and use that knowledge in online sentence processing. At the same time, the study attempts to assess whether adult L2ers are able to engage in fully specified morphosyntactic processing when it involves nonlocal structural relations. Through the use of the 2 processing tasks (comprehension-focused vs. structure-focused), this study also aims to ascertain whether task demands influence participants’ processing strategies.

This chapter is structured as follows: Section 4.1 introduces Chinese negators *bu* and *mei*, centering on their inherent aspectual features. Section 4.2 discusses two major Chinese aspect markers *-le* and *-guo*. Section 4.3 examines how negators *bu* and *mei* interact with aspect markers *-le* and *-guo*, and proposes a feature-based framework to account for their interactions. Section 4.4 provides some L2 theoretical background for this study. Sections 4.5 and 4.6 report on the 2 self-paced reading experiments of this study, conducted to examine L1 and L2 processing of negation–aspect interactions in Chinese. A general discussion is offered in Section 4.7. The last section (Section 4.8) presents the conclusion.

4.1 Negation in Chinese

Chinese has several negative markers or particles (which will be referred to as “negators” in this dissertation), such as *bu*, *mei*, *meiyou*, *bie*, *beng*, *fei*, *wu*, *fou*, *wei*, etc. Among them, only two are the most commonly used for sentential negation in contemporary standard Mandarin Chinese (henceforth “Chinese”). They are *bu* ‘not’ and

⁹ Note that a portion of this chapter appeared in Wen and Schwartz (2014).

mei ‘not’ (e.g., Lin, 2003; Zhuang & Liu, 2011).¹⁰ These two negators are also the most researched in previous studies on Chinese negation (Lee & Pan, 2001).

One interesting phenomenon concerning *bu* and *mei*, which has been observed in numerous studies (e.g., Ernst, 1995; M. Li, 1999; Lin, 2003; Xiao & McEnery, 2008), is that these two Chinese negators have inherent aspectual requirements, which reflect an inherent quality of the negators—they possess aspectual features. This is not difficult to see from the contrasts in (1).

- (1) a. Lisi **qu** Beijing.
 Lisi **go** Beijing
 i) ‘Lisi goes to Beijing.’
 ii) ‘Lisi will go to Beijing.’

¹⁰ Traditionally, four principal negators are identified in Chinese as being in common use (see C. N. Li & Thompson, 1981, pp. 415–417). They are *bu*, *bie*, *mei*, and *mei(you)*. Among them, *bu* is considered to be the most general form of negation. The negator *bie* (meaning “don’t”) is a merged form from *bu* and *yao* ‘want’ and is usually used in imperatives. The third negator, *mei*, is used for negating the lexical verb *you* (meaning “to have” or “to exist”). As shown in (i), taken from C. N. Li and Thompson (1981, p. 416), the verb *you* can be negated by *mei* but not by *bu*; and when negated by *mei*, *you* can be optionally deleted.

- (i) a. Wo **mei (you)** qian.
 I **not exist** money
 ‘I don’t have any money.’
 b. * Wo **bu (you)** qian.
 I **not exist** money

When *you* in (ia) is deleted, *mei* functions not only as a negator but also as a verb. This special use of *mei* will not be discussed further in this study. The fourth negator, *mei(you)*, is used for negating bounded events. As shown in (ii), *meiyou* can be shortened to *mei*.

- (ii) a. John **meiyou** qu Beijing.
 John **not** go Beijing
 ‘John didn’t go to Beijing.’
 b. John **mei** qu Beijing
 John **not** go Beijing
 ‘John didn’t go to Beijing.’

Clearly, *mei* and *meiyou* perform the same negation function. In contemporary Chinese, *mei* and *meiyou* are interchangeable almost all the time when they are used for sentential negation. In brief, the discussion here suggests that the four major negators can be reduced to just two: *bu* and *mei*.

- b. Lisi **bu qu** Beijing.
 Lisi **not go** Beijing
 i) ‘Lisi doesn’t go to Beijing.’
 ii) ‘Lisi won’t go to Beijing.’
- c. Lisi **mei qu** Beijing.
 Lisi **not go** Beijing
 i) ‘Lisi didn’t go to Beijing.’
 ii) ‘Lisi hasn’t gone to Beijing.’

Sentences (1b) and (1c) are the negative versions of the affirmative (1a), negated by *bu* and *mei*, respectively. As reflected by the English glosses, (1b) and (1c) have different foci in terms of temporal viewpoints. In (1b) the focus rests on what happens *before* the potential event of Lisi going to Beijing (i.e., negation of the event that Lisi goes or will go to Beijing). In contrast, (1c) focuses on the state *after* the final point of the event (i.e., negation of the state of Lisi being in Beijing now, a state that has to have resulted from going to Beijing).

The example sentences in (1) show that mere use of negators marks aspect in Chinese. In (1b), *bu* negates an unbounded (i.e., [–bounded]) event, because the event will not happen and thus has no endpoint. In fact, an event that can be negated by *bu* usually has a habitual, future, or volitional reading, and thus is unbounded (Ernst, 1995). In (1c), however, the event negated by *mei* is bounded (i.e., [+bounded]), because a final point must be imposed as the reference point so that the state beyond the final point can be discussed. Note that boundedness is a typical feature of perfective aspect. Generally speaking, an event becomes bounded if we view it in its entirety or as a whole, because in viewing it this way we place temporal, spatial, or conceptual limits on it (C. N. Li & Thompson, 1981). In addition to negating a bounded event, *mei* in (1c) focuses the viewpoint on the resultant state (i.e., [+resultant state]). The contrasts between (1a) and (1b) and (1c) indicate that Chinese negators have inherent aspectual features in them, unlike the English negator *not*, which has no aspectual feature. The aspectual features of *bu* and *mei* are summarized in (2):

- (2) a. *bu* [-bounded]
b. *mei* [+bounded], [+resultant state]

(based on Ernst, 1995; M. Li, 1999)

4.2 Aspect Markers *-le* and *-guo*

Chinese linguistics generally recognizes four overt aspectual markers: *-le*, *-guo*, *zai*, and *-zhe* (*zai* and *-zhe* are imperfective aspectual markers). This study focuses on *-le*¹¹ and *-guo*. Traditionally, *-le* is analyzed as a perfective aspect marker and *-guo* as an experiential aspect marker. According to C. N. Li and Thompson (1981, p. 185), *-le* signifies “that an event is being viewed in its entirety or as a whole” and thus is perfective, while *-guo* emphasizes that that an event has taken place or has been experienced with regard to a certain reference time. The perfective *-le* marks bounded events. For example, (3) describes a bounded event.

- (3) Wo mai-**le** na ben shu.
I buy-**LE** that CL book
i) ‘I bought that book.’
ii) ‘I have bought that book.’

In (3), buying that book has already taken place prior to the time of speech; in other words, (3) indicates that a temporal boundary has been placed on the event.

Although the perfective marker *-le* is used for marking bounded events, it does not necessarily mean past tense. This point becomes obvious when we look at (4).

- (4) Ta mai-**le** na ben shu jiu zou.
3sg buy-**LE** that CL book then leave
‘He/She will leave after buying that book.’

¹¹ Note that the perfective aspect marker *-le* may occur at the end of a sentence in cases such as when it follows an intransitive verb or an adjective. However, the perfective *-le* should not be confused with the sentence-final particle *le*, which functions to signal the relevance of a state of affairs to a particular time frame. The perfective *-le* is a bound morpheme and cannot be separated from the verb or adjective to which it is suffixed, but the sentence-final particle *le* need not be attached to any verb or adjective. For detailed discussion of the uses and interpretations of the sentence-final *le* in different contexts, see C. N. Li and Thompson (1981, pp. 240–300).

There is no past tense involved in (4), although *-le* is used in it. The perfective *-le* indicates that the event of buying that book goes before the event of leaving, and therefore the use of *-le* functions to demarcate a borderline for the first event. As C. N. Li and Thompson (1981, p. 215) pointed out, the reason why *-le* often seems to indicate past time is that “many perfective events reported in speech are events that occurred prior to the time of speaking.”

Unlike the perfective *-le*, which presents a bounded event and typically suggests that event has taken place, the experiential aspect marker *-guo* focuses on conveying the meaning that “an event has been experienced at least once” with reference to a time point (C. N. Li & Thompson, 1981, p. 227). The difference between *-le* and *-guo* is illustrated in (5).

- (5) a. Lisi qu-**le** Beijing.
 Lisi go-**LE** Beijing
 i) ‘Lisi went to Beijing.’
 ii) ‘Lisi has gone to Beijing.’
- b. Lisi qu-**guo** Beijing.
 Lisi go-**GUO** Beijing
 ‘Lisi has been to Beijing.’

The perfective *-le* in (5a) indicates that the whole event of going to Beijing occurred or has occurred. Therefore, *-le* marks bounded events. In contrast, *-guo* makes it clear not only that the event has occurred but also that it has occurred at least once prior to the time of speech. Some Chinese linguists (e.g., M. Li, 1999; Smith, 1994; Sun, 2006) argue that the experiential *-guo* focuses on a change of state subsequent to the final point of a situation in addition to presenting a situation as being closed, whereas *-le* only signals that the final state of a situation has been obtained, regardless of whether that final state lasts or not. For example, *-le* in (5a) signals that the event of going to Beijing is closed, but it provides no further information beyond that and hence it is not known whether Lisi is still in Beijing now. However, *-guo* in (5b) indicates not only that the final state (i.e.,

Lisi's being in Beijing) has been obtained but also that the final state no longer holds, and therefore (5b) would be inappropriate if Lisi is still in Beijing. The examples in (5) show that the experiential *-guo* focuses on a new state that results from the prior situation, while *-le* lacks such a focus. If we use [\pm resultant state] to denote the presence and absence of this additional focus, we can indicate the differences between *-le* and *-guo* in terms of their aspectual features, as in (6).

- (6) a. *-le* [+bounded], [–resultant state]
 b. *-guo* [+bounded], [+resultant state]

In brief, *-le* has the features of [+bounded] and [–resultant state] while *-guo* has the features of [+bounded] and [+resultant state]. These two aspect markers in Chinese are language-specific. English has no aspect morphemes of the exact equivalence (C. N. Li & Thompson, 1981; M. Li, 1999).

4.3 Negation–Aspect Interactions

As mentioned in Section 4.1, *bu* and *mei* each possess inherent aspectual features. These inherent aspectual features constrain the negators' ability to co-occur with the aspect markers *-guo* and *-le* in a [negator + V + aspect marker] construction. Consider the data in (7).

- (7) a. Ta kan-**guo** zhe-ben shu.
 3sg look-**GUO** this-CL book
 ‘He/She has read this book.’
- b. Ta **mei** kan-**guo** zhe-ben shu.
 3sg **not** look-**GUO** this-CL book
 ‘He/She hasn't read this book.’
- c.* Ta **bu** kan-**guo** zhe-ben shu.
 3sg **not** look-**GUO** this-CL book

- d. Ta kan-**le** zhe-ben shu.
 3sg look-**LE** this-CL book
 i) ‘He/She read this book.’
 ii) ‘He/She has read this book.’
- e. * Ta **mei** kan-**le** zhe-ben shu.
 3sg **not** look-**LE** this-CL book
- f. * Ta **bu** kan-**le** zhe-ben shu.
 3sg **not** look-**LE** this-CL book

Sentences (7a)–(7c) are marked with the experiential aspect marker *-guo*, and sentences (7d)–(7f) are marked with the perfective aspect marker *-le*. As can be seen, *mei* can co-occur with *-guo* (7b), but not with *-le* (7e), while *bu* cannot co-occur with either *-guo* (7c) or *-le* (7f).

To account for the negation–aspect interactions in Chinese, some researchers (e.g., Huang, 1988; Zhuang & Liu, 2011), following W. S. Y. Wang (1965), treated *mei* as a variant form of *bu* before *you* (in *meiyou*). Wang proposed that *-le* and *you* (in *meiyou*) should be considered as suppletive alternants of the same perfective morpheme. When the morpheme is realized as *you* in a negative context, *bu* is changed to *mei* by a special morphological rule; in an affirmative context, however, *you* is transposed to post-verbal position and changed to *-le*. According to this proposal, *-le* can never occur in a negative context. Wang’s *-le/you* alternation analysis can explain why *bu* and *-le* are incompatible, as in (7f), and why *-le* cannot appear in sentences negated by *mei*, as in (7e). However, as Ernst (1995) and M. Li (1999) pointed out, the *-le/you* alternation approach is a stipulative solution to accounting for how negation interacts with aspect in Chinese, and there is no independent evidence to corroborate the claim that *mei* and *bu* originate from one common negator. Besides, the *-le/you* alternation entails that any sentence that can be negated by *mei* should in principle allow the perfective *-le* in its affirmative counterpart. This is obviously not true, as (8) shows.

- (8) a. Lili **mei** *juede* *hen lei*.
 Lili **not** feel very tired
 ‘Lili didn’t feel very tired.’
- b. * Lili *juede-le* *hen lei*.
 Lili feel-LE verytired
- c. Lili *juede* *hen lei*.
 Lili feel very tired
 i) ‘Lili feels very tired.’
 ii) ‘Lili felt very tired.’

In fact, (8b), which is supposed to be the affirmative version of (8a), is ungrammatical, precisely because of the presence of *-le* after the verb *juede* ‘feel’. In contrast, (8c), which does not have *-le* after *juede*, is grammatical. The predicate *juede hen lei* ‘feel very tired’ expresses a state, which is incompatible with the [+bounded] feature of the perfective *-le*.

Because of these problems with the *-le/you* alternation analysis, the present study proposes a feature-based framework (based on Ernst, 1995; M. Li, 1999; Smith, 1994) to account for the co-occurrence restrictions of *bu* and *mei* with *-le* and *-guo*.¹² The framework is illustrated in Figure 4.1.

¹² Note that negation–aspect interactions in Chinese are much more extensive and much more complex than the few cases mentioned here. For example, Chinese negators also interact with situation types expressed by the constellation of a verb and its arguments (see M. Li, 1999; Smith, 1994, for discussions of situation types in Chinese, such as states, activities, accomplishments, and achievements). Due to the scope of the present study, the discussion on negation–aspect interactions in Chinese is confined to the interactions of negators *bu* and *mei* with aspect markers *-le* and *-guo*.

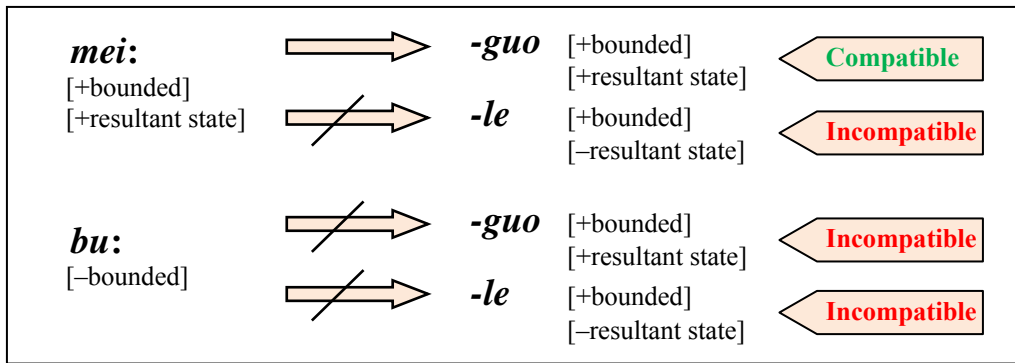


Figure 4.1. Feature-Based Framework of Negation–Aspect Interactions in Chinese.

Figure 4.1 suggests that *mei*, with the aspectual features of [+bounded] and [+resultant state], is compatible with *-guo*, which has the same features, but not with *-le*, because the [+resultant state] feature in *mei* clashes with the [–resultant state] feature in *-le*. The negator *bu* is not compatible with either *-le* or *-guo*, because the [–bounded] feature in *bu* clashes with the [+bounded] feature in *-le* and *-guo*.

One question arises from the foregoing discussions on the negation–aspect interactions: If something such as an auxiliary or adverb intervenes between a negator and V-*le/-guo* in sentences like (7), can the negation–aspect co-occurrence restrictions be altered? The answer is “yes.” There is ample evidence suggesting that the co-occurrence restrictions can be eliminated in such circumstances. For example, the modal *keneng* ‘likely’ cancels the inability of *bu* to go with *-le/-guo*, as shown in (9).

- (9) a. * Lisi **bu** qu-**le** Beijing.
Lisi **not** go-**LE** Beijing
- b. * Lisi **bu** qu-**guo** Beijing.
Lisi **not** go-**GUO** Beijing
- c. Lisi **bu keneng** qu-**le** Beijing.
Lisi **not likely** go-**LE** Beijing
i) ‘It’s not likely that Lisi went to Beijing.’
ii) ‘It’s not likely that Lisi has gone to Beijing.’

- d. Lisi **bu keneng** qu-**guo** Beijing.
Lisi **not likely** go-**GUO** Beijing
'It's not likely that Lisi has been to Beijing.'

The ungrammatical (9a) and (9b) become grammatical once *keneng* is inserted in between *bu* and *V-le/-guo*, as in (9c) and (9d). In these cases, *bu* negates the modal *keneng*, which is not bounded in nature and thus has no feature clash with *bu*.

In the general case, in summary, *mei* can co-occur with *-guo* but not *-le*, and *bu* cannot co-occur with either *-guo* or *-le*. These negation–aspect co-occurrence restrictions can be accounted for by appealing to the (in)compatibility of their aspectual features. As mentioned earlier, the English negator *not* has no aspectual feature, and hence negation–aspect interactions do not occur in English.

4.4 L2 Theoretical Background

Negation–aspect interactions in Chinese, so far, have received little or no attention in research on L2 acquisition and L2 processing (for a recent review of research on the L2 acquisition of Chinese, see Zhao, 2011). It is not known whether L1-English L2ers (or L2ers of any other language background) can acquire these negation–aspect interactions and process them in real-time sentence comprehension. The current study addresses this issue.

L2 researchers generally agree that functional morphology in the target language poses severe (if not in fact the most severe) problems, and often persistent problems, for adult L2ers (e.g., Goad & White, 2004; Hopp, 2007, 2010; Prévost & White, 2000; Slabakova, 2009). L2 representational-deficit theories (e.g., Franceschina, 2001, 2005; Hawkins & Chan, 1997; Hawkins & Liszka, 2003), in variants of the Failed Functional Features (FFF) hypothesis, state that the features of TL functional categories (or of the elements that fill those categories) are not accessible/acquirable to post-critical-period L2ers unless the features are instantiated in the L1, and hence TL functional features are not acquirable for adult L2ers whose L1 grammar lacks them. To compensate for this, L2ers have to resort to reserivicing existing L1 features or to using general cognitive strategies, which will inevitably lead to a nontarget-like grammar. In this regard, L2

representational-deficit theories view L2ers' problems with morphosyntax as being caused by deficient representations of the TL. Because negation–aspect interactions in Chinese are regulated by grammatical features not instantiated in English counterparts (e.g., *not* has no inherent aspectual feature, and thus there are no negation–aspect interactions in English), L2 acquisition theories like the FFF hypothesis predict that L1-English L2ers cannot acquire these functional features and hence the negation–aspect interactions that depend on them.

L2 processing-deficit theories like the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, 2006b, 2006c) claim that adult L2ers are unable to process complex syntax such as nonadjacent morphosyntactic relations, even when their proficiency reaches very advanced levels. L2 sentence processing research shows that although L2ers can at times perform offline tasks almost equally well as (mature) L1 speakers (“natives”), L2ers' online processing profiles often differ from those of natives' (e.g., Felser & Roberts, 2007; Jiang, 2004, 2007; Marinis et al., 2005). In addition, L2 processing difficulties are often associated with complex syntax and inflectional morphology while there is less trouble in the L2 processing of lexical semantics such as argument structures and thematic roles. Such observations led Clahsen and Felser to propose the SSH. The SSH states that the representations that L2ers build during online sentence processing lack structural details and that L2 comprehension is primarily based on lexical semantics, world knowledge, and simple heuristics. Such views of (adult) L2 processing predict that L1-English L2ers of Chinese cannot use the knowledge of Chinese negation–aspect interactions (on the assumption that the pertinent linguistic properties are acquirable) in online processing of sentences such as (7), because those sentences involve a nonlocal morphosyntactic dependency relation: The negators and the aspect markers in (7) are separated by the verb *kan* ‘read’ (e.g., ... *mei kan-guo* ...; * ... *bu kan-guo* ...).

The study to be reported in this chapter assesses whether L1-English L2ers of Chinese can use and hence acquire knowledge of negation–aspect interactions in Chinese. It includes two experiments, both of which used a noncumulative moving-window, self-paced reading paradigm (Just et al., 1982). Experiment 4.1 employed a comprehension-focused self-paced reading (CFSPR) task, while Experiment 4.2

employed an acceptability-judgment self-paced reading (AJSPR) task. As discussed in Chapter 2, comparing results from these two tasks makes it possible for us to determine whether and how different task demands influence participants' online processing strategies. From the L2 processing of subject-verb number agreement in English (see Chapter 2), we found that L2ers indeed adjust their processing strategies under different task conditions: In a holistic meaning-focused online task such as CFSPR, participants tend to overlook some structural details, while in a structure-focused online task such as AJSPR, they are more likely to engage in fully specified structural processing. In this chapter we assess whether task demand has similar influence on adult L2ers' processing strategies in reading Chinese sentences that involve negation–aspect interactions.

The experimental items used in this study are like (10).

- (10) a. Ta gang kan-**guo** zhe-ben shu.
 3sg just look-**GUO** this-CL book
 'He/She has just read this book.'
- b. Ta **mei** kan-**guo** zhe-ben shu.
 3sg **not** look-**GUO** this-CL book
 'He/She hasn't read this book.'
- c. *Ta **bu** kan-**guo** zhe-ben shu.
 3sg **not** look-**GUO** this-CL book
- d. Ta gang kan-**le** zhe-ben shu.
 3sg just look-**LE** this-CL book
 i) 'He/She just read this book.'
 ii) 'He/She has just read this book.'
- e. *Ta **mei** kan-**le** zhe-ben shu.
 3sg **not** look-**LE** this-CL book

f. * Ta **bu** kan-**le** zhe-ben shu.
 3sg **not** look-**LE** this-CL book

If L1 English-speaking L2ers can acquire and subsequently employ the requisite knowledge in real-time processing, they should spend longer time reading the regions at and/or following, first, *-guo* in ungrammatical (10c) relative to *-guo* in grammatical (10b) and to the affirmative control in (10a), and, second, *-le* in ungrammatical (10e) and (10f) relative to the affirmative control in (10d). Moreover, if L2ers exercise more fully specified processing in AJSPR than in CFSPR, these effects of ungrammaticality should show up earlier and/or more robustly in Experiment 4.2 than in Experiment 4.1.

These predictions are tested in the present study. The specific research questions are as follows:

1. Can L2ers whose L1 does not instantiate negation–aspect interactions acquire such knowledge in Chinese and use it in online Chinese sentence processing, as indicated by adult L1-English L2ers’ online sensitivity to the violations of the negation–aspect co-occurrence restrictions, relative to the grammatical controls, as in (10)? Hence, can L2ers come to engage in fully specified morphosyntactic processing, since the negation–aspect interactions involve nonlocal dependencies between negators and aspect markers in Chinese?
2. Does AJSPR engage L2ers in more fully specified processing than CFSPR does, as indicated by adult L1-English L2ers’ earlier and/or more robust online sensitivity to the ungrammatical negation–aspect interactions in Chinese in Experiment 4.2 than in Experiment 4.1?

4.5 Experiment 4.1: Comprehension-Focused Self-Paced Reading

4.5.1 Method

4.5.1.1 Participants

Participants of this study included 30 native speakers of Chinese and 54 L1-English L2ers of Chinese (see Chapter 3 for the recruiting criteria). They received a small monetary sum for taking part.

The L2 participants were divided into 2 proficiency groups (Advanced and Intermediate), based on their scores from the Chinese proficiency test developed for the present dissertation research (for a description of this quasi-C-test, see Chapter 3: General Methods).¹³ The L2ers' scores were initially used to create 2 groups of 27; this was subsequently adjusted slightly so as to have one L2 group with the same number of participants as in the Native group ($n = 30$) as well to even out the number of participants for each of the 6 presentation lists (see Section 4.5.1.2. Materials) in both L2 proficiency groups. This resulted in a cutoff point of 20: Those L2ers who scored 20 and above out of 50 were grouped as Advanced ($n = 30$); those L2ers who scored less than 20 were grouped as Intermediate ($n = 24$). Background information on the 3 groups is provided in Table 4.1.

Table 4.1. *Participants' Chinese Background Information in Experiment 4.1*

Group	Years of learning Chinese		Years of residence in China/Taiwan		Chinese proficiency scores (Max = 50)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Natives ($n = 30$)	N/A	N/A	N/A	N/A	46.4	1.8
Advanced L2ers ($n = 30$)	6.1	6.5	4.1	7.0	31.0	8.2
Intermediate L2ers ($n = 24$)	3.4	3.3	0.8	0.9	12.8	5.6

¹³ Note that the participants' mean proficiency scores reported in Wen and Schwartz (2014) were slightly different from those reported here. This is because we later made some adjustments to the scoring method for the Chinese proficiency test. In the previous method, we assigned 0.5 point in each of the following cases: (a) the answer is possible but deviates from the target answer; (b) the character is identifiable but not spelled with exact accuracy; and (c) the answer is correct but its *pinyin* is used instead of the character. In the new method (see Appendix G), we assigned 0.75 point for each close-to-correct response (i.e., failed to write the correct character but supplied the *pinyin*, misspelt character, or homophone) and 0.5 point for each plausible response (i.e., an answer that is possible but deviates a little from the target answer, including those written in *pinyin*, misspelt character, or homophone). The mean proficiency scores reported here reflect the scores based on the new scoring criteria.

The 3 groups' Chinese proficiency levels differed from each other significantly, $F(2, 81) = 219.04, p < .001$. SPSS Bonferroni corrected post hoc tests¹⁴ revealed that the Native group was significantly better than both L2 proficiency groups ($ps < .001$) and in turn the Advanced L2 group was also significantly better than the Intermediate L2 group ($p < .001$).

4.5.1.2 Materials

For the experiment, 24 experimental items like (10) were constructed, using simplified Chinese characters (see Chapter 3 for a description of how the experimental materials were developed). For the purpose of the self-paced reading task, each experimental sentence was split into 7 regions. Table 4.2 illustrates the regions as well as the experimental conditions of the sample set of stimuli (10).

Table 4.2. *Sample Set of Experimental Stimuli in Experiment 4.1*

Conditions	Regions						
	1	2	3	4	5	6	7
<i>gang-guo</i>	他	刚	看	过	这	本	书。
	Ta	gang	kan	-guo	zhe	ben	shu.
	3sg	just	look	-GUO	this	CL	book
<i>mei-guo</i>	他	没	看	过	这	本	书。
	Ta	mei	kan	-guo	zhe	ben	shu.
	3sg	not	look	-GUO	this	CL	book
<i>* bu-guo</i>	他	不	看	过	这	本	书。
	Ta	bu	kan	-guo	zhe	ben	shu.
	3sg	not	look	-GUO	this	CL	book
<i>gang-le</i>	他	刚	看	了	这	本	书。
	Ta	gang	kan	-le	zhe	ben	shu.
	3sg	just	look	-LE	this	CL	book
<i>* mei-le</i>	他	没	看	了	这	本	书。
	Ta	mei	kan	-le	zhe	ben	shu.
	3sg	not	look	-LE	this	CL	book
<i>* bu-le</i>	他	不	看	了	这	本	书。
	Ta	bu	kan	-le	zhe	ben	shu.
	3sg	not	look	-LE	this	CL	book

¹⁴ The p -values for the post hoc comparisons in this study were adjusted in SPSS (see Kinnear & Gray, 2004), by applying an adjustment that is mathematically equivalent to the Bonferroni method. Specifically, the adjustment is performed in this way: Rather than dividing the desired alpha level by the total number of pairwise comparisons, the SPSS program multiplies each of the observed t -test p -values by the total number of pairwise comparisons. In this way, the resulting p -values manifest Bonferroni adjustments while the experiment-wise alpha level of .05 can still be maintained for determining statistical significance. For the sake of convenience, such post hoc comparisons will be called "SPSS Bonferroni corrected post hoc tests" in this dissertation.

The regions of primary interest were 4, 5, and 6. Participants are likely to slow down in those regions if they detect an ungrammatical use of the (post-verbal) aspect marker in relation to the (pre-verbal) negator. As shown in Table 4.2, each set of experimental stimuli contained 6 conditions, representing 6 types of grammatical vs. ungrammatical sentences. The *gang-guo* and *gang-le* conditions are affirmative control conditions, and both are grammatical. The *mei-guo* condition is a grammatical negation condition, while the * *bu-guo*, * *mei-le*, * *bu-le* conditions are ungrammatical negation conditions. As such, the experiment makes use of the negation–aspect co-occurrence restrictions to manipulate (un)grammaticality. Out of the 6 conditions, 3 are grammatical, and 3 are ungrammatical.

Each test sentence was followed by a comprehension question. The questions were simple *yes/no* verification questions, like *Ta kan-guo zhe ben shu, dui ma?* ‘He/She has read this book, right?’ for the *-guo* sentences (10a)–(10c) or *Ta kan-le zhe ben shu, dui ma?* ‘He/She has read this book, right?’ for the *-le* sentences (10d)–(10f). Note that *ma* is a marker for *yes-no* interrogatives in Chinese. The questions for (affirmative) (10a) and (10d) require a *dui* ‘right’ answer; and for (negative) (10b), (10c), (10e), and (10f), the questions require a *bu dui* ‘not right’ answer. Note that the main clause of each comprehension question always appeared in the affirmative. If it had been in the negative, it would have amounted to supplying a correction of the ungrammatical negation–aspect interaction. By using comprehension questions in this form (i.e., affirmative main clause with a *dui ma* ‘right Q’ question tag), it avoids drawing participants’ attention to the ungrammatical negation–aspect interactions (e.g., [10c], [10e], and [10f]).

The 6 conditions of each experimental item were distributed evenly across 6 presentation lists in a counterbalanced (i.e., Latin square) design. Each participant randomly received one list and thus saw only one condition of each experimental item. (For a complete list of the experimental items, see Appendix K.)

The experiment used 64 fillers (52 grammatical and 12 ungrammatical) that were used to conceal what was being tested. They had different constructions but were similar in length to the experimental sentences. Importantly, these fillers included 8 grammatical uses of *bu* in other sentence types, such as (11).

- (11) Ta **bu** qu nali xuexi wai yu.
3sg **not** go there study foreign languages
'He/She does/will not go there to study foreign languages.'

The purpose of these 8 grammatical *bu* items is to prevent participants from making any association between the appearance of *bu* and ungrammaticality because the experimental sentences with *bu* were all ungrammatical. (For the 8 grammatical *bu* filler items, see Appendix L.) In addition, the fillers also included 8 items that were specifically devoted to assessing participants' attentiveness in performing the self-paced reading task (for a description of this type of filler, see Chapter 3; for these 8 filler items, see Appendix H). Comprehension accuracy from this type of filler formed the basis of a screening procedure, so as to ascertain which participants were not attentive enough in performing the reading task (resulting in the exclusion of 2 participants in the Advanced L2 group). This type of filler also served to assess participants' comprehension accuracy in this experiment. These 8 filler sentences are all grammatical and should better measure comprehension accuracy than the experimental items, which contained ungrammatical sentences; the ungrammaticalities may reduce comprehension accuracy. The other 48 fillers were used as experimental items in some other studies, and therefore they will not be reported on here. All the fillers were also each followed by a comprehension question. These questions are also simple verification questions. For example, the comprehension question for (11) is *Ta qu nali xuexi wai yu, dui ma?* 'He/She will go there to study foreign languages, right?' The answer is *bu dui* 'not right'. The overall number of *dui* 'right' and *bu dui* 'not right' answers in the experiment was kept in balance.

Altogether, each participant read 88 sentences plus 8 practice sentences. The sentences were pseudo-randomized when presented to participants.

4.5.1.3 Procedure

Participants were naive as to the purposes of the experiment. They were tested individually. Specific instructions on the experimental procedure were given before the self-paced reading experiment started. The participants were told that they would read some Chinese sentences on a computer monitor at their own pace in a word-by-word

(character-by-character) fashion and at the end of each sentence answer a *dui* ‘right’/*bu dui* ‘not right’ (i.e., *yes/no*) comprehension question by pressing the designated *Dui* (Yes) or *Bu dui* (No) key on the computer keyboard. The instructions also gave a specific example (for the exact instructions, see Appendix I). The experimental materials were presented to participants in simplified Chinese characters (font: STSong; font size: 14). Each sentence appeared in a single line on the computer screen.

The session started with 8 practice items. The DMDX software (Forster & Forster, 2003) was used to present materials and to collect response data. After the self-paced reading experiment, participants were asked to fill out a language background questionnaire first (see Appendix E) and then take the Chinese proficiency test (see Appendix F). To avoid making participants feel nervous about taking a test, the instructions for the proficiency test said that it was a fill-in-the-blank exercise.

4.5.2 Results

For data analysis we examined both comprehension accuracy rates and Reading Times (RTs). The results from these two types of data are reported separately.

4.5.2.1 Comprehension Accuracy

As mentioned earlier, the experiment used a special type of filler (simple factual statements) to check participants’ attentiveness in performing the self-paced reading task as well as to assess participants’ comprehension accuracy. As can be seen from Table 4.3, the 3 groups’ mean accuracy rates for this type of filler ($k = 8$) are very high (at least 92%). The results thus show high attentiveness and high comprehension accuracy for all the 3 groups.

Table 4.3. Mean Comprehension Accuracy Rates for the Fillers in Experiment 4.1

Group	Type		
	Special ($k = 8$)	<i>bu</i> ($k = 8$)	Other ($k = 48$)
Natives ($n = 30$)	.96 (.06)	1.00 (.02)	.91 (.05)
Advanced L2ers ($n = 30$)	.97 (.05)	.98 (.04)	.85 (.07)
Intermediate L2ers ($n = 24$)	.92 (.09)	.94 (.08)	.77 (.07)

Note. Standard deviations are in parentheses.

The 3 groups' mean accuracy rates of the grammatical *bu* fillers were also very high (at least 94%; see Table 4.3). The other fillers had relatively lower accuracy rates. Those fillers included different constructions, some of which might have caused a little comprehension difficulty for the participants.

As for the comprehension accuracy rates for the experimental sentences, they are visually shown for each group in Figure 4.2. The descriptive statistics are given in Appendix M. Overall, the mean comprehension accuracy rate of each group was at least 91%, suggesting that the participants did pay attention in performing the task and comprehended the sentences accurately 91% of the time (for each condition, accuracy was at least 86%—see below).

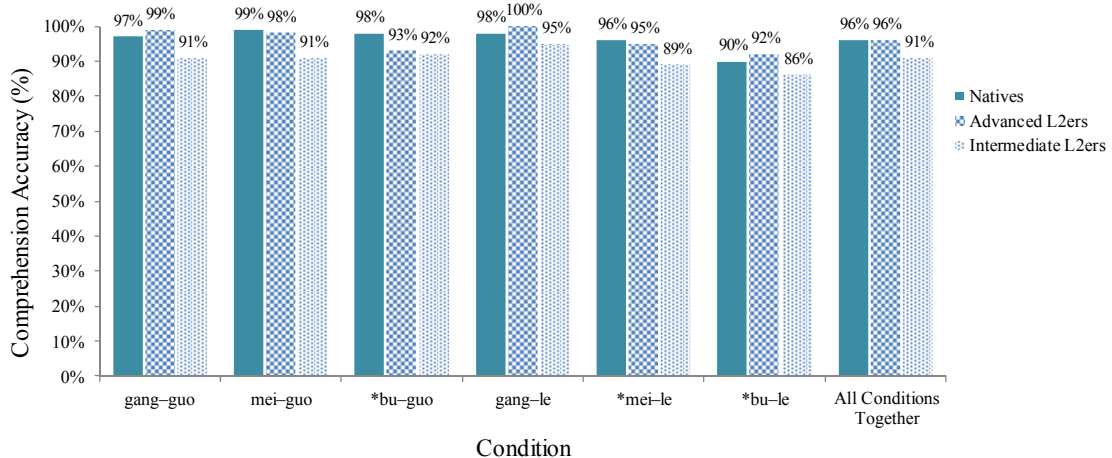


Figure 4.2. Mean Comprehension Accuracy Rates for the Experimental Sentences of Experiment 4.1.

A 3×6 (Group [Native, Advanced L2, Intermediate L2] \times Condition [*gang-guo*, *mei-guo*, **bu-guo*, *gang-le*, **mei-le*, **bu-le*]) repeated measures ANOVA carried out on the arcsine-transformed accuracy data showed a main effect of group, $F(2, 81) = 6.99$, $p = .002$, $\eta_p^2 = .15$, and a main effect of condition, $F(5, 405) = 4.44$, $p = .001$, $\eta_p^2 = .05$. The Group \times Condition interaction was not significant, $F(10, 405) = 1.10$, $p = .361$, $\eta_p^2 = .03$. These main effects mean that the participants performed differently in the different experimental conditions, and that the 3 groups differed from each other in their comprehension accuracy. From Figure 4.2 we can see that the 3 groups showed higher comprehension accuracy for grammatical *-le* sentences than for ungrammatical *-le*

sentences. This is not difficult to understand; ungrammaticality may derail a participant's comprehension process and disrupt retention of the sentence meaning. This also means that an incorrect response to the comprehension question does not necessarily represent a failure to process the sentence. This last point is important, as it relates to how to analyze the RT data. In the traditional practice of psycholinguistic research, RT data from trials on which participants answer the comprehension questions incorrectly (i.e., "incorrect trials") are often discarded. Yet, as just argued, if an incorrect trial does not necessarily represent a failure to process the linguistic elements under investigation, then dropping RT data from incorrect trials is likely to result in loss of crucial information. This is especially possible for studies employing a violation paradigm (also a standard processing task in event-related potential [ERP] studies). Admittedly, this argument only holds of the ungrammatical items here. However, even for grammatical items, an incorrect trial does not necessarily indicate a failure in processing. Other factors could lead to incorrect trials as well. For example, under online reading pressure without the possibility of going back, participants can have difficulty retaining the meaning of a sentence they read, and hence may answer the comprehension question incorrectly. Further, even if participants retain the meaning very well, answering a comprehension question may involve additional processes, such as weighing the propositional content against pragmatic appropriateness (also see Chapters 5 and 6). Therefore, failing to answer a comprehension question accurately cannot be equated with an inability to process the sentence. For the above reasons, the data analysis of this study did not exclude participants' RTs for incorrect question trials.

4.5.2.2 Reading Times

For the raw RT data, a two-step data-trimming procedure was performed. First, participants' RTs of a region (i.e., word or character) longer than 2,000 ms—the cutoff point—were replaced with that cutoff point. Second, each participant's RTs that were 2 standard deviations or more away from the participant's own mean RT were replaced with the corresponding cutoff point. Overall, data trimming affected 4.8% of the Natives' data, 7.8% of the Advanced L2ers' data, and 9.9% of the Intermediate L2ers' data. Most adjustments pertained only to the RTs of the last region. The resulting RT data for each

region were used for further analyses. The trimmed raw RT means are provided in Appendix N.

Figures 4.3–4.8 show the 3 groups' mean RTs by region for the *-guo* sentences and the *-le* sentences.¹⁵ Overall, the Natives and Advanced L2ers, but not the Intermediate L2ers, clearly slowed down at the post-critical *zhe* and *ben* regions (i.e., Regions 5 and 6) when reading the ungrammatical sentences. These RT patterns indicate that the Natives and Advanced L2ers were sensitive to the illegal negation–aspect interactions involved in the sentences.

To test the statistical significance of these patterns, repeated-measures ANOVAs were conducted separately for the *-guo* sentences and the *-le* sentences. The analyses focused on the RTs at Regions 5 and 6. Post hoc pairwise comparisons used paired-samples *t* tests (two-tailed). The ANOVAs and *t* tests were performed on both participant (F_1 and t_1) and item (F_2 and t_2) means. Following standard self-paced reading convention, the significance level was set at $\alpha = 0.05$ for all statistical tests.

Figure 4.3 depicts the Native group's mean RTs by region for the 3 types of *-guo* sentences.

¹⁵ We can see that the 3 groups differed in their reading speed: The Advanced L2 group was generally slower than the Native group, and the Intermediate group was generally slower than the Advanced L2 group. These differences are not surprising; they indicate that these participant groups are different in their Chinese proficiency levels.

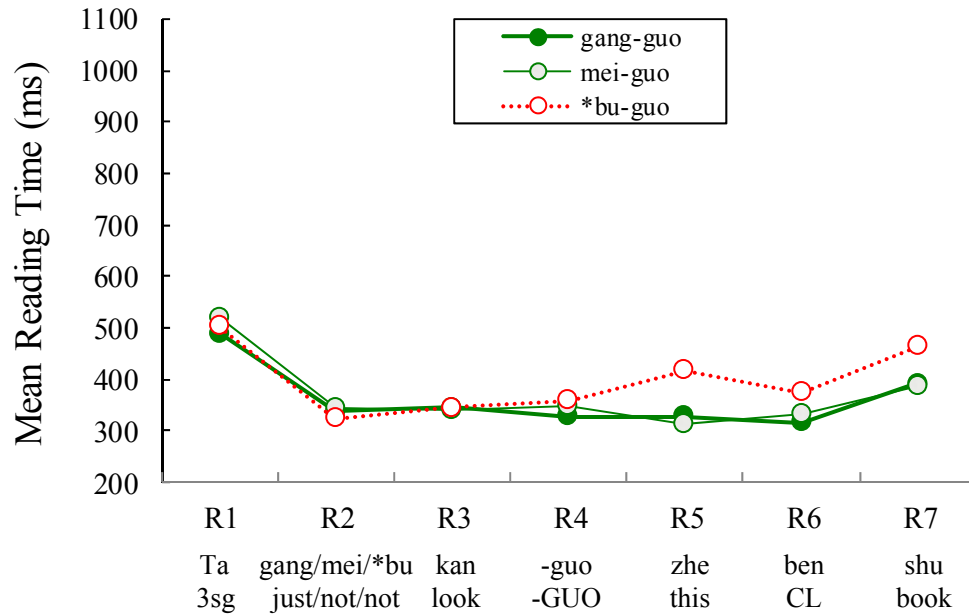


Figure 4.3. Natives' Mean RT (msec) Profile for the *-guo* Sentences of Experiment 4.1.

At Region 5 there was a main effect of condition, $F_1(2, 58) = 16.71, p < .001, \eta_p^2 = .37$; $F_2(2, 46) = 19, p < .001, \eta_p^2 = .45$. The effect spilled over to Region 6, $F_1(2, 58) = 6.37, p = .003, \eta_p^2 = .18$; $F_2(2, 46) = 8.04, p = .001, \eta_p^2 = .26$. Subsequent pairwise comparisons indicated significant differences at Region 5 between the **bu-guo* and *gang-guo* conditions, $t_1(29) = 4.93, p < .001$; $t_2(23) = 4.66, p < .001$, and between the **bu-guo* and *mei-guo* conditions, $t_1(29) = 4.49, p < .001$; $t_2(23) = 5.45, p < .001$, and at Region 6 between the **bu-guo* and *gang-guo* conditions, $t_1(29) = 3.24, p = .003$; $t_2(23) = 3.41, p = .002$, and between the **bu-guo* and *mei-guo* conditions, $t_1(29) = 2.15, p = .04$; $t_2(23) = 2.70, p = .013$. The Natives were significantly slower reading the **bu-guo* type of sentences than the other two types of sentences at these two regions.

The Natives' mean RT patterns for the 3 types of *-le* sentences are presented in Figure 4.4.

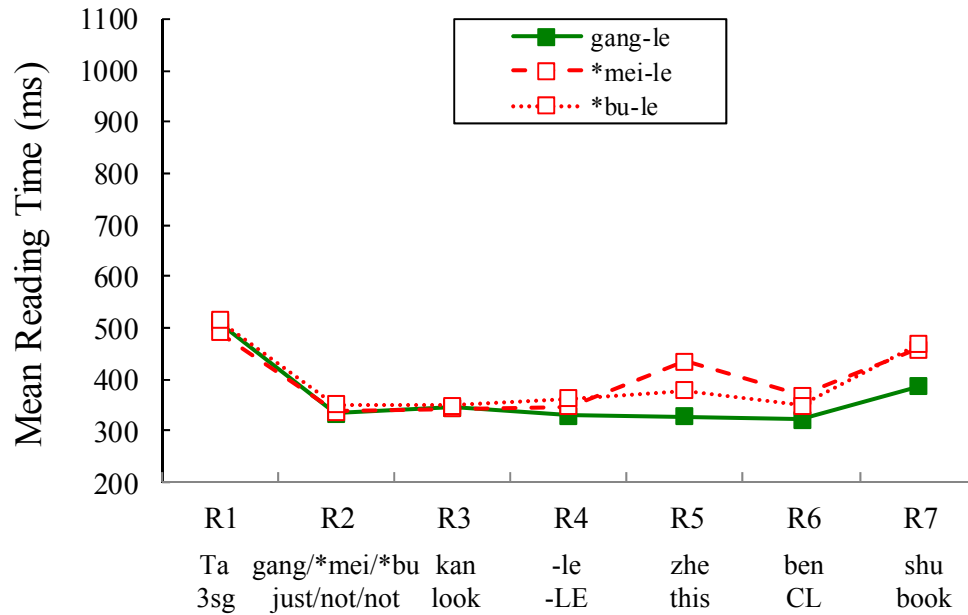


Figure 4.4. Natives' Mean RT (msec) Profile for the *-le* Sentences of Experiment 4.1.

Again, a main effect of condition was found at Region 5, $F_1(2, 58) = 17.95, p < .001, \eta_p^2 = .38$; $F_2(2, 46) = 16.43, p < .001, \eta_p^2 = .42$, as well as at Region 6, $F_1(2, 58) = 6.25, p = .003, \eta_p^2 = .18$; $F_2(2, 46) = 6.1, p = .004, \eta_p^2 = .21$. Subsequent pairwise comparisons revealed significant differences at Region 5 between the **mei-le* and *gang-le* conditions, $t_1(29) = 5.9, p < .001$; $t_2(23) = 5.07, p < .001$, and between the **bu-le* and *gang-le* conditions, $t_1(29) = 3.25, p = .003$; $t_2(23) = 4.15, p < .001$, and at Region 6 between the **mei-le* and *gang-le* conditions, $t_1(29) = 3.35, p = .002$; $t_2(23) = 3.2, p = .004$, and between the **bu-le* and *gang-le* conditions, $t_1(29) = 2.35, p = .026$; $t_2(23) = 2.6, p = .016$. The Natives spent significantly longer reading the **bu-le* and **mei-le* sentences than the affirmative control sentences at these two regions.

In sum, the Natives were sensitive to the ungrammatical negation–aspect interactions in both the *-guo* sentences and the *-le* sentences.

The Advanced L2ers' mean RT profiles for the *-guo* sentences and the *-le* sentences are presented in, respectively, Figures 4.5 and 4.6.

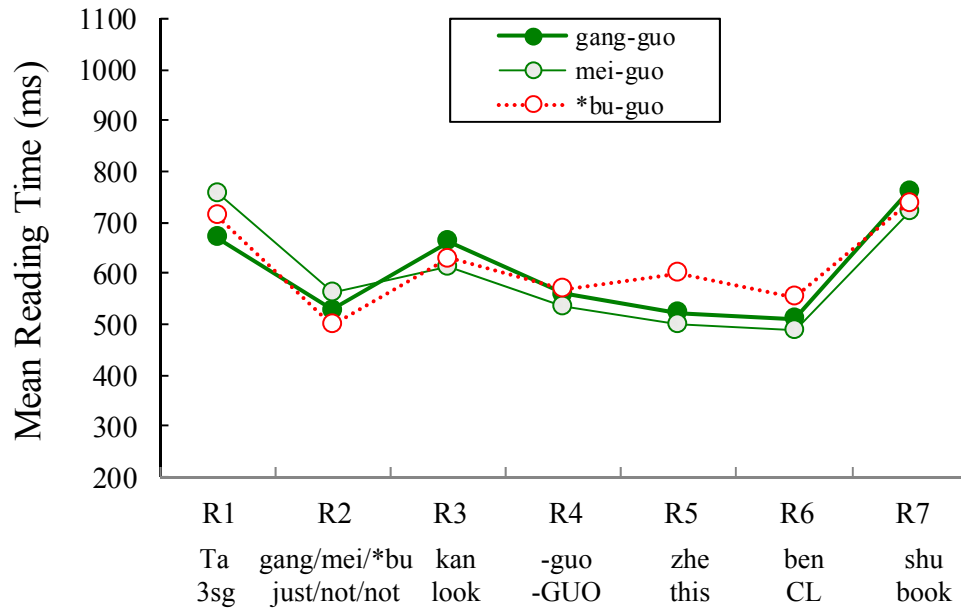


Figure 4.5. Advanced L2ers' Mean RT (msec) Profile for the *-guo* Sentences of Experiment 4.1.

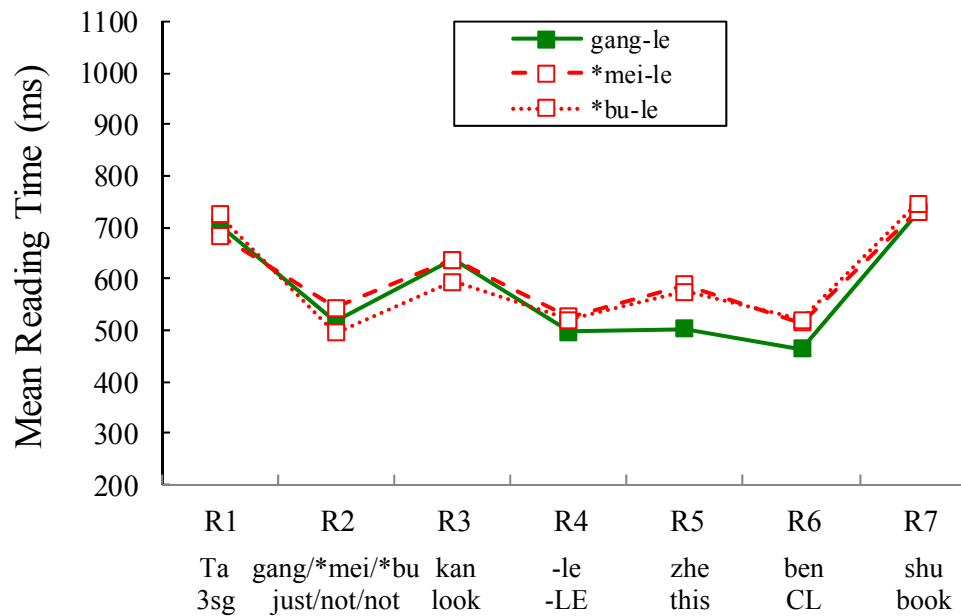


Figure 4.6. Advanced L2ers' Mean RT (msec) Profile for the *-le* Sentences of Experiment 4.1.

Similar to the Natives, the Advanced L2ers showed a main effect of condition for the *-guo* sentences at Region 5, $F_1(2, 58) = 4.77, p = .012, \eta_p^2 = .14$; $F_2(2, 46) = 7.09$,

$p = .002$, $\eta_p^2 = .24$, and at Region 6 in the participant analysis, $F_1(2, 58) = 3.34$, $p = .042$, $\eta_p^2 = .10$, but (unlike Natives) not in the item analysis, $F_2(2, 46) = 1.92$, $p = .158$, $\eta_p^2 = .08$. Subsequent pairwise comparisons revealed significant (or marginally significant) differences at Region 5 between the * *bu-guo* and *gang-guo* conditions, $t_1(29) = 1.83$, $p = .077$; $t_2(23) = 2.2$, $p = .038$, and between the * *bu-guo* and *mei-guo* conditions, $t_1(29) = 2.63$, $p = .013$; $t_2(23) = 3.79$, $p = .001$; and at Region 6, the difference between the * *bu-guo* and *gang-guo* conditions, although numerically quite large, did not reach statistical significance, $t_1(29) = 1.48$, $p = .149$; $t_2(23) = 1.18$, $p = .25$, but the difference between the * *bu-guo* and *mei-guo* conditions did reach significance in the participant analysis, $t_1(29) = 2.83$, $p = .008$, and marginal significance in the item analysis, $t_2(23) = 1.95$, $p = .064$.

For the *-le* sentences, the Advanced L2ers also showed a main effect of condition at Region 5, $F_1(2, 58) = 4.83$, $p = .011$, $\eta_p^2 = .14$; $F_2(2, 46) = 5.05$, $p = .01$, $\eta_p^2 = .18$, and at Region 6, it was significant in the participant analysis, $F_1(2, 58) = 3.31$, $p = .044$, $\eta_p^2 = .10$, but marginal in the item analysis, $F_2(2, 46) = 2.48$, $p = .095$, $\eta_p^2 = .10$. Subsequent pairwise comparisons showed significant differences at Region 5 between the * *mei-le* and *gang-le* conditions, $t_1(29) = 2.39$, $p = .024$; $t_2(23) = 3.71$, $p = .001$, and between the * *bu-le* and *gang-le* conditions, $t_1(29) = 2.63$, $p = .014$; $t_2(23) = 2.15$, $p = .042$; and at Region 6 a marginally significant difference was found between the * *mei-le* and *gang-le* conditions, $t_1(29) = 1.97$, $p = .058$; $t_2(23) = 1.84$, $p = .079$, and a significant difference was found between the * *bu-le* and *gang-le* conditions, $t_1(29) = 2.42$, $p = .022$; $t_2(23) = 2.07$, $p = .05$.

In brief, the Advanced L2 group, like the Native group, spent significantly longer, at Regions 5 and 6, reading the ungrammatical *-guo* and *-le* sentences than the corresponding grammatical sentences, which indicates that the Advanced L2ers were sensitive to the ungrammatical negation–aspect interactions.

The Intermediate L2ers' mean RT profiles for the *-guo* sentences and the *-le* sentences are shown in Figures 4.7 and 4.8, respectively.

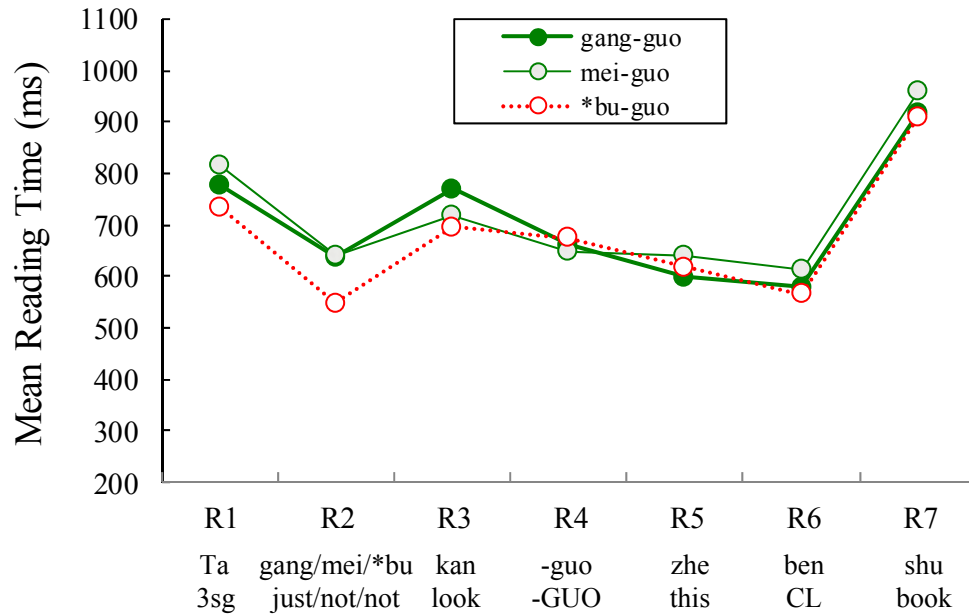


Figure 4.7. Intermediate L2ers' Mean RT (msec) Profile for the *-guo* Sentences of Experiment 4.1.

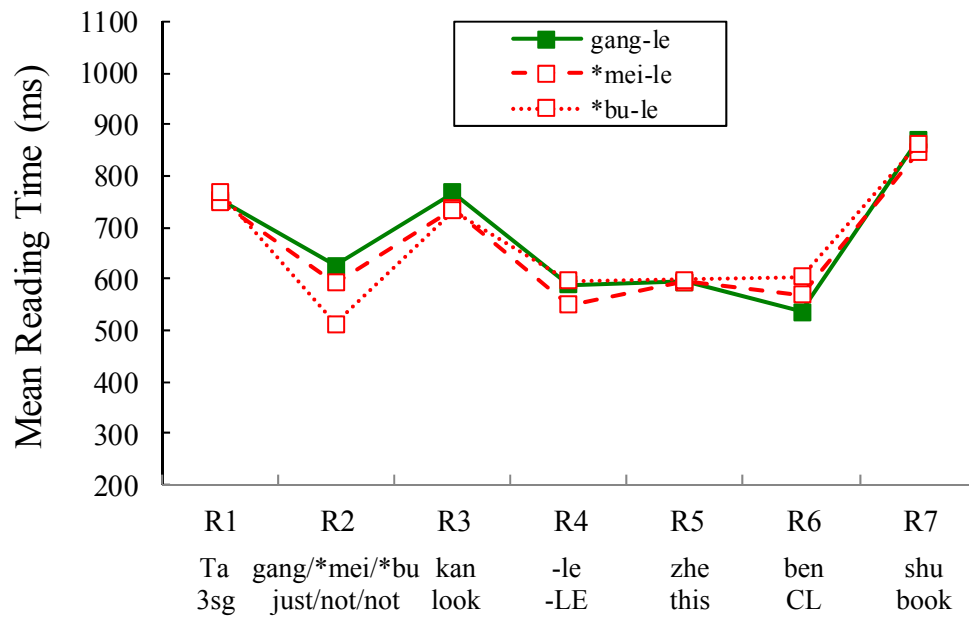


Figure 4.8. Intermediate L2ers' Mean RT (msec) Profile for the *-le* Sentences of Experiment 4.1.

No significant main effect of condition was found for the *-guo* sentences (at Region 5, $F_s < 0.85$, $p_s > .43$; at Region 6, $F_s < 0.65$, $p_s > .52$) or the *-le* sentences (at

Region 5, $F_s < 0.29$, $ps > .75$; at Region 6, $F_s < 1.02$, $ps > .37$). The Intermediate L2ers showed no reliable sensitivity to the ungrammaticalities in the experimental sentences.

Notice that the Intermediate L2ers showed some RT differences between the conditions at Region 2 (i.e., the *gang/mei/bu* region) for both the *-guo* sentences, $F_1(2, 46) = 8.27$, $p = .001$, $\eta_p^2 = .27$; $F_2(2, 46) = 5.73$, $p = .006$, $\eta_p^2 = .20$, and the *-le* sentences, $F_1(2, 46) = 11.68$, $p < .001$, $\eta_p^2 = .34$; $F_2(2, 46) = 12.31$, $p < .001$, $\eta_p^2 = .35$. Subsequent pairwise comparisons revealed significant differences, for the *-guo* sentences, between the **bu-guo* and *gang-guo* conditions, $t_1(23) = 3.80$, $p = .001$; $t_2(23) = 2.99$, $p = .007$, and between the **bu-guo* and *mei-guo* conditions, $t_1(23) = 3.43$, $p = .002$; $t_2(23) = 3.73$, $p = .001$; and for the *-le* sentences, between the **bu-le* and *gang-le* conditions, $t_1(23) = 4.04$, $p = .001$; $t_2(23) = 4.49$, $p = .001$, and between the **bu-le* and **mei-le* conditions, $t_1(23) = 4.16$, $p < .001$; $t_2(23) = 3.17$, $p = .004$. These results indicate that Intermediate L2ers read *bu* significantly faster than *gang* and *mei* for both types of sentences. This could be due to the differences of these 3 words in number of strokes (*gang*: 6; *mei*: 7; *bu*: 4) and in word frequency (*gang*: 8,617; *mei*: 61,435; *bu*: 408,160).¹⁶ Since Region 2 is not adjacent to the critical region (Region 4), it is not likely to exert any significant influence on the RTs at the critical and post-critical regions.

4.5.3 Discussion

Experiment 4.1 employed a CFSPR task to examine L1 and L2 processing of negation–aspect interactions in Chinese. The results show that the Advanced L2ers patterned like the Natives in their processing profiles: For both the *-guo* sentences and the *-le* sentences, they spent significantly longer time reading the post-critical regions (Regions 5 and 6) in the ungrammatical conditions than in the grammatical conditions, demonstrating reliable sensitivity to violations of the negation–aspect co-occurrence restrictions in the Chinese sentences. Unlike the Advanced L2ers, the Intermediate L2ers did not show sensitivity to the ungrammatical negation–aspect interactions. The results of Experiment 4.1 suggest that adult L2ers whose L1 lacks negation–aspect interactions are

¹⁶ The frequency counts are based on the 50,000,000-word Routledge Chinese Frequency Dictionary Corpus (Xiao, Rayson, & McEnery, 2009).

able to acquire such knowledge and use it in real-time sentence processing as their proficiency rises.

Since the experimental sentences used in this study involve nonlocal syntactic dependencies between negators and aspect markers, the results of this experiment also show that (advanced) adult L1-English L2ers were sensitive to nonlocal syntactic relations. Although CFSPR requires participants to focus on comprehending meaning and thus may discourage fully specified, incremental processing, the Advanced L2ers nevertheless showed online sensitivity to the negation–aspect interactions in Chinese. This indicates L2ers are able to engage in fully specified morphosyntactic processing. This finding speaks against the SSH (Clahsen & Felser, 2006a, 2006b, 2006c), which states that L2ers are unable to process nonlocal structural relations.

Notice that in this experiment the Natives and Advanced L2ers showed effects of negation–aspect interactions at the post-critical regions (Regions 5 and 6) but not at the critical region (Region 4). We suspected that this might have to do with the nature of the CFSPR task. In Experiment 4.2, we tested whether in an AJSPR task participants would more strongly show fully specified, incremental processing and thus exhibit earlier (at Region 4) and more robust effects of negation–aspect interactions.

4.6 Experiment 4.2: Acceptability-Judgment Self-Paced Reading

4.6.1 Method

4.6.1.1 Participants

This experiment included 30 native speakers of Chinese and 53 adult L1-English L2ers of Chinese (see Chapter 3 for the recruiting criteria). None of the participants had taken part in Experiment 4.1. They received a small payment for participating in the experiment. As in Experiment 4.1, the L2 participants were split into 2 proficiency groups (Advanced and Intermediate) according to their Chinese proficiency scores on the quasi-C-test (see Chapter 3 for a description of this test). The same cutoff point in Experiment 4.1 (i.e., 20) was applied: Those L2ers who scored 20 and above were grouped as Advanced ($n = 28$), and those L2ers who scored less than 20 were grouped as Intermediate ($n = 25$). Participants' background information is given in Table 4.4.

Table 4.4. *Participants' Chinese Background Information in Experiment 4.2*

Group	Years of learning Chinese		Years of residence in China/Taiwan		Chinese proficiency scores (Max = 50)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Natives (<i>n</i> = 30)	N/A	N/A	N/A	N/A	46.1	1.6
Advanced L2ers (<i>n</i> = 28)	4.0	3.3	1.7	2.9	25.9	5.7
Intermediate L2ers (<i>n</i> = 25)	2.8	2.2	1.0	1.6	13.9	5.0

The 3 groups' Chinese proficiency levels were significantly different from each other, $F(2, 80) = 374.14, p < .001$. SPSS Bonferroni corrected post hoc tests indicate that the Native group was significantly higher in Chinese proficiency than both the L2 groups ($ps < .001$), and the Advanced L2 group was significantly higher than the Intermediate L2 group ($p < .001$).

Note that the Advanced L2 group in Experiment 4.1 scored a little higher in the Chinese proficiency test ($M = 31.0, SD = 8.2$) than the Advanced L2 group in this experiment ($M_{\text{difference}} = 5.0, t(52) = 2.71, p = .009$). The Chinese proficiency scores of the Intermediate L2 group in Experiment 4.1 ($M = 12.8, SD = 5.6$), however, was comparable to those of the Intermediate L2 group in Experiment 4.2 ($M_{\text{difference}} = -1.1, t(46) = 0.74, p = .446$).

4.6.1.2 Materials

The same 24 experimental items from Experiment 4.1 were used in Experiment 4.2, but participants in this experiment, after reading each sentence, were asked to make an acceptability judgment instead of answering a comprehension question. As in Experiment 4.1, the 6 experimental conditions (3 grammatical and 3 ungrammatical) of each item were evenly distributed across 6 presentation lists in a Latin square design. In addition to the experimental items, this experiment also used 64 fillers of different constructions (44 grammatical and 20 ungrammatical), thus yielding a total of 88 items in each list. The fillers included the 8 grammatical *bu* fillers of Experiment 4.1 (see above). In addition, the experiment also used the 8 special fillers designed for assessing participants' attentiveness in performing the self-paced reading task (for a description of the fillers, see Chapter 3; for the 8 special fillers, see Appendix H). The other 48 fillers were the same as in Experiment 4.1.

4.6.1.3 Procedure

The procedure of Experiment 4.2 was the same as that of Experiment 4.1 except that participants were asked to judge the acceptability of each sentence. To prompt judgments, participants were always asked this question: *Juzi tongshun ma?* ‘Is this sentence acceptable? Participants made acceptability judgments by pressing the designated *Tongshun* (‘well-formed, acceptable’) or *Bu tongshun* (‘not well-formed, not acceptable’) key on the computer keyboard (for a description of the AJSPR task, see Chapter 3). To help participants understand what to do, the instructions provided one ungrammatical example (for the instructions, see Appendix I). Also, participants were given 8 practice items at the beginning of the experiment (for the practice items, see Appendix J). The practice items contained 4 ungrammatical sentences, to help illustrate what *bu tongshun* really means. After the self-paced reading task, participants completed a language background questionnaire (see Appendix E) and the Chinese proficiency test (see Appendix F).

4.6.2 Results

As in Experiment 4.1, both participants’ accuracy data and online RTs were analyzed.

4.6.2.1 Acceptability-Judgment Accuracy

Table 4.5 gives the descriptive statistics for the acceptability-judgment accuracy data of the fillers in Experiment 4.2.

Table 4.5. *Mean Acceptability-Judgment Accuracy Rates for the Fillers in Experiment 4.2*

Group	Type			
	Special (ungramm; $k = 8$)	<i>bu</i> (gramm; $k = 8$)	Other (gramm; $k = 36$)	Other (ungramm; $k = 12$)
Natives ($n = 30$)	.93 (.12)	.96 (.08)	.51 (.13)	.73 (.24)
Advanced L2ers ($n = 28$)	.91 (.11)	.93 (.11)	.54 (.18)	.52 (.20)
Intermediate L2ers ($n = 25$)	.85 (.13)	.87 (.11)	.54 (.13)	.50 (.24)

Note. Standard deviations are in parentheses. ungramm = ungrammatical; gramm = grammatical.

The 3 participant groups’ accuracy for the special type of filler was very high ($M \geq 85\%$), indicating that all the participants were attentive in performing the reading

task. This is also corroborated by their accuracy for the grammatical *bu* fillers ($M \geq 87\%$). The accuracy rates of the other fillers were much lower, but similar to those for the experimental sentences in some conditions (e.g., the * *mei-le* condition for the L2 groups; see Figure 4.9).

The mean acceptability-judgment accuracy rates for the experimental sentences are presented for each participant group in Figure 4.9. (For the descriptive statistics, see Appendix O.)

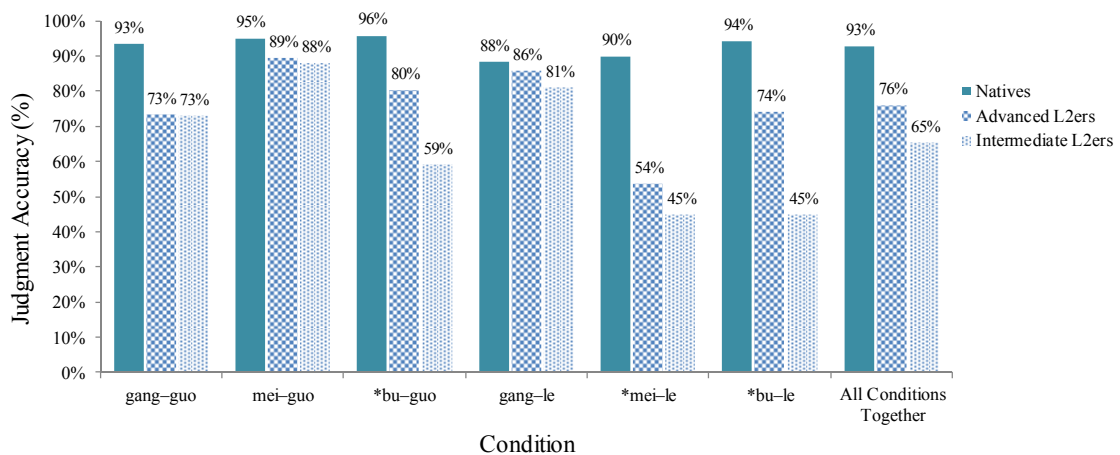


Figure 4.9. Mean Acceptability-Judgment Accuracy Rates for the Experimental Sentences of Experiment 4.2.

As can be seen from Figure 4.9, the Native group has a much higher overall accuracy rate than the 2 L2 groups. The 2 L2 groups performed similarly and fairly well for the grammatical sentences (at least 73%), but not for the ungrammatical sentences. Both L2 groups performed poorly in the * *mei-le* condition. The reason for this is unknown, and these accuracy results did not seem to match the RT results of this condition (see Figures 4.13 & 4.15). A 3×6 (Group [Native, Advanced L2, Intermediate L2] \times Condition [*gang-guo*, *mei-guo*, * *bu-guo*, *gang-le*, * *mei-le*, * *bu-le*]) repeated measures ANOVA performed on the arcsine-transformed accuracy data showed a main effect of group, $F(2, 80) = 48.67, p < .001, \eta_p^2 = .55$, a main effect of condition, $F(5, 400) = 11.90, p < .001, \eta_p^2 = .13$, and a significant Group \times Condition interaction, $F(10, 400) = 4.73, p < .001, \eta_p^2 = .11$. These effects attest to the fact that the 3 groups

performed differently in the different experimental conditions, suggesting that both proficiency and condition are at work.

4.6.2.2 Reading Times

Prior to statistical analysis, the raw RTs were trimmed by applying the procedure we used in Experiment 4.1. Data trimming affected 5.5% of the Natives' data, 11.1% of the Advanced L2ers' data, and 14.7% of the Intermediate L2ers' data. Appendix P gives all the trimmed raw RT means.

As in Experiment 4.1, the data analysis of this experiment did not exclude those trials on which the participants were incorrect in their judgments.¹⁷ In deciding whether incorrect trials should be removed from data analysis, it is crucial to establish the actual cause(s) of the incorrect trials. The principal reasons (or perhaps the only reasons) for excluding error trials should be that the errors are due to (a) participants' carelessness or inattentiveness in performing the task and (b) participants' lack of the relevant knowledge. In the present experiment, the participants demonstrated a high degree of attentiveness in performing the self-paced reading task (assessed through the special type of filler). Although the judgment accuracy of L2 participants was relatively low, they nevertheless demonstrated the desired effects in the online task—at least the Advanced L2 group does (see Figures 4.12, 4.13, and 4.15); the online effects would be difficult to explain if the participants did not have the relevant knowledge. It seems that failing to make an accurate online acceptability judgment after reading a sentence may not always suggest lack of knowledge. Other factors could also cause incorrect responses to experimental trials (also see Section 4.5.2.1). First, the pressure of self-paced reading tasks may impose an additional difficulty for participants to maintain the meaning and structure of a whole sentence they read, and as a result this may lead to an incorrect acceptability judgment (even if they process the sentence successfully). Second, when making an acceptability judgment, participants may consider things other than just the form of the sentence. As

¹⁷ Our data analysis, however, did explore whether removing incorrect trials affects the data patterns. The results indicate that, generally speaking, such data cleaning did not improve the significance levels, as can be seen from Appendix Q. In addition, we found that removal of a large proportion of incorrect trials could lead to a shift of the data pattern. This is evidenced in the item analysis of the *-guo* sentences at Region 4 for the Intermediate L2ers, which revealed an unexpected significant effect of condition in the opposite direction (for details, see Appendix Q).

will be discussed in Chapters 5 and 6, participants may consider, for instance, propositional coherence and pragmatic appropriateness. Third, incorrect trials may also be indicative of things such as a speed-accuracy tradeoff (see Chapter 6), shallow or good-enough processing (see Chapter 5), structural complexity of the experimental sentences (see Chapter 5), etc. For these reasons, we argue that removing RT data for error trials may increase the chances of distorting the true pictures of participants' online processing (for other arguments on why online data analysis should not exclude incorrect trials, see Hopp, 2007).¹⁸ Hence, our RT data analysis did not remove the error trials.

Figures 4.10–4.15 present the 3 participant groups' mean RT profiles by region. By visually inspecting these figures we can see that when reading the ungrammatical *-guo* sentences, both the Native and Advanced L2 groups slowed down at Region 4 (i.e., the critical *-guo* region); the slowdown became conspicuous at Region 5 (i.e., the post-critical *zhe* region) for the Native group, but at Region 4 for the Advanced L2 group. The Intermediate L2 group did not show any slowdown in reading the ungrammatical *-guo* sentences at Region 4 or 5. For the *-le* sentences, all 3 groups read more slowly at Regions 4 and 5 (i.e., the critical *-le* region and the post-critical *zhe* region) for the ungrammatical ** mei-le* and ** bu-le* sentences than for the grammatical *gang-le* sentences; at Region 6, the participant groups (except the Advanced L2 group) did not seem to show any interesting RT pattern.

To test the statistical significance of these RT patterns at Regions 4 and 5, we ran repeated-measures ANOVAs separately for the *-guo* sentences and the *-le* sentences. Post hoc pairwise comparisons used paired-samples *t* tests (two-tailed). As was done for Experiment 4.1, the statistical analyses were performed on both participant (F_1 and t_1) and item (F_2 and t_2) means. The significance level was again set at $\alpha = 0.05$.

Figures 4.10 and 4.11 show the Native group's mean RT profiles for the *-guo* sentences and the *-le* sentences, respectively.

¹⁸ In view of the standard psycholinguistic practice of excluding the data of error trials, the point we are trying to make may be a significant one. Since it has become a convention, people seldom question why they do so. However, without determining whether the incorrect responses are caused by inattentiveness and/or ignorance, excluding incorrect trials may result in discarding critical information. If such data removal affects a large portion of the overall data, it is likely to alter the actual data pattern(s).

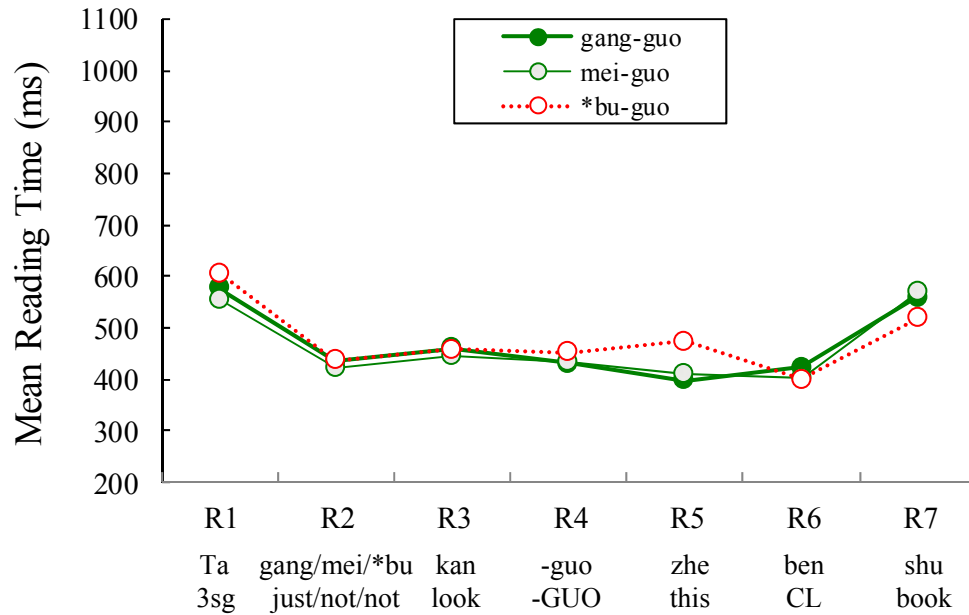


Figure 4.10. Natives' Mean RT (msec) Profile for the *-guo* Sentences of Experiment 4.2.

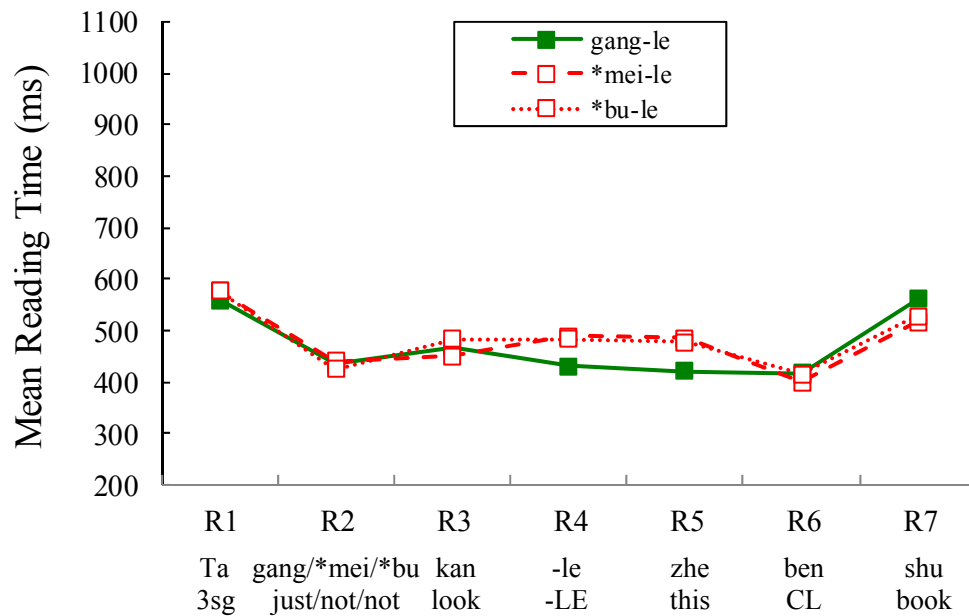


Figure 4.11. Natives' Mean RT (msec) Profile for the *-le* Sentences of Experiment 4.2.

For the *-guo* sentences, the Native group's main effect of condition was not significant at Region 4, $F_1(2, 58) = 1.01, p = .370, \eta_p^2 = .03$; $F_2(2, 46) = 1.32, p = .278, \eta_p^2 = .05$, but it reached significance at Region 5, $F_1(2, 58) = 12.34, p < .001, \eta_p^2 = .30$; $F_2(2, 46) = 13.25, p < .001, \eta_p^2 = .37$. Subsequent pairwise comparisons yielded

significant differences at Region 5 between the * *bu-guo* and *gang-guo* conditions, $t_1(29) = 4.69, p < .001$; $t_2(23) = 4.22, p < .001$, and between the * *bu-guo* and *mei-guo* conditions, $t_1(29) = 3.27, p = .003$; $t_2(23) = 3.61, p = .001$. Clearly, the Natives spent significantly longer time reading the ungrammatical *-guo* sentences than the grammatical *-guo* sentences at Region 5.

For the *-le* sentences, the Natives showed a main effect of condition at Region 4, $F_1(2, 58) = 5.32, p = .008, \eta_p^2 = .16$; $F_2(2, 46) = 5.64, p = .006, \eta_p^2 = .20$, as well as at Region 5, $F_1(2, 58) = 5.94, p = .005, \eta_p^2 = .17$; $F_2(2, 46) = 5.80, p = .006, \eta_p^2 = .20$. Subsequent pairwise comparisons yielded significant differences at Region 5 between the * *mei-le* and *gang-le* conditions, $t_1(29) = 2.64, p = .013$; $t_2(23) = 2.76, p = .011$, and between the * *bu-le* and *gang-le* conditions, $t_1(29) = 2.89, p = .007$; $t_2(23) = 3.12, p = .005$, and at Region 6 between the * *mei-le* and *gang-le* conditions, $t_1(29) = 2.80, p = .009$; $t_2(23) = 3.22, p = .004$, and between the * *bu-le* and *gang-le* conditions, $t_1(29) = 2.77, p = .010$; $t_2(23) = 2.73, p = .012$. It took significantly longer time for the Natives to read the ungrammatical *-le* sentences than the grammatical *-le* sentences at Regions 4 and 5.

At Region 6, the Native group did not show any main effect or interaction for either the *-guo* sentences ($F_s < 1.46, p_s > .24$) or the *-le* sentences ($F_s < 0.50, p_s > .61$).

The Native group's RT patterns for the *-guo* and *-le* sentences suggest that in doing the AJSPR task, the Natives were again sensitive to the violations of the negation–aspect co-occurrence restrictions.

The Advanced L2ers' mean RT profiles for the *-guo* sentences and the *-le* sentences are shown in Figures 4.12 and 4.13, respectively.

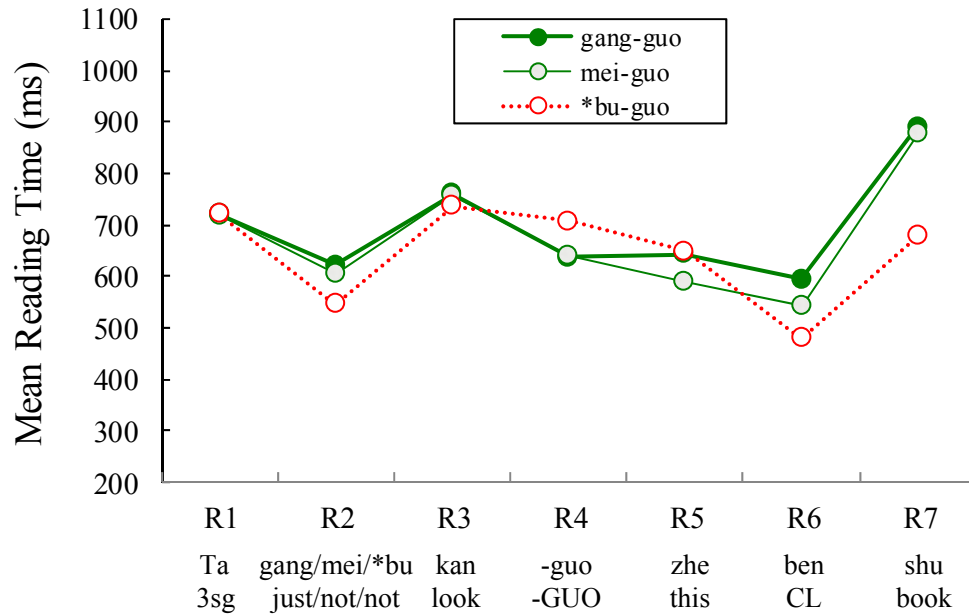


Figure 4.12. Advanced L2ers' Mean RT (msec) Profile for the *-guo* Sentences of Experiment 4.2.

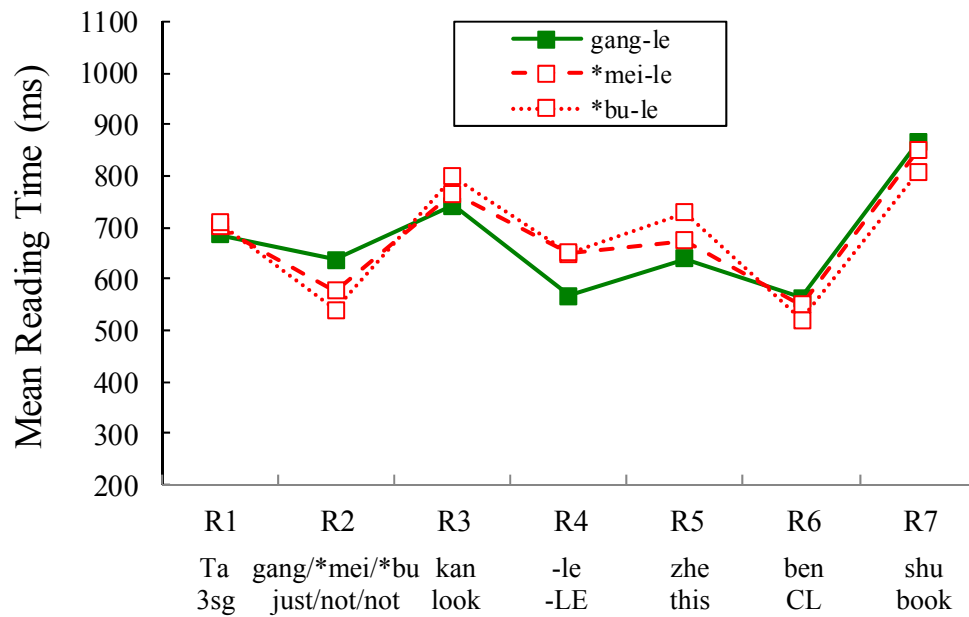


Figure 4.13. Advanced L2ers' Mean RT (msec) Profile for the *-le* Sentences of Experiment 4.2.

For the *-guo* sentences, the Advanced L2ers showed a main effect of condition at Region 4, $F_1(2, 54) = 4.29, p = .019, \eta_p^2 = .14$; $F_2(2, 46) = 3.25, p = .048, \eta_p^2 = .12$, but

not at Region 5, $F_1(2, 54) = 1.15, p = .326, \eta_p^2 = .04$; $F_2(2, 46) = 0.77, p = .468, \eta_p^2 = .03$. Subsequent pairwise comparisons for Region 4 showed that the difference between the **bu-guo* and *gang-guo* conditions was significant in the participant analysis, $t_1(27) = 2.96, p = .006$, and marginal in the item analysis, $t_2(23) = 2.02, p = .055$, and the difference between the **bu-guo* and *mei-guo* conditions was also significant in the participant analysis, $t_1(27) = 2.08, p = .047$, and marginal in the item analysis, $t_2(23) = 2.05, p = .052$. The Advanced L2ers spent reliably longer time reading the ungrammatical -*guo* sentences than the grammatical -*guo* sentences at Region 4. The Advanced L2 group also showed RT differences between the conditions at Region 6, $F_1(2, 54) = 3.64, p = .033, \eta_p^2 = .12$; $F_2(2, 46) = 6.36, p = .004, \eta_p^2 = .22$, but in the opposite direction. It appears that the Advanced L2ers, after detecting the ungrammatical negation-aspect interaction at Region 4 in the **bu-guo* condition, subsequently spent less effort searching for further unacceptable elements in the sentences of this condition; as a result, this ungrammatical condition was read faster at Region 6 and also Region 7.

Note that at Region 2, the Advanced L2ers read *bu* faster than *gang* and *mei*, $F_1(2, 54) = 5.70, p = .006, \eta_p^2 = .17$; $F_2(2, 46) = 3.29, p = .046, \eta_p^2 = .13$. As discussed in Section 4.5.2.2, these RT results probably reflect the differences of these 3 words in terms of number of strokes and/or word frequency. Since these effects are unlikely to influence the RTs at the critical and post-critical regions (Regions 4 and 5), we will not pursue this matter further.

For the -*le* sentences, the Advanced L2ers showed a main effect of condition at Region 4, $F_1(2, 54) = 4.67, p = .013, \eta_p^2 = .15$; $F_2(2, 46) = 3.99, p = .025, \eta_p^2 = .15$, and a marginal main effect of condition at Region 5, $F_1(2, 54) = 2.82, p = .069, \eta_p^2 = .09$; $F_2(2, 46) = 2.43, p = .100, \eta_p^2 = .10$. No main effect or interaction was found at Region 6 ($F_s < 1.04, p_s > .34$). Subsequent pairwise comparisons showed that at Region 4, the difference between the **mei-le* and *gang-le* conditions was significant in the participant analysis, $t_1(27) = 2.33, p = .028$, and marginal in the item analysis, $t_2(23) = 2.03, p = .054$, and the difference between the **bu-le* and *gang-le* conditions was significant in both participant and item analyses, $t_1(27) = 3.23, p = .003$; $t_2(23) = 2.86, p = .009$. At Region 5 no significant difference was found between the **mei-le* and *gang-le* conditions, $t_1(27) = 0.96, p = .334$; $t_2(23) = 0.71, p = .484$, but the difference between the **bu-le* and

gang-le conditions was significant in the participant analysis, $t_1(27) = 2.10, p = .045$, and marginal in the item analysis, $t_2(23) = 1.98, p = .060$. The Advanced L2ers were thus significantly slower in reading the ungrammatical *-le* sentences than the grammatical *-le* sentences at Region 4; the effects spilled over to Region 5, although becoming statistically less reliable.

In brief, the Advanced L2ers were very much like the Natives in showing sensitivity to the negation–aspect interactions, although such sensitivity was not always equally robust.

Figures 4.14 and 4.15 show the Intermediate L2ers’ mean RT profiles for the *-guo* sentences and the *-le* sentences, respectively.

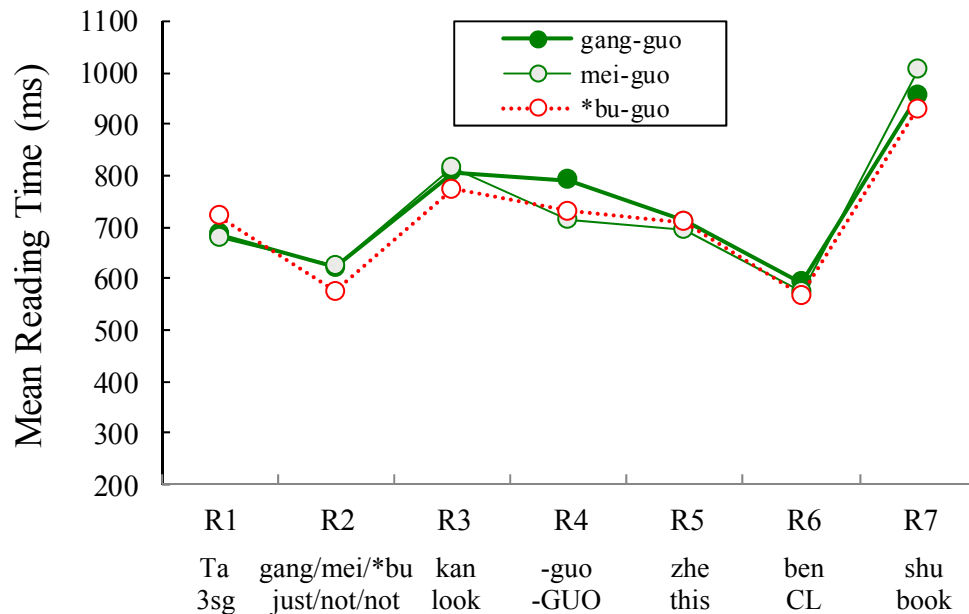


Figure 4.14. Intermediate L2ers’ Mean RT (msec) Profile for the *-guo* Sentences of Experiment 4.2.

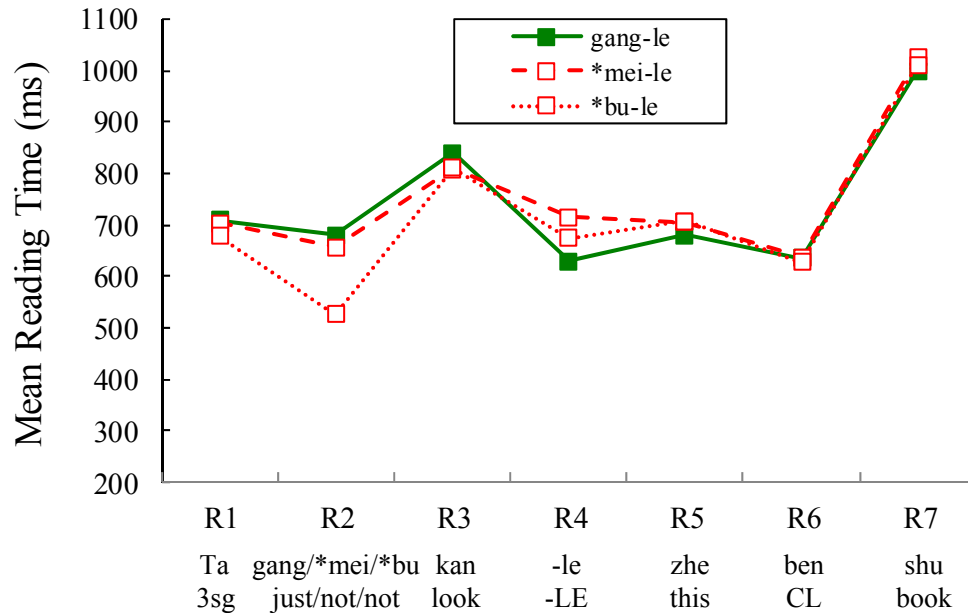


Figure 4.15. Intermediate L2ers' Mean RT (msec) Profile for the *-le* Sentences of Experiment 4.2.

For the *-guo* sentences, the Intermediate L2 group showed no significant main effect of condition at Region 4 ($F_s < 1.96, p_s > .15$), Region 5 ($F_s < 0.21, p_s > .81$), or Region 6 ($F_s < 0.41, p_s > .67$).

For the *-le* sentences, Region 4 showed a trend in the right direction for the Intermediate L2 group, and the main effect of condition came close to marginal significance in the participant analysis, $F_1(2, 48) = 2.38, p = .104, \eta_p^2 = .09$, but not in the item analysis, $F_2(2, 46) = 1.89, p = .163, \eta_p^2 = .08$. No main effect of condition was found at Region 5 ($F_s < 0.61, p_s > .54$) or Region 6 ($F_s < 0.19, p_s > .83$). Subsequent pairwise comparisons for Region 4 yielded a marginal difference between the **mei-le* and *gang-le* conditions, $t_1(24) = 1.94, p = .065$; $t_2(23) = 1.97, p = .061$, but not between the **bu-le* and *gang-le* conditions, $t_1(24) = 1.52, p = .141$; $t_2(23) = 1.12, p = .274$.

In short, the Intermediate L2ers showed just a little sensitivity to the negation–aspect interactions for the *-le* sentences, but not any for the *-guo* sentences.

4.6.3 Discussion

Experiment 4.2 employed an AJSPR task to examine L1 and L2 processing of negation–aspect interactions in Chinese. The results of this experiment are very similar to

those of Experiment 4.1 except that the effects of ungrammatical negation–aspect interactions appeared earlier in Experiment 4.2, at the critical region (for the Natives in processing *-le* sentences and for the Advanced L2ers in processing both *-guo* and *-le* sentences). In addition, the Intermediate L2ers in Experiment 4.2 showed a small effect of negation–aspect interactions in processing the *-le* sentences. These earlier and more robust L2 results in Experiment 4.2 than Experiment 4.1 suggest that L2ers do engage in more fully specified, incremental morphosyntactic processing in AJSPR than in CFSPR. The Intermediate L2ers’ marginal sensitivity to ungrammatical negation–aspect interactions (viz., **mei-le* vs. *gang-le* conditions at Region 4) may also be indicative that AJSPR can indeed reveal some L2 structural processing ability that CFSPR may not be able to reveal. The results of these two experiments together confirm that task demand influences participants’ online processing strategies and that L2 shallow processing exhibited in CFSPR may simply indicate a performance strategy rather than lack of the relevant processing ability.

In summary, the results of Experiment 4.2 further demonstrate that L2ers can acquire morphosyntactic knowledge that their L1 does not instantiate and they can exercise fully specified morphosyntactic processing even when it involves nonlocal structural relations.

4.7 General Discussion

The present study employed both CFSPR and AJSPR to examine how adult L2ers process interactions between negation and aspect in Chinese. It attempted to find out whether adult L2ers of Chinese whose L1 (i.e., English) lacks negation–aspect interactions can (come to) acquire such morphosyntactic knowledge and use this knowledge in online sentence processing. By employing experimental sentences targeting nonlocal structural relations, the study also assessed whether adult L2ers can (come to) engage in fully specified morphosyntactic processing.

The results of both experiments show that L2ers whose L1 does not instantiate negation–aspect interactions are indeed able to acquire such knowledge and use it in online sentence processing as their proficiency rises. In reading the *-guo* and *-le* sentences, which were designed to test the participants’ abilities to process negation–aspect

interactions in Chinese, the Advanced L2ers, but not yet the Intermediate L2ers, patterned like the Natives in their processing profile. They showed clear online sensitivity to violations of the negation–aspect co-occurrence restrictions at issue. The sensitivity was reflected by their longer RTs for the ungrammatical conditions than the grammatical conditions in Experiment 4.1 at the post-critical *zhe* ‘this’ and *ben* ‘CL’ regions (i.e., Regions 5 and 6) and in Experiment 4.2 at the critical *-guo/-le* region and the post-critical *zhe* region (i.e., Regions 4 and 5).

These results are incompatible with L2 representational-deficit theories such as the Failed Functional Features hypothesis (e.g., Franceschina, 2001, 2005; Hawkins & Chan, 1997; Hawkins & Liszka, 2003), which hold that L2ers’ nontarget-like behavior results from their deficient L2 representations. The FFF hypothesis claims that functional (i.e., grammatical) features that are not activated in early life become unavailable to adult acquirers. This claim is challenged by the data of the current study. Despite that fact that the English negator *not* has no aspectual feature and that no negation–aspect interactions occur in English, adult L1-English L2ers are nevertheless able to acquire such morphosyntactic knowledge and use that knowledge in parsing L2 sentences. They are not permanently impaired in this area of L2 acquisition, contrary to what the FFF hypothesis predicts.

The results of this study also demonstrate that L2ers can come to engage in fully specified morphosyntactic processing, contrary to what the SSH (Clahsen & Felser, 2006a, 2006b, 2006c) claims. According to the SSH, adult L2ers’ processing differs fundamentally from that of (mature) native speakers in that it is restricted to shallow parsing, depending not on morphosyntactic information but instead on lexical-semantic, contextual, and pragmatic information. The SSH argues that L2ers, even at high levels of L2 proficiency, cannot process complex syntactic relations such as nonlocal structural dependencies. Despite the fact that this study involved processing nonlocal structural dependencies between negators and aspect markers in Chinese, the Advanced L2ers displayed clear online sensitivity. L2 processing, as evidenced by the Advanced L2ers’ results in this study, is not necessarily shallow in the sense of the SSH.

Proponents of L2 representational-deficit or processing-deficit theories might be tempted to argue that the results of this study can instead be interpreted as indicating that

L2ers rely merely on a simple, linear morpheme co-occurrence association/restriction, rather than on underlying knowledge of negation–aspect interactions. However, such an interpretation becomes untenable if we consider the data in (9), repeated below as (12).

- (12) a. * Lisi **bu** qu-**le** Beijing.
 Lisi **not** go-**LE** Beijing
- b. * Lisi **bu** qu-**guo** Beijing.
 Lisi **not** go-**GUO** Beijing
- c. Lisi **bu** **keneng** qu-**le** Beijing.
 Lisi **not** **likely** go-**LE** Beijing
 i) ‘It’s not likely that Lisi went to Beijing.’
 ii) ‘It’s not likely that Lisi has gone to Beijing.’
- d. Lisi **bu** **keneng** qu-**guo** Beijing.
 Lisi **not** **likely** go-**GUO** Beijing
 ‘It’s not likely that Lisi has been to Beijing.’

If L2ers adopt a linear strategy in processing such sentences, they will attend only to the co-occurrence associations or restrictions linearly (i.e., not structurally), regardless of what occurs *in between* the relevant elements (i.e., a negator and an aspect marker). A linear association account would predict that L2ers treat (12a)/(12c) and (12b)/(12d) in the same way, although (12a) and (12b) are grammatical while (12c) and (12d) are ungrammatical. Sentences like (12c) and (12d), with an intervening element between negator and aspect marker, are not at all infrequent in Chinese; three common types of this are given in (13) and (14).

- (13) a. Lisi **bu** **yiding** qu-**le** Beijing.
 Lisi **not** **certainly** go-**LE** Beijing
 i) 'It's not certain that Lisi went to Beijing.'
 ii) 'It's not certain that Lisi has gone to Beijing.'
- b. Lisi **bu** **yiding** qu-**guo** Beijing.
 Lisi **not** **certainly** go-**GUO** Beijing
 'It's not certain that Lisi has been to Beijing.'
- (14) Lisi **mei** **keneng** qu-**le** Beijing.
 Lisi **not** **likely** go-**le** Beijing
 i) 'It's not likely that Lisi went to Beijing.'
 ii) 'It's not likely that Lisi has gone to Beijing.'

Supposing that L2ers process such sentences merely by linear associations and have learned, through input sentences such as (13) and (14), that the co-occurrence of *bu* with *-guo/-le* and *mei* with *-le* are grammatical; this could lead them to treat experimental sentences that violate the negation–aspect co-occurrence restrictions (**bu V-guo/-le* and **mei V-le*) as grammatical. On the other hand, if L2ers have learned, by linear associations, both that *bu* cannot co-occur with *-guo/-le* and that *mei* cannot co-occur with *-le*, then upon encountering sentences such as (13) and (14) they could treat them as ungrammatical. However, the results of the present study demonstrate that advanced L2ers do not treat the negation–aspect co-occurrence restrictions indiscriminately or randomly. On the contrary, they know very well which negator can and cannot co-occur with which aspect markers in a [negator + V + aspect marker] construction (i.e., *mei* can co-occur with *-guo* but not *-le*, and *bu* cannot co-occur with either *-guo* or *-le*). The L2ers' differentiated treatments of these co-occurrence patterns can hardly be reconciled to a mere linear association interpretation.

Finally, one possible question that can be raised about this study concerns the predicates used in the experimental sentences. To ensure that the experiment adequately assesses participants' processing of nonadjacent negation–aspect interactions, all of the

experimental sentences were created using verbal predicates that in principle allow either negator, not predicates that can be negated only by *bu* or only by *mei*. If, for example, one of the negators at issue is not permitted with a particular predicate, there is a danger of it leading to an earlier, local effect of ungrammaticality. Table 4.6 gives the 24 predicates of the experimental sentences, categorized according to their situation types (based on M. Li, 1999; Smith, 1994; Vendler, 1967).

Table 4.6. *Situation Types of the Predicates Used in the Experimental Sentences*

Situation Type	Predicates used
State ([+Static], [+Durative])	None
Activity ([-Static], [+Durative], [-Telic])	None
Accomplishment ([-Static], [+Durative], [+Telic])	<ol style="list-style-type: none"> 1. <i>biaoda zhe ge xiangfa</i> ‘express this idea’ 2. <i>chang na shou ge</i> ‘sing that song’ 3. <i>chi nayang de mifan</i> ‘eat that kind of rice’ 4. <i>da na ge tongxue</i> ‘beat that classmate’ 5. <i>du na pian wenzhang</i> ‘read that article’ 6. <i>jian na ge laoshi</i> ‘see that teacher’ 7. <i>jiao na ge che</i> ‘get that car’ 8. <i>jieshao zhe wei xiansheng</i> ‘introduce this gentlemen’ 9. <i>jieshi na ge danci</i> ‘explain that word’ 10. <i>kan zhe ben shu</i> ‘read this book’ 11. <i>lianxi na ju hua</i> ‘practice that sentence’ 12. <i>ma zhe ge ren</i> ‘curse that person’ 13. <i>mai zhe jian yifu</i> ‘buy this piece of clothing’ 14. <i>qingzhu ta de shengri</i> ‘celebrate his birthday’ 15. <i>qu na ge difang</i> ‘go to that place’ 16. <i>wan zhe ge youxi</i> ‘play this game’ 17. <i>wen na ge wenti</i> ‘ask that question’ 18. <i>xi na tiao kuzi</i> ‘wash that pair of pants’ 19. <i>xie na ge timu</i> ‘write that topic’ 20. <i>xue na men ke</i> ‘study that subject’ 21. <i>yong zhe tai diannao</i> ‘use this computer’ 22. <i>zuo na ge cai</i> ‘cook that dish’ 23. <i>zuo nali de ditie</i> ‘take that subway’
Achievement ([-Static], [-Durative], [+Telic])	<ol style="list-style-type: none"> 1. <i>tingshuo zhe jian shi</i> ‘hear of this thing’

Of the 24 predicates, 23 fell into the accomplishment type, and the other into the achievement type. All of the 23 accomplishment predicates can be negated either by *bu* or by *mei*. The achievement predicate as used in this study can be negated by *mei* (i.e., it is grammatical to say *mei tingshuo zhe jian shi* ‘didn’t hear/haven’t heard of this thing’)

but not by *bu* (i.e., it is ungrammatical to say * *bu tingshuo zhe jian shi*). However, *bu* can still negate this achievement predicate in a sentence such as that in (15).

- (15) Wo bu tingshuo zhe jian shi ye hui xinfan.
I not hear-of this CL thing still can trouble
'Even if I didn't hear of this thing, my heart could still be troubled.'

In light of this, the data from this achievement predicate were not excluded from the data analysis. The likelihood of this single predicate affecting the data patterns is negligible and not worrisome, and therefore the conclusion of this study regarding advanced L2ers' processing of the nonlocal structural relations remains intact.¹⁹

4.8 Conclusion

This study compared native and nonnative processing of negation–aspect interactions in Chinese and found that adult L2ers whose L1 lacks the relevant morphosyntactic features can both acquire them and process them in online sentence comprehension. This finding challenges deficit theories of L2 acquisition/processing but is easily accommodated by L2 proposals that reject the idea that post-puberty L2ers are confined to the grammatical features of their L1 in the acquisition of a nonnative language. These proposals, such as the Full Transfer/Full Access/Full Parse model (Dekydtspotter et al., 2006; Schwartz & Sprouse, 1996) and the Fundamental Identity Hypothesis (Hopp, 2007), hold that adult L2 knowledge/representation is not of an epistemologically different type from native-speaker knowledge and that adult L2 processing is, or at least can become, fundamentally the same as native-speaker processing. In addition, the current study demonstrates that adult L2ers are able to engage in fully specified morphosyntactic processing, even when it involves nonlocal structural dependencies (in the sense of Clahsen & Felser, 2006c, p. 564: “local dependencies, typically involving adjacent words or constituents”). Further, through the use of a holistic

¹⁹ For both experiments, analyses of the RT data with the achievement-predicate item excluded did not change the RT patterns. Some minor differences between the statistical tests including vs. excluding the item are shown in Appendix R. In most cases, those differences indicate that the statistical significance actually suffered a little bit from the item removal.

comprehension-focused processing task (i.e., CFSPR) vis-à-vis a structure-focused processing task (i.e., AJSPR), this study demonstrates that task demands influence participants' processing strategies. AJSPR tends to prompt incremental and fully specified structural processing while CFSPR is comparatively more likely to induce shallower and less detailed structural processing. The task effects of this study suggest that L2 shallow processing is not always an indication of inability; rather, it can simply be a performance strategy.

The present study illuminates the ways natives and L2ers process Chinese morphosyntax and contributes to a better understanding of (native and nonnative) sentence processing in Chinese. To date, published studies on Chinese sentence processing are almost exclusively concerned with how mature natives process the language. Indeed, studies on sentence-level processing of Chinese in general are still sparse (see Zhou, Ye, Cheung, & Chen, 2009, for a recent review). Chinese is a language well known for its scarcity of inflectional morphology (Zhou & Shu, 2011). It is distinct from typical European languages in crucial ways (e.g., lacking inflectional morphology for marking verb tense, subject-verb agreement, case; not marking word boundaries in the written form). Given these language-specific properties, lexical semantics and discourse context have been said to play primary roles over syntax in Chinese, much more so than in English or other European languages (e.g., C. N. Li & Thompson, 1981; P. Li, 1998). However, the present study suggests that at least in certain grammatical domains (such as negation and aspect marking), morphosyntax plays an indispensable role in Chinese sentence processing.

CHAPTER 5

PROCESSING SEMANTIC INFORMATION IN CHINESE

This chapter examines how native (“L1”) and nonnative (“L2”) speakers process semantic information in Chinese in order to assess whether, with respect to sentence processing in a particular target language (TL), adult L2 learners (“L2ers”) rely more on semantic information than L1 speakers (“natives”) do, as the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, 2006b, 2006c) contends. Phillips and Ehrenhofer (in press) pointed out that

the evidence for or against SSH generally involves demonstrations of L2ers’ sensitivity or insensitivity to structural cues that native speakers attend to in parsing, rather than evidence that L2ers are more strongly guided by lexical and conceptual associations.

This study probes the latter type of evidence. We report on two self-paced reading experiments that compare L1 and L2 online sensitivity to local semantic (im)plausibilities²⁰ that were created by making use of the selectional restrictions of Chinese verbs. We hypothesize that if, as the SSH claims, adult L2 processing is more strongly guided by semantics-based information than is the case in L1 processing, L2ers should be more sensitive to the local plausibility manipulation than natives are.

This chapter consists of five sections, arranged as follows: Section 5.1 reviews previous studies on L1 and L2 processing of lexical-semantic information at the sentence level. Sections 5.2 and 5.3 report on the two experiments that examine how natives and L2ers use lexical-semantic information in Chinese sentence processing. Section 5.4 presents a general discussion of the study. Section 5.5 is the conclusion.

²⁰ The (im)plausibilities at issue are actually (in)compatibilities. For the convenience of discussion, this study uses these two terms interchangeably.

5.1 Native and Nonnative Processing of Lexical-Semantic Information

Lexical semantics is one of the major sources of non-morphosyntactic information that the parser employs in sentence processing. Results from L1 processing research generally converge on the conclusion that semantic processing at the sentence level is incremental and lexical-semantic information is used immediately in online sentence interpretation (for a recent review, see Frisson & Traxler, 2013; but for reviews of counter evidence, see Pickering & van Gompel, 2006; van Gompel & Pickering, 2007).

L1 studies on semantic processing in Chinese have mostly been conducted in the context of assessing whether syntax takes precedence over semantics in sentence processing and whether semantic information produces rapid effects on sentence interpretation at the initial stage, employing techniques such as event-related potentials (ERP; e.g., S. Wang, Mo, Xiang, Xu, & Chen, 2013; Ye, Luo, Friederici, & Zhou, 2006; Yu & Zhang, 2008; Zhang et al., 2013; Zhang, Yu, & Boland, 2010; Zhou et al., 2010) and eye tracking (e.g., J. Yang, Wang, Chen, & Rayner, 2009). The semantic processes were investigated in these studies by using the violation paradigm. The semantic anomalies were usually introduced by replacing the critical verb in a normal control sentence with a semantically inappropriate verb so that the object would violate a selectional requirement of that verb. The results of these studies almost always overwhelmingly indicate that the use of semantic information is immediate, which is reflected in ERP studies by the presence of an N400 effect (i.e., a negative brain wave that peaks at around 400 milliseconds following a semantic anomaly, indexing difficulties in semantic processing) and in eye-tracking studies by longer word viewing times following semantic anomalies.

Since the present study investigates L1 and L2 processing of selectional restrictions in Chinese, the literature review below will focus on how selectional restrictions are processed in online sentence interpretation.

Selectional restrictions generally refer to the semantic constraints that a verb imposes on its arguments. Chomsky (1965) discussed selectional restrictions in terms of a small set of verb-general categories such as animacy and abstractness. However, later work extended the scope of selectional restrictions to include a wider range of categories.

Selectional restrictions have been investigated in studies of adult natives' semantic processing (e.g., Altmann & Kamide, 1999; Myers & Blumstein, 2005; Warren & McConnell, 2007). The studies show that adult natives are sensitive to a verb's selectional requirements and that an immediate processing difficulty is incurred if they are violated. For example, the selectional restriction of *blackmail* is violated in (1a) because spaghetti cannot be "blackmailed," whereas in (1b) there is no selectional-restriction violation because it is quite normal for spaghetti to be drained.

- (1) a. ? The man used a photo to blackmail the thin spaghetti yesterday evening.
b. The man used a strainer to drain the thin spaghetti yesterday evening.

Relative to controls as (1b), violations of selectional restrictions as in (1a) caused immediate disruption in readers' eye movements and resulted in longer processing times at the target (i.e., *spaghetti*) region (Warren & McConnell, 2007).

L1 studies also suggest that selectional-restriction information can be used by the parser to predict a verb's upcoming referent. In an eye-tracking study using the visual-world paradigm, Altmann and Kamide (1999) showed that auditory stimuli such as *The boy will eat ...* immediately led participants to look at the picture of an edible object (e.g., a cake) rather than a picture of a distractor object (e.g., a balloon, a toy train set, or a toy car) in the visual scene on the computer screen. In other words, saccadic eye movements were launched earlier to a picture of the semantically appropriate direct object (DO) of the verb, even before the DO had been heard. In contrast, no such anticipatory eye movements were observed when participants heard *The boy will move ...* although the auditory stimulus was paired with the same visual scene.

The immediate influence of selectional restrictions is also seen in priming studies. In the short verb-noun pair *mail-letter*, the noun is compatible with the selectional requirements of the verb, because a letter is something that can be mailed. In the verb-noun pair *persuade-letter*, however, *letter* does not fit the selectional restrictions of *persuade*, because a letter is not persuadable. A priming study by Myers and Blumstein (2005) used such word pairs to test the effects of selectional restrictions. For the nouns that fit the verbs' selectional restrictions, priming effects were obtained both at the lexical

level (the verb-noun pairs occurred without a syntactic context) and at the sentential level (the verb-noun pairs appeared as the verb and the object in simple present-tense, subject-verb-object sentences, e.g., *The men mail the letter* vs. *The men persuade the letter*).

Briefly, previous studies have demonstrated that lexical-semantic information such as selectional restrictions has an immediate effect on L1 sentence processing. The effect is seen in elevated processing times when selectional restrictions are violated during sentence comprehension, in anticipatory looks at the referent cued by the selectional restrictions of a verb, and in verbs' priming of noun targets that are compatible with the verbs' selectional requirements.

Turning to studies on L2 processing of semantic information, L2ers generally are able to make use of lexical semantics in interpreting L2 sentences (for a recent review, see Roberts, 2013). This is especially evidenced in studies using ERP technologies (for reviews of these ERP studies, see Frenck-Mestre, German, & Foucart, 2014 and Moreno, Rodríguez-Fornells, & Laine, 2008). Most of the ERP studies employed a semantic violation paradigm. They compared L1 and L2 semantic processing in such sentences as *Die Tür wurde geschlossen* 'The door was being closed' vs. *Der Ozean wurde geschlossen* 'The ocean was being closed' (Hahne, 2001) and *The bread was eaten* vs. *The volcano was eaten* (Hahne & Friederici, 2001). The studies observed that semantic anomalies (resulting from violations of selectional restrictions) reliably elicit N400 effects for L2ers, similar to those found in adult natives, although the effects in L2ers are often delayed and/or reduced in amplitude (e.g., Bowden, Steinhauer, Sanz, & Ullman, 2013; Clahsen & Felser, 2006a, 2006b, 2006c).

This type of ERP evidence in L2 semantic processing is a major impetus for the SSH (Clahsen & Felser, 2006a, 2006b, 2006c), according to which L2ers' sentence comprehension involves, essentially, the computation of "predicate-argument structure representations of the input that capture thematic roles and other aspects of lexical-semantic structure, but which lack hierarchical detail and more abstract elements of syntactic structure" (Clahsen & Felser, 2006b, p. 32). In essence, the SSH contends that L2 sentence processing depends primarily on meaning-based information, including lexical-semantic information such as thematic roles and selectional restrictions.

However, the differences between L1 and L2 semantic processing do not always seem to be quantitative in nature. For example, the study by Newman, Tremblay, Nichols, Neville, and Ullman (2012) revealed some qualitative differences between L1 and L2 semantic processing. In this study, both natives and adult L2ers showed N400 and P600 effects²¹ when processing lexical-semantic violations in English sentences such as *The Irishman sipped Todd's whiskey at the party* vs. *The Irishman sipped Todd's thunder at the party*. However, the N400 effect was more prominent in natives, whereas the P600 effect was more prominent in L2ers; moreover, the asymmetrical distribution of these effects was not affected by the L2ers' English proficiency. These processing patterns suggest that a qualitative difference exists between L1 and L2 semantic processing.

In a review of recent ERP studies, Frenck-Mestre et al. (2014, p. 248) also pointed out some qualitative differences between L1 and L2 semantic processing:

When processing is examined for different levels of semantic integration difficulty, the L1 results consistently show a frontal positivity subsequent to the more typically reported N400 modulation, in contrast to the results for L2 processing which show only modulations in the N400 response.

It looks like L2 semantic processing is not always the same as that of L1, and there may be some qualitative differences between the two. Apparently, the issue is open to further investigation.

Probably because most L2 researchers assume that adult L2ers have little problem with lexical semantics (in the TL), behavioral studies on L2 sentence processing have almost exclusively focused on syntactic processing (for a review, see Roberts, 2013). Only a few studies have looked at L2 semantic processing when researching L2 syntactic processing (e.g., Felser et al., 2003; Jackson & Roberts, 2010; Papadopoulou & Clahsen, 2003; Williams, 2006; Williams et al., 2001). The studies indicate that lexical-semantic

²¹ ERP studies on L1 processing of lexical-semantic information at the sentence level often find a P600 component (a centroparietal positivity that peaks around 600 ms, typically indexing difficulty in syntactic processing) in addition to an N400 effect (e.g., DeLong, Urbach, Groppe, & Kutas, 2011; Kuperberg, 2007; Kuperberg et al., 2007; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003); this is the so-called "semantic P600" in the literature.

information influences L2 sentence processing and hence is not as problematic as morphosyntactic information.

So far, little research attention has been directed to specifically assessing whether adult L2ers are more strongly guided by lexical-semantic information than natives are in the course of sentence processing. A self-paced reading study by Roberts and Felser (2011) seems to suggest that L2 processing does indeed rely more on lexical-semantic information than L1 processing does. In their study, advanced L1-Greek L2ers of English appeared to be more easily influenced by lexical-semantic information than natives in reading locally ambiguous (i.e., “garden-path” [GP]) sentences such as (2) and (3).

- (2) a. The inspector **warned the boss** would destroy very many lives.
b. The inspector **warned the crimes** would destroy very many lives.
- (3) a. While the band **played the song** pleased all the customers.
b. While the band **played the beer** pleased all the customers.

The NPs *the boss* and *the crimes* in (2) and *the song* and *the beer* in (3) are temporarily ambiguous between objects and subjects. While the sentence fragments *warned the boss* in [2a] and *played the song* in [3a] are plausible, the fragments *warned the crimes* in [2b] and *played the beer* in [3b] are implausible. The L2ers showed reliable plausibility effects: Their reading times (RTs) were significantly shorter at and also immediately following the critical region in the plausible conditions (*boss* in [2a] and *song* in [3a]), compared to the same regions (*crimes* in [2b] and *beer* in [3b]) in the implausible conditions. Moreover, the L2ers’ plausibility effects for sentences such as (2) reversed not at the disambiguating region (e.g., the modal *would* in [2]) but the one following it (e.g., *destroy* in [2]), indicating that it was difficult for the L2ers to recover from an initial misanalysis. In contrast, the natives only showed some weak plausibility effects for sentences such as (2) (showing up at the post-critical modal region and statistically significant only by participants, not by items); for sentences such as (3), they failed to show any significant effect of plausibility. In sum, the natives were much less subject to

the influence of plausibility. The temporarily plausible DOs (e.g., *the boss* in [2a] and *the song* in [3a]) did not seem to garden-path the natives as much as they did the L2ers.

However, as Roberts and Felser (2011) pointed out, natives normally show significant plausibility effects (or “GP effects” here) in reading such sentences (e.g., Pickering & Traxler, 1998). The differences between natives and L2ers in Roberts and Felser (2011) might have to do with the short distance between the ambiguous NP and the disambiguating element (i.e., the modal or the tensed verb), perhaps making it very easy for natives to abandon the initial incorrect analysis, with the result that the effects were not large enough to be detectable online. In addition, the natives read the experimental stimuli much faster than the L2ers did, which might have masked the plausibility effects (e.g., the effects went to after the end of the sentence and hence could not be seen online). Another possibility is that the semantic and reanalysis effects, which are in opposed directions, cancelled each other out. In any case, the natives’ less prominent (or null) effects of plausibility could simply be due to a Type II error (i.e., failing to reject the null hypothesis). Undoubtedly, the issue of whether adult L2ers are more strongly guided by lexical-semantic information in sentence processing than natives are remains inconclusive and needs further investigation.

The present study attempts a systematic investigation of this issue. It compares how natives and L2ers process lexical-semantic information in interpreting Chinese sentences, capitalizing on the selectional requirements of Chinese transitive verbs to manipulate temporary, or local, (im)plausibilities in sentences such as (4).

(4) a. Locally Plausible Condition

Tamen **haipa** [NP [NP **shouji**] de²² [N yonghu]] youxie fanchang.
 3pl **fear** **cellphone** DE user some unusual
 ‘It’s a little unusual that they feared the cellphone’s user.’

²² This *de* is a prenominal modifier marker (Huang, 1998). It is sometimes called a nominalizer (e.g., C. N. Li & Thompson, 1981). Here, it functions to realize the genitive case, similar to possessive -’s in English. The analysis of the *de* structure in (4) is based on Huang, Li, and Li (2009, pp. 14–36).

b. Locally Implausible Condition

Tamen **ganxie** [_{NP} [_{NP} **shouji**] de [_N yonghu]] youxie fanchang.
3pl **thank** **cellphone** DE user some unusual
'It's a little unusual that they thanked the cellphone's user.'

In the NP *shouji de yonghu* 'the cellphone's user' in (4), the first nominal *shouji* 'cellphone' (which is also an NP) is locally ambiguous between a DO and a prenominal modifier. In (4a), the verb *haipa* 'fear' can take either inanimate or animate DOs (e.g., *shouji* 'cellphone' and *yonghu* 'user'); the selectional restrictions of the verb can be satisfied by *shouji* 'cellphone', and hence *shouji* can serve as a locally plausible DO. In (4b), however, the verb *ganxie* 'thank' normally takes only human DOs (e.g., *yonghu* 'user'); when *shouji* 'cellphone' acts as a local DO, the selectional restrictions of the verb are violated because *shouji* is inanimate. The violation gives rise to a temporary semantic incompatibility in (4b). If semantic processing is incremental, the local implausibility in (4b) should interrupt comprehenders' processing and result in longer processing time in (4b) relative to (4a) at and/or immediately following the critical region (i.e., the ambiguous NP *shouji* 'cellphone' region).

Since the SSH contends that L2 processing, compared to mature L1 processing, is directed less by structural information and more by semantic information, it predicts that local implausibility effects will be even stronger in L2ers than in natives and will also lead L2ers to experience increased reanalysis difficulty at and/or immediately following the disambiguating region. Such a prediction was actually made in Roberts and Felser (2011):

If learners are more strongly guided by plausibility information during L2 processing than native speakers [are], then effects of plausibility should be more pronounced and/or prolonged for the learners, who should also have more difficulty than native speakers recovering from an initially plausible misanalysis. (p. 304)

Note that in the current study, recovery from the initial incorrect analysis involves reanalyzing the ambiguous NP (e.g., *shouji* ‘cellphone’ in [4]) as a modifier rather than a DO. This necessitates using the structural information of *de* (at the post-critical region) as a prenominal modification marker and reanalyzing the NP *shouji* as the specifier of the head N *yonghu* ‘user’, rather than as the DO itself.²³ Thus, a reanalysis difficulty indexes inability to use the structural information.

The current study tests the predictions of Roberts and Felser (2011), following the SSH claim, in two experiments. Experiment 5.1 uses a comprehension-focused self-paced reading (CFSPR) task, in which participants are asked to focus on comprehending meaning. Experiment 5.2 uses an acceptability-judgment self-paced reading (AJSPR),²⁴ to see whether expected results from Experiment 5.1 could be extended to a task that focuses participant more on detailed structural processing.

The major research questions of the present study are as follows:

1. Do native and nonnative adult speakers of Chinese make immediate use of lexical-semantic information in sentence processing, as indicated by L1 and L2 participants’ sensitivity to temporary semantic incompatibilities at the critical and/or post-critical regions?
2. Does L2 processing over-rely on lexical-semantic information, as indicated (a) by more robust and prolonged effects of temporary semantic incompatibilities in L2ers than in natives at and/or immediately following the critical region and (b) by increased reanalysis difficulty (in the locally plausible condition) for L2ers than for natives at the disambiguating regions?

²³ Note that *de* is not a disambiguating word because prior to the appearance of the head N, *de* is ambiguous between a modifier marker and a (relative) clause marker. As the latter, *de* may not always lead to a reanalysis of *shouji* ‘cellphone’ from a DO to a modifier. For example, in *Tamen haipa shouji de yuanyin* [3pl fear cellphone DE reason] ‘the reason that they feared the cellphone’, *shouji* remains a DO.

²⁴ Both sentences in (4) are acceptable; for the acceptability-judgment task to work, we used unacceptable fillers (see Section 5.3 for details).

5.2 Experiment 5.1: Comprehension-Focused Self-Paced Reading

This experiment compares L1 and L2 processing of lexical-semantic information in Chinese using a CFSPR task. The main purpose was to assess whether in a meaning-focused online task, L2 processing, in comparison to L1 processing, indeed exhibits an over-dependence on lexical-semantic information.

5.2.1 Method

5.2.1.1 Participants

Thirty adult natives and 38 adult L2ers of Chinese (20 L1-English and 18 L1-Japanese) participated in this experiment. They were paid a fee for participation. One Japanese L2er was excluded due to a high error rate for the type of filler designed for the purpose of assessing participants' attentiveness in performing the self-paced reading task ($M_{\text{error rate}} = 37.5\%$).²⁵

For the purpose of assessing L2 proficiency effects, L2 participants were divided into 2 proficiency groups (Advanced and Intermediate) according to their scores on a Chinese proficiency test, a quasi-C-test developed for this dissertation research (for a description of the test, see Chapter 3). The participants in the Advanced L2 group (10 L1-English and 8 L1-Japanese) all scored 22 and above on the proficiency test. Those who scored less than 22 were grouped as Intermediate (10 L1-English; 9 L1-Japanese). The cutoff point was based on the L2ers' median score. Note that the L2 participants were not grouped by L1. In this study, we did not expect L1 background to be a factor that consistently influences the final results. On the one hand, the Chinese verbs used in this study were not chosen according to similarities or differences in the selectional restrictions of the corresponding verbs in English or Japanese. On the other hand, those Chinese verbs often lack a one-to-one correspondence with the verbs in English or Japanese. A Chinese verb may correspond to multiple verbs in English and/or Japanese. It is difficult to know which one of such multiple verbs exerts the most influence in the processing of the Chinese verb, thereby making it practically impossible to design an experiment based on consistent match or mismatch patterns in selectional restrictions across these three languages. Let us look at one specific example. The Chinese verb

²⁵ For details about this type of filler ($k = 8$), see Chapter 3: General Methods.

wuhui (误会) corresponds to at least two English verbs, namely *misunderstand* and *misinterpret*, and also to at least two Japanese verbs, *gokai suru* (誤解する) and *kanchigai suru* (勘違いする). Moreover, the selectional restrictions of these verbs differ in very subtle ways in these languages: The Chinese verb *wuhui* (误会) is mostly used to mean ‘misunderstand someone’ and hence requires a [+human] direct object (e.g., *wo wuhui ta le* [我误会他了] [I misunderstand him LE] ‘I misunderstood him’); English *misunderstand* and *misinterpret*, when used transitively, can have either [+human] or [–human] DOs, although *misunderstand someone* is probably more common than *misinterpret someone*; Japanese *gokai suru* (誤解する) and *kanchigai suru* (勘違いする) can take either [+human] or [–human] objects, but *kanchigai suru* (勘違いする) is often used as an intransitive verb.²⁶ Differences such as these make it very difficult, if not impossible, to check similarities or differences in selectional restrictions of the relevant verbs across these three languages. Since we did not control for L1 in this study, any possible L1-based influence on the final results should be considered to be due to chance.

Table 5.1 gives the 3 participant groups’ Chinese background information.

Table 5.1. *Chinese Background Information of the Participants in Experiment 5.1*

Group	Years of learning Chinese		Years of residence in China/Taiwan		Chinese proficiency scores (Max = 50)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	Natives (<i>n</i> = 30)	N/A	N/A	N/A	N/A	45.9
Advanced L2ers (<i>n</i> = 18)	5.2	4.7	3.0	3.7	35.1	8.8
Intermediate L2ers (<i>n</i> = 19)	2.5	2.2	0.5	0.3	14.4	5.3

A one-way ANOVA performed on the participants’ mean Chinese proficiency scores revealed significant differences between the 3 groups’ proficiency levels, $F(2, 64) = 193.26, p < .001$. SPSS Bonferroni corrected post hoc tests²⁷ indicate that the Native group exceeded the 2 L2 groups in Chinese proficiency ($ps < .001$), and the Advanced L2 group exceeded the Intermediate L2 group ($p < .001$).

²⁶ The Japanese data are based on my personal communications with M. Kaohu, Y. Kobayashi, M. Miyao, A. Takeda, and S. Yamaguchi, December 2–5, 2014.

²⁷ The *p*-values reported for the post hoc comparisons in this study were adjusted in SPSS (see Kinneer & Gray, 2004), using the Bonferroni method. The original *p*-values were each multiplied by the total number of pairwise comparisons. The derived *p*-values reflect Bonferroni adjustments while the experiment-wise alpha level of .05 remains unchanged.

5.2.1.2 Materials

The study used 16 experimental items like (4). The two versions of each item were exactly the same except for the verbs used for manipulating local plausibilities. The verbs differ in selectional restrictions. The use of two verbs with different selectional restrictions gives rise to two experimental conditions: *locally plausible* condition vs. *locally implausible* condition. In the locally plausible condition, the verbs allow both animate and inanimate objects; in the locally implausible condition, the verbs only allow human objects. The two versions of each item were evenly distributed across 2 lists in a Latin square design.

In addition to the 16 experimental trials ($k = 8$ per condition per list), the experiment used 72 filler sentences to conceal what was being tested. The fillers were sentences of different types: ungrammatical due to semantic incompatibilities ($k = 8$); ungrammatical due to syntactic violations ($k = 12$); grammatical but containing local semantic incompatibilities ($k = 16$); grammatical, containing no semantic incompatibilities ($k = 28$). Moreover, one special type of filler ($k = 8$; grammatical) was also used for assessing participants' attentiveness in performing the self-paced reading task (see Chapter 3). Each participant was given one list (a total of 88 sentences plus 8 practice sentences). The sentences were pseudo-randomized when presented to participants.

Each item was followed by a comprehension question. For one half of experimental sentences, the comprehension questions were like these for (4): *Tamen haipa yonghu, dui ma?* 'They feared the user, right?' (for [4a]) and *Tamen ganxie yonghu, dui ma?* 'They thanked the user, right?' (for [4b]). The correct answer to these questions was *dui* 'right'. Correct answers to these questions require participants to understand the actual DO of the verbs (because the questions asked about the head of the whole NP) and thus can index full recovery from any initial misanalysis (in the locally plausible condition). For the other half of experimental sentences, the comprehension questions were designed in a different way. Consider, for example, the experimental sentences given in (5).

(5) a. Locally Plausible Condition

Tamen liaojie gongchang de gongren shi yinggaide.
3pl understand factory DE worker be necessary
'It was necessary that they understood the factory's workers.'

b. Locally Implausible Condition

Tamen yongbao gongchang de gongren shi yinggaide.
3pl hug factory DE worker be necessary
'It was necessary that they hugged the factory's workers.'

The comprehension question for the locally plausible condition (5a) was like this: *Tamen liaojie gongchang, dui ma?* 'They understood the factory, right?' The correct answer to this question should be *bu dui* 'not right'. This question asked about the nominal (*gongchang* 'factory') that immediately follows the verb, instead of the head (*gongren* 'worker') of the whole NP. Questions like this were used to test whether the participants practiced shallow processing (e.g., Christianson et al., 2001; Ferreira et al., 2002). If participants attempted to integrate the NP *gongchang* 'factory' with the verb *liaojie* 'understand' and treated the temporarily plausible DO of the verb as the ultimate DO, they would incorrectly answer *dui* 'right'. In this case, it would suggest that participants only understood the experimental sentence superficially and could not abandon the initial misanalysis of *gongchang* as the DO. For the locally implausible condition, the comprehension questions asked about the head of the NPs, like this: *Tamen mei yongbao gongren, dui ma?* 'They didn't hug the workers, right?' (for [5b]). The correct answer to this question was *bu dui* 'not right' because the truth of (5b) is that they did hug the workers. By comparing participants' answers to these two types of question, we can assess the degree to which participants exercise shallow processing. (We will return to this in the Results and Discussion sections.)

The overall *dui* 'right' and *bu dui* 'not right' answers for the experimental sentences were balanced. (For a complete list of the experimental sentences and the corresponding comprehension questions, see Appendix S.) As for the comprehension

questions on the fillers, 32 required *dui* ‘right’ answers, and 40 required *bu dui* ‘not right’ answers.

A sample pair of experimental sentences, with regions indicated, is provided in Table 5.2.

Table 5.2. *Sample Pair of Experimental Stimuli in Experiment 5.1*

Conditions	Regions						
	1	2	3	4	5	6	7
Locally plausible	她们	害怕	手机	的	用户	有些	反常。
	Tamen	haipa	shouji	de	yonghu	youxie	fanchang.
	3pl	fear	cellphone	DE	user	some	unusual
Locally implausible	她们	感谢	手机	的	用户	有些	反常。
	Tamen	ganxie	shouji	de	yonghu	youxie	fanchang.
	3pl	thank	cellphone	DE	user	some	unusual

Each experimental sentence was split into 7 regions for the purpose of the self-paced reading task. The regions of interest are Regions 3–6. Region 3 is the critical region, and Region 4 is the post-critical region. If there is any local plausibility effect, it should show up at these regions. The RT pattern should be that the locally plausible condition takes shorter time than the locally implausible condition. Regions 5–6 are the disambiguating regions. If participants experience any difficulty in reanalysis (the locally plausible condition), it should appear at these regions in the form of a reversed RT pattern (i.e., the locally plausible condition taking longer time than the locally implausible condition), because the initial commitment to a DO analysis of the temporarily ambiguous NP in the locally plausible condition should result in revision difficulty later on at the disambiguating region.

To test whether the verbs used in each sentence pair were appropriate for the experimental conditions, we conducted a norming study with 30 Chinese adult natives. They were all recruited in Beijing and the vicinities, and they did not participate in the main experiments. The norming study included two parts. The first part tested whether all the verbs can take [+human] DOs. Participants ($n = 10$) were presented with 32 simple sentences like (6), in which all 32 verbs and all 16 human DOs were embedded.

- (6) Tamen changchang **ganxie** naxie **yonghu.**
3pl often **thank** those **users**
'They often thank those users.'

Participants were asked to decide whether each sentence was acceptable or unacceptable. The sentences were randomized before being presented to the participants. To ensure that participants made judgments based on their first impression, each sentence was read aloud to them individually by a female native speaker of standard Mandarin Chinese. The sentences were read only once at normal speed. After hearing each sentence, participants made their judgment by saying *tongshun* 'well-formed, acceptable' or *bu tongshun* 'not well-formed, not acceptable' (for a discussion of these terms for acceptability judgments in Chinese, see Lu et al., 2000, as summarized in Section 3.3). Participants' responses were recorded by the experimenter. This part of the norming study did not use fillers (for a complete list of the norming sentences in Part 1, see Appendix T).

The acceptability-judgment accuracy results are presented in Table 5.3. Note that the acceptability-judgment accuracy for this set of sentences indicates participants' correctly accepting a sentence as well-formed, coherent, and appropriate, and hence shall be called *acceptance* hereafter.

Table 5.3. *Participants' (n = 10) Acceptance Rates of the Norming Sentences (k = 32) Used for Assessing Whether the Verbs Can Take Human Objects*

Verb + Human object NP	<i>M</i>	<i>SD</i>
Verbs used in the locally plausible condition		
1. <i>aihu</i> 'take care of' + <i>tongxue</i> 'classmate'	1.00	0.00
2. <i>bangzhu</i> 'help' + <i>jingli</i> 'manager'	1.00	0.00
3. <i>chengzan</i> 'praise' + <i>zuozhe</i> 'author'	1.00	0.00
4. <i>gaobie</i> 'bid farewell to' + <i>linju</i> 'neighbor'	0.80	0.42
5. <i>guanzhu</i> 'give attention to' + <i>xuesheng</i> 'student'	1.00	0.00
6. <i>haipa</i> 'fear' + <i>yonghu</i> 'user'	1.00	0.00
7. <i>hen</i> 'hate' + <i>yisheng</i> 'doctor'	1.00	0.00
8. <i>huainian</i> 'miss' + <i>pengyou</i> 'friend'	1.00	0.00
9. <i>liaojie</i> 'understand' + <i>gongren</i> 'worker'	1.00	0.00
10. <i>likai</i> 'leave' + <i>jumin</i> 'resident'	1.00	0.00
11. <i>qu-zhao</i> 'go to find' + <i>zhuren</i> 'owner'	1.00	0.00
12. <i>taoyan</i> 'dislike' + <i>zuozhe</i> 'author'	1.00	0.00
13. <i>xi'ai</i> 'love' + <i>yanyuan</i> 'actor'	1.00	0.00
14. <i>xiang-kan</i> 'want to see' + <i>siji</i> 'driver'	1.00	0.00
15. <i>xiangnian</i> 'miss' + <i>tongshi</i> 'colleague'	1.00	0.00
16. <i>xihuan</i> 'like' + <i>tongban</i> 'friend'	1.00	0.00
Verbs used in the locally implausible condition		
1. <i>da</i> 'beat' + <i>yisheng</i> 'doctor'	1.00	0.00
2. <i>daji</i> 'attack' + <i>zuozhe</i> 'author'	1.00	0.00
3. <i>daying</i> 'promise' + <i>tongxue</i> 'classmate'	1.00	0.00
4. <i>fandui</i> 'oppose' + <i>jumin</i> 'resident'	1.00	0.00
5. <i>ganxie</i> 'thank' + <i>yonghu</i> 'user'	1.00	0.00
6. <i>huanying</i> 'welcome' + <i>xuesheng</i> 'student'	1.00	0.00
7. <i>qinwen</i> 'kiss' + <i>jingli</i> 'manager'	1.00	0.00
8. <i>quanshuo</i> 'persuade' + <i>zuozhe</i> 'author'	1.00	0.00
9. <i>qu-song</i> 'go to send off' + <i>tongshi</i> 'colleague'	1.00	0.00
10. <i>renshi</i> 'get to know' + <i>tongban</i> 'friend'	1.00	0.00
11. <i>shanghai</i> 'hurt' + <i>pengyou</i> 'friend'	1.00	0.00
12. <i>tongyi</i> 'agree with' + <i>yanyuan</i> 'actor'	0.70	0.48
13. <i>wuhui</i> 'misunderstand' + <i>linju</i> 'neighbor'	1.00	0.00
14. <i>xiangxin</i> 'trust' + <i>zhuren</i> 'owner'	1.00	0.00
15. <i>yongbao</i> 'hug' + <i>gongren</i> 'worker'	1.00	0.00
16. <i>yuanliang</i> 'forgive' + <i>siji</i> 'driver'	1.00	0.00
Overall	0.98	0.12

As can be seen from Table 5.3, the participants' average acceptance rates for the 32 sentences were at least 70%, with an overall mean of 98%, indicating that the verbs go well with the [+human] DOs.

The second part of the norming study, which appeared in the form of a paper-and-pencil questionnaire survey, tested whether the 16 verbs that were used in the locally plausible condition can indeed also take [-animate] DOs. Twenty adult natives

were asked to read 16 incomplete sentences like (7) and choose one verb in each verb pair to fill in the blank.

- (7) Tamen ____ (A. ganxie B. haipa) shouji youxie fanchang.
 3pl ____ (A. thank B. fear) cellphone some unusual
 ‘It’s a little unusual that they ____ (A. thanked B. feared) the cellphone.’

The NP immediately following the blanks were the 16 [-animate] nominals used in this study. To counterbalance the answers, half of the intended correct options appeared as A, and half as B (for a complete list of the norming sentences in Part 2, see Appendix U). The sentences were mingled with 128 filler sentences (used as materials for other norming studies). The norming stimuli were presented to participants randomly, in simplified Chinese characters. Table 5.4 gives the participants’ mean accuracy rates for the 16 relevant norming sentences.

Table 5.4. *Participants’ (n = 20) Verb-Selection Accuracy Rates for the Norming Sentences (k = 16) Used for Assessing Whether the Verbs in the Locally Plausible Condition Can Take Inanimate Objects*

Verb + Inanimate object NP	<i>M</i>	<i>SD</i>
1. <i>aihu</i> ‘take care of’ + <i>sushe</i> ‘dormitory’	0.95	0.22
2. <i>bangzhu</i> ‘help’ + <i>gongsi</i> ‘company’	0.75	0.44
3. <i>chengzan</i> ‘praise’ + <i>wenzhang</i> ‘article’	0.85	0.37
4. <i>gaobie</i> ‘bid farewell to’ + <i>xiaoqu</i> ‘district’	1.00	0.00
5. <i>guanzhu</i> ‘give attention to’ + <i>nongcun</i> ‘rural area’	0.95	0.22
6. <i>haipa</i> ‘fear’ + <i>shouji</i> ‘cellphone’	0.95	0.22
7. <i>hen</i> ‘hate’ + <i>yiyuan</i> ‘hospital’	1.00	0.00
8. <i>huainian</i> ‘miss’ + <i>jiexiang</i> ‘hometown’	1.00	0.00
9. <i>liaojie</i> ‘understand’ + <i>gongchang</i> ‘factory’	1.00	0.00
10. <i>likai</i> ‘leave’ + <i>loufang</i> ‘building’	1.00	0.00
11. <i>qu-zhao</i> ‘go to find’ + <i>fangzi</i> ‘house’	1.00	0.00
12. <i>taoyan</i> ‘dislike’ + <i>xiaoshuo</i> ‘novel’	1.00	0.00
13. <i>xi’ai</i> ‘love’ + <i>dianshiju</i> ‘TV drama’	1.00	0.00
14. <i>xiang-kan</i> ‘want to see’ + <i>qiche</i> ‘car’	1.00	0.00
15. <i>xiangnian</i> ‘miss’ + <i>xuexiao</i> ‘school’	1.00	0.00
16. <i>xihuan</i> ‘like’ + <i>gaozhong</i> ‘high school’	0.70	0.47
Overall	0.95	0.22

As can be seen from Table 5.4, the mean accuracy for each verb was at least 70%.

Overall, the participants chose the target verbs accurately 95% of the time, suggesting

that the verbs used in the locally plausible condition can indeed take [-animate] DOs and therefore were appropriate for this study.

In addition to the norming study, the verbs used in this study were also controlled for word length, number of strokes, and frequency. The verbs in each pair always contain the same number of syllables (i.e., Chinese characters). The verbs in the two experimental conditions do not differ from each other in number of strokes, $t(15) = 1.06$, $p = .305$, nor in word frequency (based on the 50,000,000-word Routledge Chinese Frequency Dictionary Corpus; Xiao et al., 2009), $t(15) = 0.47$, $p = .643$.

5.2.1.3 Procedure

The CFSPR task used a noncumulative moving-window paradigm (Just et al., 1982). Participants read the sentences on a computer monitor in a segment-by-segment fashion at their own natural pace. A segment contained either a word or character. The sentences were displayed in simplified Chinese characters (font: STSong; font size: 14). Each test sentence occupied a single line on the computer screen. Participants were tested individually, and they were naïve to the purposes of the experiment. Before the experiment started, participants were given specific instructions as well as a specific example (see Appendix I). The experiment was preceded by 8 practice items (see Appendix J). After reading each sentence, participants responded to a *dui* ‘right’/*bu dui* ‘not right’ comprehension question by pressing one of the two designated keys for *Dui* (Yes) and *Bu dui* (No) on the keyboard. The DMDX software (Forster & Forster, 2003) was used to run the experiment. After the self-paced reading experiment, participants were asked to complete a language background questionnaire (see Appendix E) and the Chinese proficiency test (see Appendix F). The entire experimental session lasted about 25–45 minutes.

5.2.2 Results

Data analysis covered both participants’ comprehension accuracy and reading times (RTs), which are reported separately in the following sections.

5.2.2.1 Comprehension Accuracy

As mentioned in Section 5.2.1.2, one special type of filler ($k = 8$) was used for checking participants' attentiveness in performing the self-paced reading task. From Table 5.5, we can see that all the 3 participant groups obtained high accuracy rates for this type of filler (at least 93%). The high accuracy indicates high attentiveness for all the groups in doing the online reading task. The accuracy rates of the other fillers were also very good (at least 81%).

Table 5.5. Mean Comprehension Accuracy Rates for the Fillers in Experiment 5.1

Group	Type	
	Special ($k = 8$)	Other ($k = 64$)
Natives ($n = 30$)	.97 (.06)	.94 (.05)
Advanced L2ers ($n = 18$)	.93 (.09)	.92 (.05)
Intermediate L2ers ($n = 19$)	.96 (.07)	.81 (.08)

Note. Standard deviations are in parentheses.

For the experimental sentences, the 3 groups' mean comprehension accuracy rates by condition are presented in Figure 5.1 (for the descriptive statistics, see Appendix V).

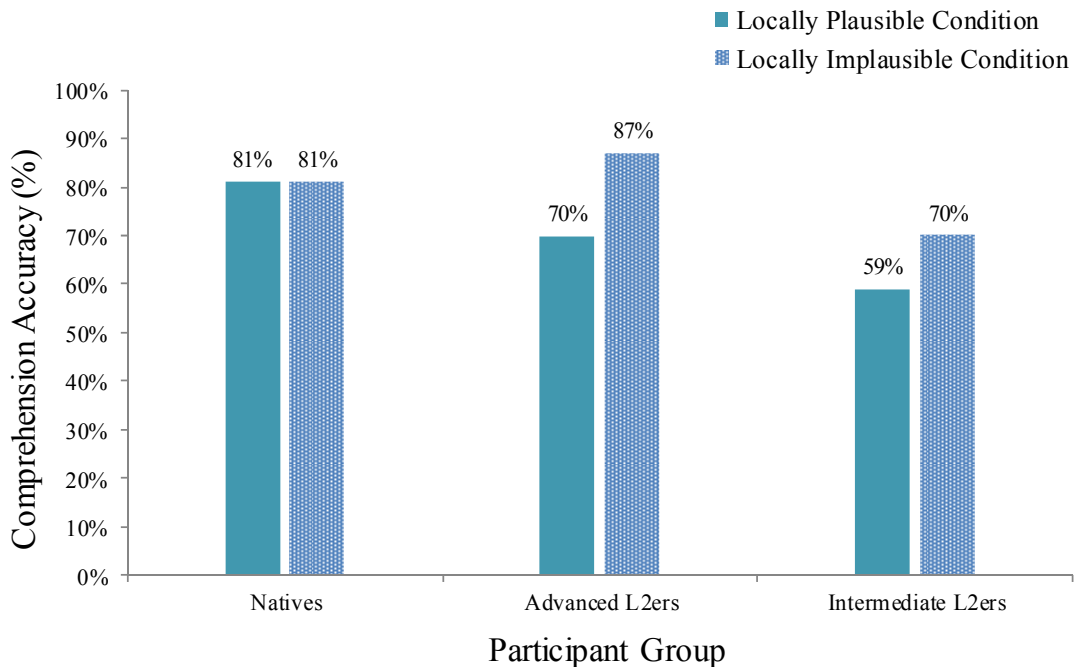


Figure 5.1. Mean Comprehension Accuracy Rates for the Experimental Sentences of Experiment 5.1.

As can be seen from Figure 5.1, the Natives' mean comprehension accuracy rates are the same between the conditions (81%). For the L2ers, there was a difference between the conditions: Both groups had lower accuracy rates in the locally plausible condition than in the locally implausible condition (the Advanced L2 group: 70% vs. 87%; the Intermediate L2 group: 59% vs. 70%). In addition, the accuracy data also reveal some effects of Chinese proficiency on the L2ers' accuracy rates.

Note that the accuracy data reported above do not specify the specific types of comprehension questions involved (for a complete list of the comprehension questions, see Appendix S). As described in Section 5.2.1.2 (Materials), half of the experimental sentences were followed by comprehension questions designed to examine whether participants exercised coarse and shallow processing strategies. Let us look at such experimental sentences as in (5), repeated here as (8):

(8) a. Locally Plausible Condition

Tamen liaojie gongchang de gongren shi yinggai de.
3pl understand factory DE worker be necessary
'It was necessary that they understood the factory's workers.'

b. Locally Implausible Condition

Tamen yongbao gongchang de gongren shi yinggai de.
3pl hug factory DE worker be necessary
'It was necessary that they hugged the factory's workers.'

The comprehension question for the locally plausible condition (8a) was *Tamen liaojie gongchang, dui ma?* 'They understood the factory, right?' A *dui* 'right' answer to this question would suggest shallow processing because the locally possible DO *gongchang* 'factory' has been treated as the ultimate DO. Since this type of question targets shallow processing, we refer to it, for the sake of convenience, as *shallow-targeted questions*. For the locally implausible condition (8b), the comprehension question was *Tamen mei yongbao gongren, dui ma?* 'They didn't hug the workers, right?' This question type tests

whether participants have knowledge of selectional restrictions in Chinese; it does not target shallow processing and hence is referred to as *non-shallow-targeted questions*.²⁸

For the other half of the experimental sentences, the comprehension questions were similar to the non-shallow-targeted questions, except that the main clause of each question was affirmative rather than negative (see the examples in Section 5.2.1.2). For clarity, we refer to those questions as *common verification questions*. Of this type of question, the locally plausible condition tests whether participants can fully recover from any initial misanalysis because correct answers to the questions require participants to understand the real DOs; the locally implausible condition tests participants' knowledge of selectional restrictions in Chinese.

In the following, we analyze the accuracy data from these different types of questions. Figure 5.2 shows the 3 participant groups' mean comprehension accuracy rates for the shallow-targeted questions ($k = 8$) vs. the non-shallow-targeted questions ($k = 8$). (The descriptive statistics are given in Appendix V.)

²⁸ The *shallow-targeted questions* were all positive questions and the *non-shallow-targeted questions* were all negative questions, but the intended correct answer to both question types was *bu dui* 'not right'. Since negative questions are known to be harder than positive questions, there is a potential confound between these two types of question. However, the data did not show such a confound. Actually, L2ers did better on the negative questions than on the positive questions (see Figure 5.2).

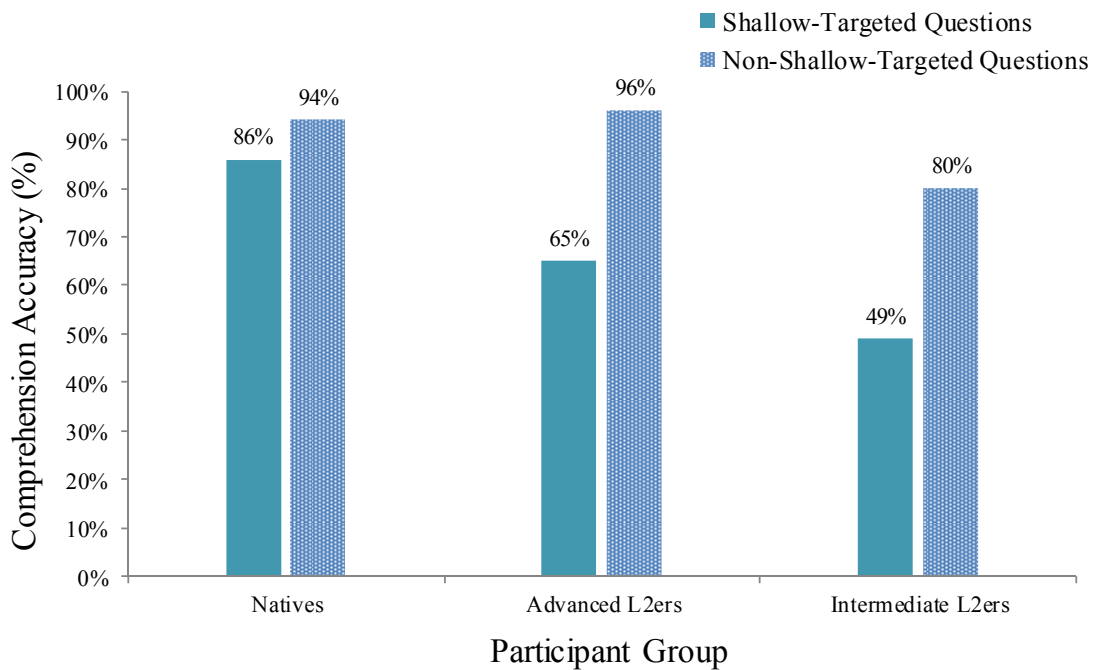


Figure 5.2. Mean Accuracy Rates for the Shallow-Targeted Questions vs. the Non-Shallow-Targeted Questions.

A 3×2 (Group [Native, Advanced L2, Intermediate L2] \times Question Type [shallow-targeted question, non-shallow-targeted question]) repeated measures ANOVA performed on the arcsine-transformed accuracy data yielded a main effect of group, $F(2, 64) = 19.12, p < .001, \eta_p^2 = .37$, a main effect of question type, $F(1, 64) = 39.55, p < .001, \eta_p^2 = .38$, and a Group \times Question Type interaction, $F(2, 64) = 4.55, p = .014, \eta_p^2 = .12$. SPSS Bonferroni corrected post hoc comparisons indicate that when the data were collapsed across conditions, the Native group and the Advanced L2 group did not differ significantly from each other ($p = .143$), and both of these groups had higher accuracy than the Intermediate L2 group ($ps \leq .001$). In the following, we will look at the accuracy data of these two types of questions separately.

For the shallow-targeted questions, the Native group had significantly higher accuracy than the L2 groups ($ps \leq .022$), and the Advanced L2 group had marginally higher accuracy than the Intermediate L2 group ($p = .098$). These results suggest that in reading the experimental sentences in the locally plausible condition (i.e., the condition in which the [-animate] nominal immediately following the verb is a plausible DO), L2 participants indeed practiced shallow processing to a larger extent than the Natives did,

but the L2ers' use of shallow processing strategies is modulated by proficiency in Chinese.

As for the non-shallow-targeted questions, the Native and Advanced L2 groups did not differ from each other ($p = 1.000$), and both groups had higher accuracy than the Intermediate L2 group ($ps \leq .007$), again suggesting a proficiency effect for the L2 groups. However, the participants' mean accuracy rates for this type of question are all very high (at least 80%), indicating that the participants could understand the experimental sentences very well (we will come back to this shortly).

To see whether each participant group had significantly different mean accuracy rates for the above two types of question, we performed separate paired-samples t tests for each group. The difference was marginal for the Native group, $t(29) = 1.85, p = .074$, and significant for both the Advanced L2 group, $t(17) = 4.53, p < .001$, and the Intermediate L2 group, $t(18) = 3.74, p = .001$. For all groups, then, the mean accuracy rates for the shallow-targeted questions were lower than for the non-shallow-targeted questions. This suggests that all groups, including the Natives, practiced shallow processing to some extent when the task demands encouraged them to do so.

Now let us turn to the common verification questions. Figure 5.3 presents the 3 groups' mean accuracy rates for this type of questions ($k = 8$ for each condition). (Appendix V provides the descriptive statistics.)

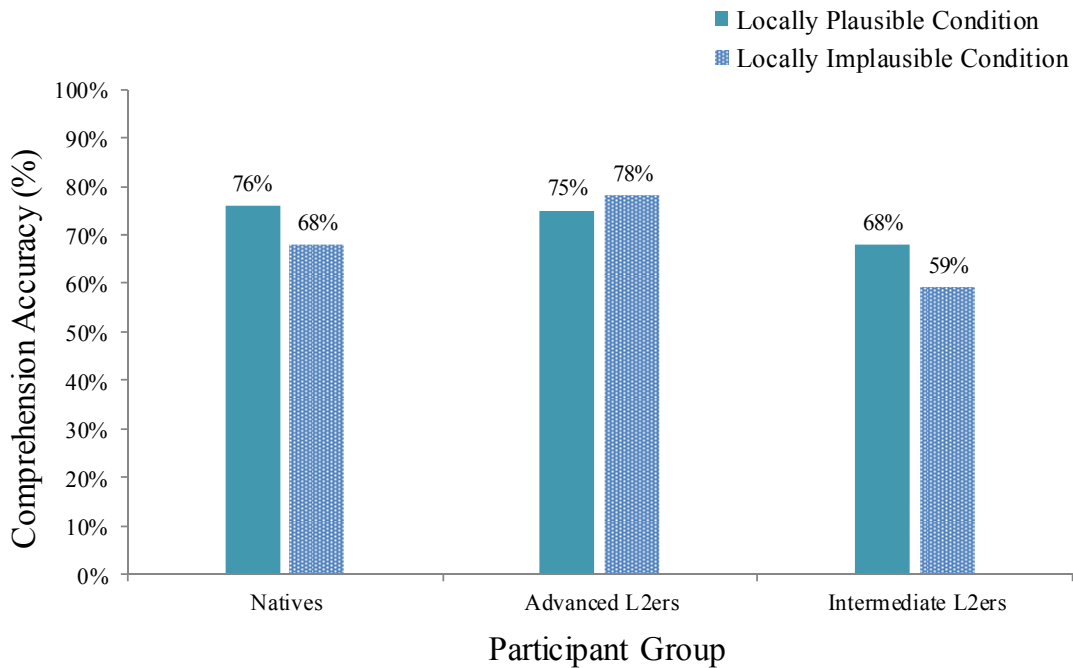


Figure 5.3. Mean Accuracy Rates for the Common Verification Questions.

A 3×2 (Group [Native, Advanced L2, Intermediate L2] \times Local Plausibility [locally plausible, locally implausible]) repeated measures ANOVA performed on the arcsine-transformed accuracy data did not reveal any significant main effect or interaction ($F_s < 1.79$, $ps > .17$). Notice that the participants' accuracy rates for the common verification questions were relatively low, especially when we consider the Natives' responses. Follow-up informal interviews suggested that participants had problems with the way the comprehension questions were phrased. For example, for the experimental sentences in (4), the comprehension questions were, respectively, *Tamen haipa yonghu, dui ma?* 'They feared the user, right?' (for [4a]) and *Tamen ganxie yonghu, dui ma?* 'They thanked the user, right?' (for [4b]). Some participants told the interviewer that they were not sure of what *yonghu* 'user' the question was talking about. They said that the question was too inclusive and could refer to a user of anything rather than just a user of a cellphone. The reason the comprehension questions were formulated in this way (i.e., by deleting the prenominal modifier, e.g., *shouji de* 'cellphone's' in the comprehension questions for [4]) was to reduce sentence length, but unexpectedly this introduced a problem for some participants. It appears that without the modifier, the lone

nominal in those comprehension questions can be interpreted as either definite or indefinite, although the experimental sentences did provide a context biasing toward a definite interpretation.

Despite the relatively low accuracy rates for the common verification questions, the Native and Advanced L2 groups' responses were all above chance level (50%) (one-sample t test performed on each group's accuracy data by condition: $p_s \leq .001$). The Intermediate L2 group's responses were also above chance in the locally plausible condition ($p = .015$), though not in the locally implausible condition ($p = .130$). Since the Natives and the Advanced L2ers performed both above chance and very similarly in the locally plausible condition (76% vs. 75%; see Figure 5.3), the results thus suggest that the Advanced L2ers were equally as able to recover from an initial misanalysis as the Natives were. Since the Intermediate L2ers' responses to the comprehension questions in the locally plausible condition were also above chance, this suggests that they, too, were able to recover from initial misanalysis, at least 68% of the time.²⁹

We return now to the non-shallow-targeted comprehension questions (see Figure 5.2). The follow-up interview also asked about these. For this type ($k = 8$), the interviewed participants did not raise any issues. What is special about the non-shallow-targeted questions is that they were all negated (with the negator *mei*), which eliminates the problem of ambiguity between definite and indefinite interpretations of the DO in those questions. To illustrate how this happens, let us look at the comprehension question for (8b): *Tamen mei yongbao gongren, dui ma?* 'They didn't hug (the) workers, right?' When *gongren* 'worker' is interpreted as indefinite, the question is suggesting that they didn't hug any type of workers; the factory's workers are no exception. This is obviously opposite to what the experimental sentence means (i.e., 'they hugged the factory's workers'). If *gongren* 'worker' is interpreted as definite, the question is suggesting that they didn't hug the factory's workers; this is also opposite to the meaning of the experimental sentence. Thus, regardless of whether *gongren* is interpreted as definite or indefinite, the correct answer to the comprehension question is always *bu dui* 'not right'. Since the non-shallow-targeted questions are not problematic,

²⁹ Both the Intermediate L2ers and the Natives performed somewhat better on the locally plausible condition than the locally implausible condition, but we will not explore this further since for neither group was the difference between the conditions significant.

the accuracy data from this type of question are arguably more representative of the participants' actual online comprehension than the accuracy data from the other types of comprehension questions. Recall that the participants' accuracy rates for this type of question are all high ($M \geq 80\%$; see Figure 5.2). Thus, the results suggest that all participant groups, including the Intermediate L2 group, could understand the experimental sentences quite well.

Furthermore, accurately answering a non-shallow-targeted question requires participants to possess adequate knowledge of the selectional restrictions of the relevant verbs (i.e., the verbs only allow a [+human] DO and disallow a [-animate] DO). In this respect, the accuracy data from this type of question can inform us as to whether the participants had knowledge of those verbs' selectional restrictions. Since the participants' accuracy rates from this type of questions are high ($\geq 80\%$ for each group), this reveals that for the most part, they indeed had the relevant knowledge at issue. Also, given that the verbs used in this experiment were all taken from textbooks for beginning-to-intermediate L2ers of Chinese and moreover that the frequencies of the verbs were matched across the two experimental conditions, we argue that the verbs in the experimental sentences with non-shallow-targeted questions should not necessarily be any more (or less) familiar to participants than the other verbs used in the experiment. In other words, the verbs used in the experiment should be viewed as coming from the same pool in terms of participants' verb familiarity. Hence, we infer that the accuracy results from the non-shallow-targeted questions indicate that the participants knew the selectional restrictions of (at least most of) the verbs used in this experiment.

In conclusion, the accuracy data of this experiment demonstrate that the L2ers were often able to recover from initial misanalysis and that they had knowledge of the selectional restrictions of most of the verbs used in this experiment.

5.2.2.2 Reading Times

Following the data-trimming procedure described in Chapter 3, we trimmed the raw RTs before conducting the RT data analyses so as to screen outlier responses. First, RTs longer than 2,500 ms were replaced with 2,500 ms; then, for each participant, the RTs that were 2 standard deviations (*SDs*) or more away from the participant's RT mean

were adjusted to that cutoff. Altogether, data trimming affected 5.7% of the Natives' data, 6.7% of the Advanced L2ers' data, and 10.2% of the Intermediate L2ers' data. (The descriptive statistics of the trimmed raw RT means for the experimental sentences are given in Appendix W.)

For the RT data analyses, this study only included trials on which participants answered the comprehension questions correctly. We removed the incorrect trials from our data analyses because there is a possibility that incorrect responses might have resulted from participants' lack of relevant knowledge (i.e., of the selectional restrictions of particular verbs). Since incorrect trials accounted for a substantial percentage of all the trials (Natives: 19%; Advanced L2ers: 22%; Intermediate L2ers: 36%), it is necessary to check whether removing the corresponding RT data might affect the final RT patterns. Appendix X provides a comparison between the results based on all trials and the results based on correct trials only. As is turned out, the RT patterns basically remain the same after the incorrect trials were removed.

In the following, we thus report the results from correct trials only.

Figures 5.4–5.6 present the 3 participant groups' region-by-region RT profiles. (For the descriptive statistics, see Appendix W.)

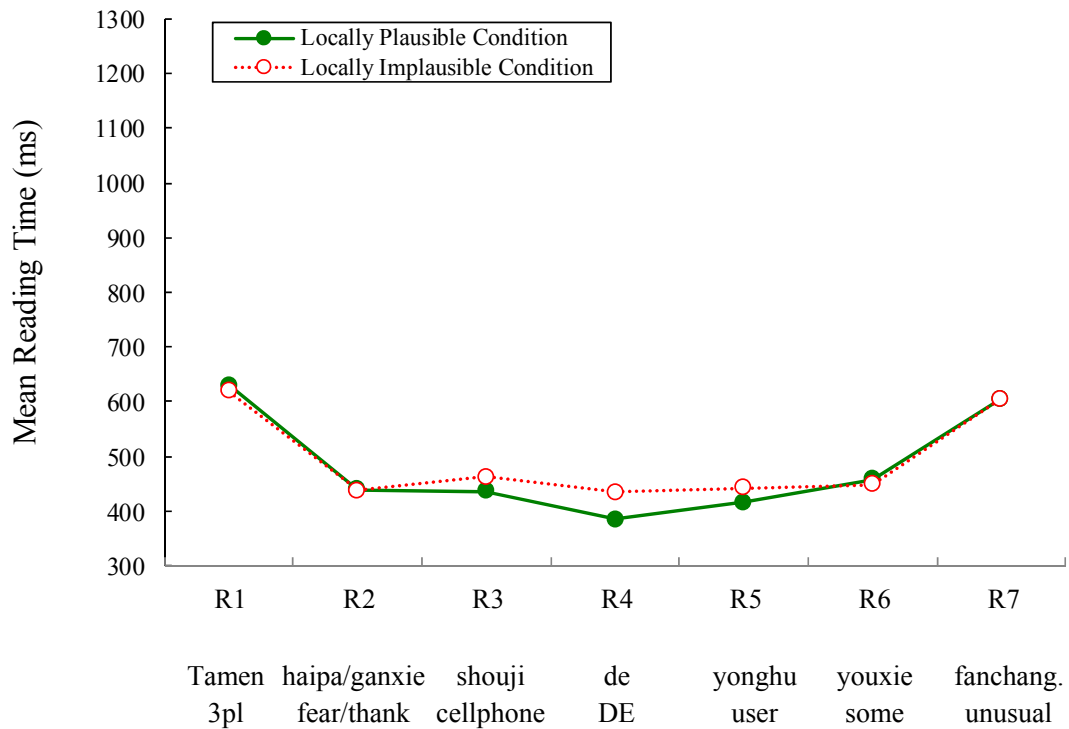


Figure 5.4. Natives' Mean RT (msec) Profile of Experiment 5.1.

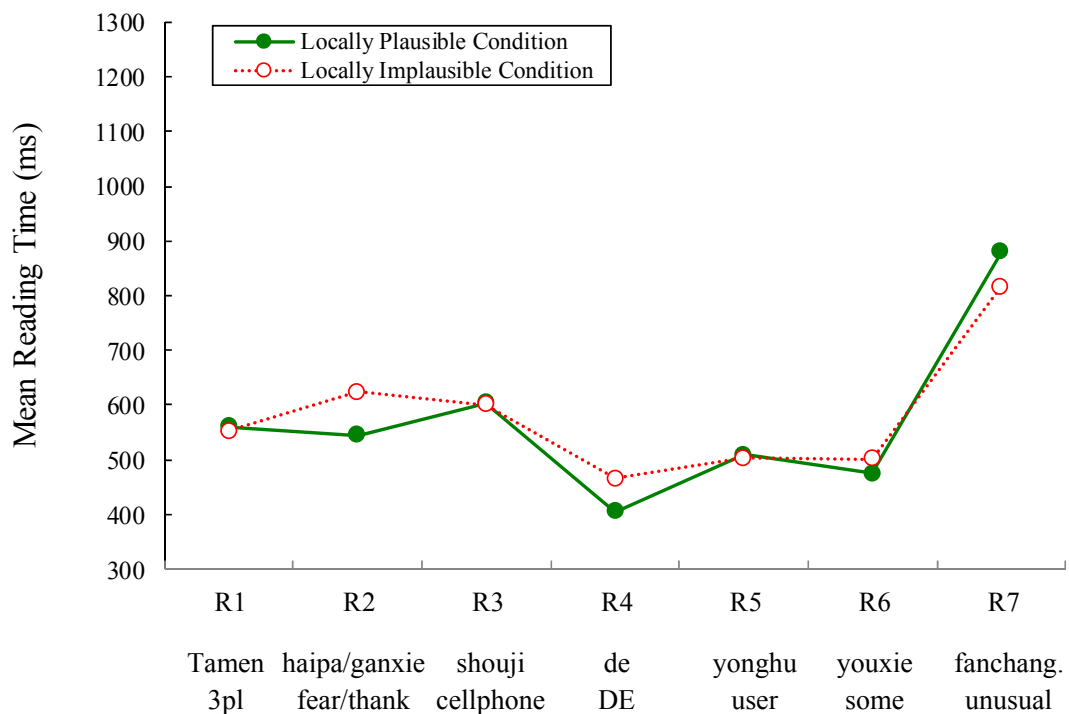


Figure 5.5. Advanced L2ers' Mean RT (msec) Profile of Experiment 5.1.

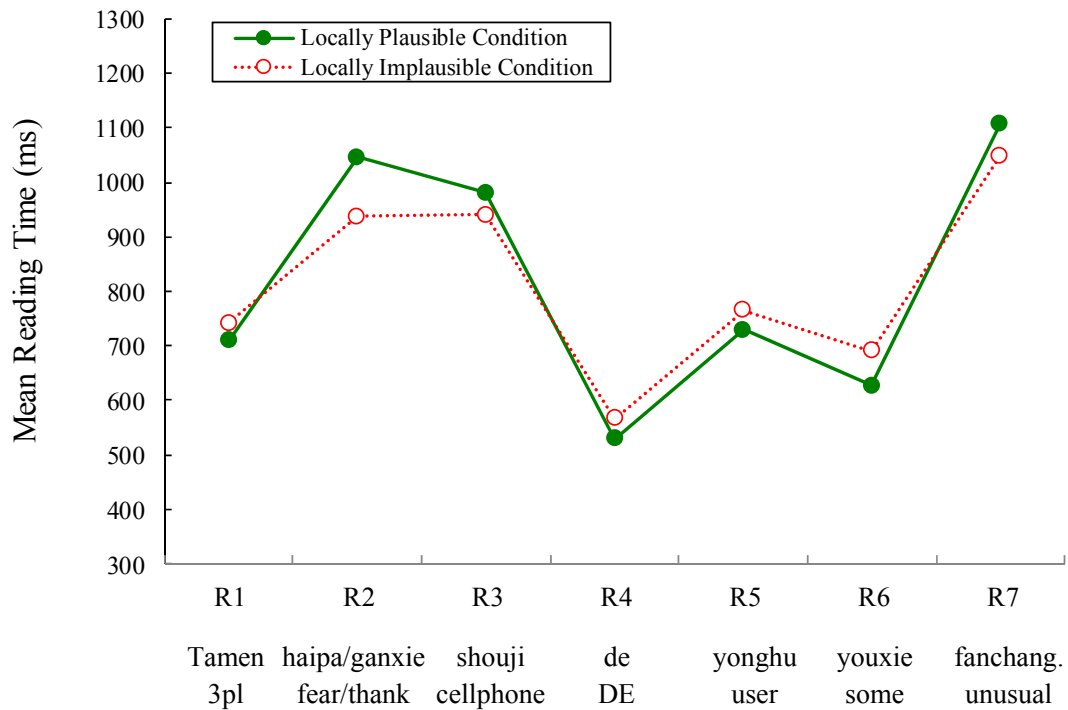


Figure 5.6. Intermediate L2ers’ Mean RT (msec) Profile of Experiment 5.1.

Visual inspection of these figures reveals that the Natives read slightly faster at Region 3 (i.e., the critical *shouji* ‘cellphone’ region) in the locally plausible condition than in the locally implausible condition. All 3 participant groups read faster at Region 4 (i.e., the post-critical *de* region) in the locally plausible condition than in the locally implausible condition, especially the Native and Advanced L2 groups. However, none of the participant groups demonstrated an obvious reversed RT pattern at Regions 5 and 6. Since the RT data were based only on participants’ correctly answered trials—hence trials for which reanalysis was successful—the data showed that none of the participant groups had heightened reanalysis effects in the locally plausible condition. Notice that the L2 groups did not show elevated reanalysis difficulty, contrary to what was predicted following Roberts and Felser (2011). We will come back to this in the next section.

To test the statistical significance of these patterns, we conducted separate paired-samples *t* tests (two-tailed) for each group at each of these regions of interest. The tests were run on both participant (t_1) and item (t_2) means. Following the standard practice of self-paced reading studies, we set the alpha level at .05 for all tests.

For the Native group, the difference between the locally plausible and locally implausible conditions was nonsignificant at Region 3 ($ts < 1.63, ps > .11$), but it reached significance at Region 4, $t_1(29) = 2.97, p = .006$; $t_2(15) = 2.77, p = .014$, revealing an effect of local plausibility. As for Regions 5 and 6, there was no significant difference between the two conditions ($ts < 1.59, ps > .13$), showing there was no reanalysis difficulty for the Natives.

As for the Advanced L2 group, the difference between the two conditions was also nonsignificant at Region 3 ($ts < 1.14, ps > .27$), but it became significant at Region 4 in the participant analysis, $t_1(17) = 3.17, p = .006$, and marginal in the item analysis, $t_2(15) = 1.98, p = .066$. At Regions 5 and 6, the difference between the conditions was nonsignificant ($ts \leq 1.13, ps \geq .28$). Like the Natives, the Advanced L2ers showed an effect of local plausibility but experienced no difficulty in reanalysis.

The Intermediate L2 group did not show any significant difference between the conditions at Regions 3–6 ($ts < 1.14, ps > .27$), demonstrating neither a local plausibility effect nor a reanalysis effect.

In sum, the data show the Natives and Advanced L2ers were sensitive to the local plausibilities, but the Intermediate L2ers were not. No group showed any reanalysis difficulty.

5.2.3 Discussion

Employing a CFSPR task, Experiment 5.1 examined how natives and L2ers process lexical-semantic information in online sentence comprehension in Chinese. The results show that the Natives and Advanced L2ers were sensitive to the temporary violations of selectional restrictions in that the participants' processing time was significantly elevated when the first nominal in the DO violated the selectional constraints of the verb. The results suggest that both natives and proficient L2ers can make immediate use of lexical-semantic information in Chinese sentence processing. These results are in line with those of previous studies suggesting incremental effects of lexical-semantic information in L1 sentence processing (e.g., S. Wang et al., 2013; Warren & McConnell, 2007; J. Yang et al., 2009) and in L2 sentence processing (e.g.,

Bowden et al., 2013; Hahne, 2001; Hahne & Friederici, 2001; Roberts & Felser, 2011; Williams, 2006; Williams et al., 2001).

In contrast to the Advanced L2ers, the Intermediate L2ers did not show any sensitivity to the selectional-restriction violations. The RT data demonstrate a proficiency effect in L2 lexical-semantic processing. The results do not lend support to proposals of a (necessary) qualitative difference between L1 and L2 processing of semantic information (e.g., Newman et al., 2012) in that proficiency increases can and do reshape L2 (lexical-)semantic processing.

In addition, the experiment did not produce evidence of more robust or prolonged effects of local plausibility for L2ers than for natives, and L2ers did not experience increased reanalysis difficulty (contra the explicit predictions of Roberts & Felser, 2011). In fact, only the Advanced L2ers patterned with the Natives, and they showed local plausibility effects only at Region 4 (see Figures 5.4–5.6). The Intermediate L2 group did not show any reliable sensitivity to the local plausibility manipulation at Regions 3 or 4, despite the fact that this group achieved 80% mean accuracy for the non-shallow-targeted comprehension questions (i.e., the type of question arguably reflecting participants' actual online comprehension and knowledge of selectional restrictions—refer to Figure 5.2). Since our data analysis here only included trials for which participants answered the comprehension questions correctly, this means that our analysis was based exclusively on the trials for which participants have knowledge of the relevant verbs' selectional restrictions and for which their reanalysis was successful. In addition, no group showed any significant RT difference between the experimental conditions at or following Region 5—the disambiguating region (see Figures 5.4–5.6). Echoing the accuracy results that indicate that the Advanced L2ers, just like the Natives, were able to recover from initial misanalyses (see Figure 5.3), the RT results here suggest that the Natives and Advanced L2ers were not so different from each other in their ability to recover from initial misanalyses.

As for why none of the groups showed reanalysis effects at the disambiguating regions, one very likely explanation is that the reanalysis proceeded easily and quickly for the participants, as a result of which it did not lead to any measurable effect online. First, the reanalysis does not involve substantial structural reconfiguration. It occurs

within the same “thematic domain” (i.e., from the DO position to the specifier position of the DO). Second, the ambiguous NP and the disambiguating elements (immediately after the prenominal modifier marker *de*) are not far away from each other, and in this case participants should be able to revise their initial misanalysis very quickly. Since the L2ers did not show reanalysis difficulty at the disambiguating regions, this suggests that they were able to use the available structural information to compute the correct structural representations.

Overall, the results of Experiment 5.1 challenge the SSH (as put forward by Roberts & Felser, 2011) in that its predictions about L2ers’ over-dependence on lexical-semantic information in L2 sentence processing were not borne out.

Since Experiment 5.1 used a CFSPR task, which might encourage participants to overuse lexical-semantic information in sentence processing, it is necessary to test whether the findings from this experiment can be extended to task conditions that engage participants in more detailed structural processing. This was examined in Experiment 5.2.

5.3 Experiment 5.2: Acceptability-Judgment Self-Paced Reading

This experiment compares L1 and L2 use of lexical-semantic information in Chinese sentence processing in an AJSPR task in order to see whether the findings from Experiment 5.1 can be generalized to this structure-focused task. In AJSPR, each sentence is followed by a prompt that asks participants to decide whether the sentence is acceptable in terms of its well-formedness. Since this task encourages fully specified structural processing, participants should be less likely to use shallow processing strategies and thus should be more successful in recovering from initial misanalyses than in a CFSPR task. By contrast, the SSH predicts that L2ers should not be able to benefit from this task in that their processing is primarily based on nonstructural information. If the SSH holds, the L2 reanalysis difficulty predicted by Roberts and Felser (2011) (but not found in Experiment 5.1) should not be alleviated in this task. If there is reanalysis difficulty, it will be exhibited as longer RTs at Regions 5 and 6 in the locally plausible condition than in the locally implausible condition. Since reanalysis difficulty means that participants cannot easily recover from initial misanalyses in the locally plausible

condition, it should also lead to lower judgment accuracy in the locally plausible condition than in the locally implausible condition.

5.3.1 Method

5.3.1.1 Participants

Thirty adult natives and 37 adult L2ers of Chinese (18 L1-English and 19 L1-Japanese) participated in Experiment 5.2, and none participated in Experiment 5.1. The participants were paid a fee for taking part. One Japanese participant was excluded due to the unusually poor performance on the special type of filler ($M_{\text{error rate}} = 62.5\%$), which was designed for assessing attentiveness in doing the self-paced reading task (for a description of this type of filler, see Chapter 3).

As in Experiment 5.1, L2ers were divided into 2 proficiency groups (Advanced and Intermediate) according to their proficiency scores on the quasi-C-test. The cutoff point was the same as in Experiment 5.1 (i.e., 22). The Advanced L2 group consisted of 9 L1-English and 9 L1-Japanese L2ers, and they all scored 22 and above. The Intermediate L2 group comprised 12 L1-English and 6 L1-Japanese L2ers, who all scored less than 22. Table 5.6 provides the Chinese background information of all 3 participant groups.

Table 5.6. *Chinese Background Information of the Participants in Experiment 5.2*

Group	Years of learning Chinese		Years of residence in China/Taiwan		Chinese proficiency scores (Max = 50)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Natives ($n = 30$)	N/A	N/A	N/A	N/A	45.9	2.0
Advanced L2ers ($n = 18$)	3.4	2.7	2.6	4.2	31.5	9.5
Intermediate L2ers ($n = 18$)	2.8	2.1	0.9	1.6	13.8	5.8

A one-way ANOVA performed on the mean proficiency scores showed that the 3 groups differed significantly in their Chinese proficiency, $F(2, 63) = 166.16, p < .001$. SPSS Bonferroni corrected post hoc tests revealed that the Natives were better than both L2 groups ($ps < .001$) and that the Advanced L2 group was better than the Intermediate L2 group ($p < .001$).

The L2ers in Experiments 5.1 and 5.2 were comparable in Chinese proficiency, as revealed by their mean Chinese proficiency test scores (the Advanced L2 groups:

$M_{\text{difference}} = 3.6$, $t(34) = 1.18$, $p = .248$; the Intermediate L2 groups: $M_{\text{difference}} = 0.6$, $t(35) = 0.33$, $p = .742$).

5.3.1.2 Materials

Experiment 5.2 used the same experimental items from Experiment 5.1. However, Experiment 5.2 differs from Experiment 5.1 in that participants in this experiment were required to make an acceptability judgment at the end of each experimental sentence instead of answering a comprehension question. The experimental sentences were combined with 72 filler sentences, of which 44 were grammatical and 28 ungrammatical. The ungrammatical fillers are important in that they enabled the acceptability-judgment task to work properly. The fillers involved different constructions so as to prevent participants from discerning what was actually being tested. Of the ungrammatical fillers, one type ($k = 8$) was used specifically for evaluating participants' attentiveness in doing the self-paced reading task (see the description of this type of filler in Chapter 3). Notice that the ratio of grammatical to ungrammatical fillers was not even. Our main consideration in this design was that some experimental items and fillers alike contain structural ambiguities and unusual propositional content, which we reasoned might cause low acceptability (see the discussion in Section 5.3.2.1). If we had used an equal number of ungrammatical fillers and grammatical fillers, participants might have formed an impression that most of the sentences were unacceptable, and this should be avoided.

5.3.1.3 Procedure

Experiment 5.2 used an AJSPR task, in which participants were asked to judge the acceptability of each sentence they read (for a description of the AJSPR task, see Chapter 3). Other than this, the procedure was identical to that of Experiment 5.1. To trigger an acceptability judgment, the experiment always used this prompt: *Juzi tongshun ma?* 'Is this sentence acceptable?' (for a discussion of proper instructions for acceptability-judgment tasks in Chinese, see Lu et al., 2000). Participants pressed the designated *Tongshun* ('well-formed, acceptable') key or the *Bu tongshun* ('not well-formed, not acceptable') key on the computer keyboard to make the judgment.

After completing the self-paced reading task, participants were asked to fill out a language background questionnaire (see Appendix E) and take the Chinese proficiency test (see Appendix F).

5.3.2 Results

As in Experiment 5.1, the data analysis was performed on both acceptability judgments and online RTs.

5.3.2.1 Acceptability Judgments

All groups achieved very good acceptability-judgment results for the 8 special fillers designed for assessing participants' attentiveness in performing the self-paced reading task. As can be seen in Table 5.7, the accuracy rates for this filler type were at least 80%, suggesting that all groups were attentive in performing the self-paced reading task. In comparison, each group's mean accuracy for the other fillers was generally lower than for the special type. Overall, we can see from Table 5.7 that for all types of filler, the acceptability-judgment accuracy rates were modulated by participants' Chinese proficiency.

Table 5.7. Mean Acceptability-Judgment Accuracy Rates for the Fillers in Experiment 5.2

Group	Type		
	Special (ungrammatical; $k = 8$)	Other (grammatical; $k = 44$)	Other (ungrammatical; $k = 20$)
Natives ($n = 30$)	.93 (.09)	.80 (.12)	.85 (.14)
Advanced L2ers ($n = 18$)	.84 (.11)	.82 (.09)	.64 (.17)
Intermediate L2ers ($n = 18$)	.80 (.12)	.71 (.12)	.46 (.18)

Note. Standard deviations are in parentheses.

The participants' mean acceptance rates (i.e., accuracy rates) for the experimental sentences are presented in Figure 5.7 (for the descriptive statistics, see Appendix Y).

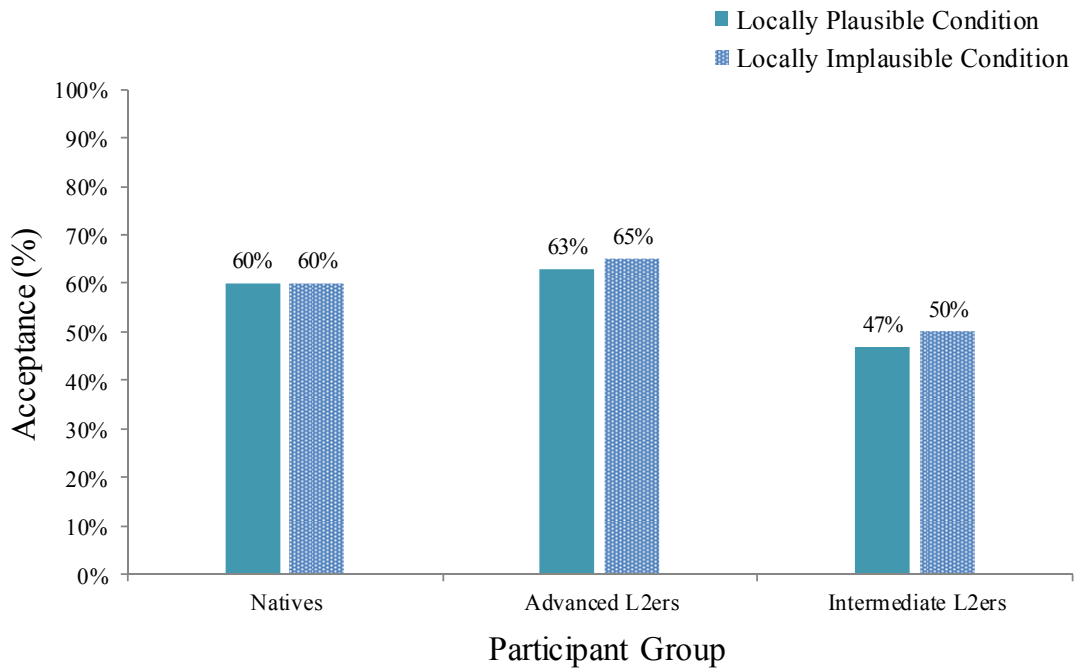


Figure 5.7. Mean Acceptance Rates for the Experimental Sentences of Experiment 5.2.

A 3×2 (Group [Native, Advanced L2, Intermediate L2] \times Local Plausibility [locally plausible, locally implausible]) repeated measures ANOVA performed on the arcsine-transformed data yielded only a main effect of group, $F(2, 63) = 4.07, p = .022, \eta_p^2 = .12$. SPSS Bonferroni corrected post hoc tests showed that the Intermediate L2 group differed from both the Native ($p = .027$) and Advanced L2 groups ($p = .009$). The Intermediate L2ers' acceptance rates for the experimental sentences were lower than those of the Native and Advanced L2 groups. No main effect of local plausibility or interaction effect of Group \times Local Plausibility was observed ($F_s < .23, p_s > .63$).

Overall, the participants' accuracy rates for the experimental sentences were not high (between 47% and 65%). Previous studies (e.g., Juffs, 1998b; Juffs & Harrington, 1996) indicate that online acceptability-judgment tasks involving GP sentences tend to produce lower acceptance rates. For example, in the Juffs and Harrington (1996) study, the mean acceptance rate for sentences such as (9) was as low as 35% by adult native English speakers and 22% by L2ers of English.

(9) Sam warned the student cheated on the exam.

Also, the experimental sentences in the current study are structurally complex, containing a sentential subject, which could be another source leading to the relatively low acceptance rates. Previous studies (e.g., Omaki & Schulz, 2011, p. 575) suggest that structurally complex sentences tend to receive acceptance ratings below the normal level. In addition, as shown in another study of this dissertation research (see Chapter 6), participants doing an AJSPR task may base their judgments not only on the structural well-formedness but also on nonstructural factors such as propositional content and pragmatic appropriateness (cf. Hsu & Hsieh, 2013). These multiple factors may render the acceptability-judgment results noticeably lower than what is typically expected.

Nevertheless, one-sample *t* tests performed on the acceptability-judgment data indicate that the Natives' and the Advanced L2ers' accuracy rates for experimental sentences were all above the 50% chance level, in both conditions ($ps \leq .04$), although the Intermediate L2ers' accuracy rates were not different from chance ($ps > .50$). Since the Native and Advanced L2 groups performed so similarly across the two experimental conditions (respectively, no difference vs. a 2% difference) and, in particular, since the 2 groups' accuracy rates in the locally plausible condition were very similar (60% vs. 63%), the results suggest that the Natives and the Advanced L2ers were equally able to recover from initial misanalysis in this structure-focused task. As for the Intermediate L2ers, their mean accuracy rates showed a nonsignificant (3%) difference between the two conditions ($p = .610$). And since these accuracy rates were not above chance level, we cannot make unequivocal inferences about whether the Intermediate L2ers were able to recover from initial misanalysis. We will return to this issue in Section 5.4 (General Discussion).

5.3.2.2 Reading Times

In this experiment, RT data from the trials on which participants responded incorrectly were not excluded in the data analysis. As mentioned in the above section, several factors might have led to participants' inaccurate acceptability-judgments, and those factors might not have much to do with participants' inattentiveness in performing the task or their ignorance of selectional restrictions in Chinese (especially in light of the Natives' data).

As with Experiment 5.1, the analyses of the RT data in this experiment focused on Regions 3–6. The same data-trimming procedure to raw RTs as in Experiment 5.1 was applied in Experiment 5.2, affecting 8.2% of the Natives’ data, 12.2% of the Advanced L2ers’ data, and 15.5% of the Intermediate L2ers’ data.

The mean RT profiles by region for each participant group are shown in Figures 5.8–5.10. (The descriptive statistics are in Appendix Z.)

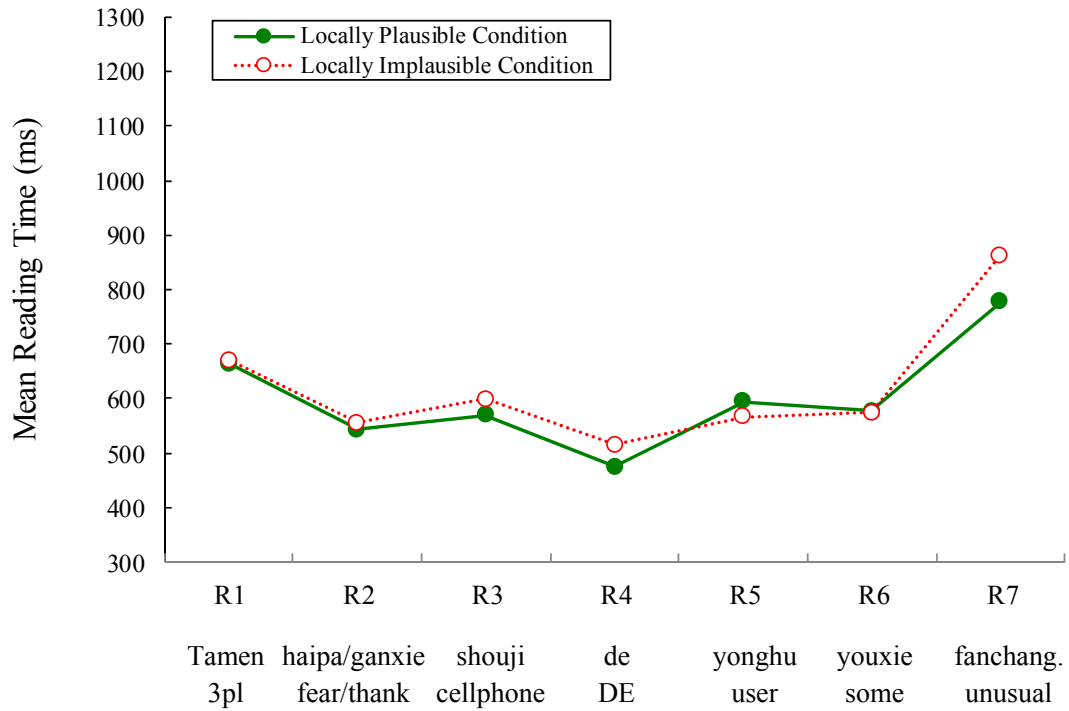


Figure 5.8. Natives’ Mean RT (msec) Profile of Experiment 5.2.

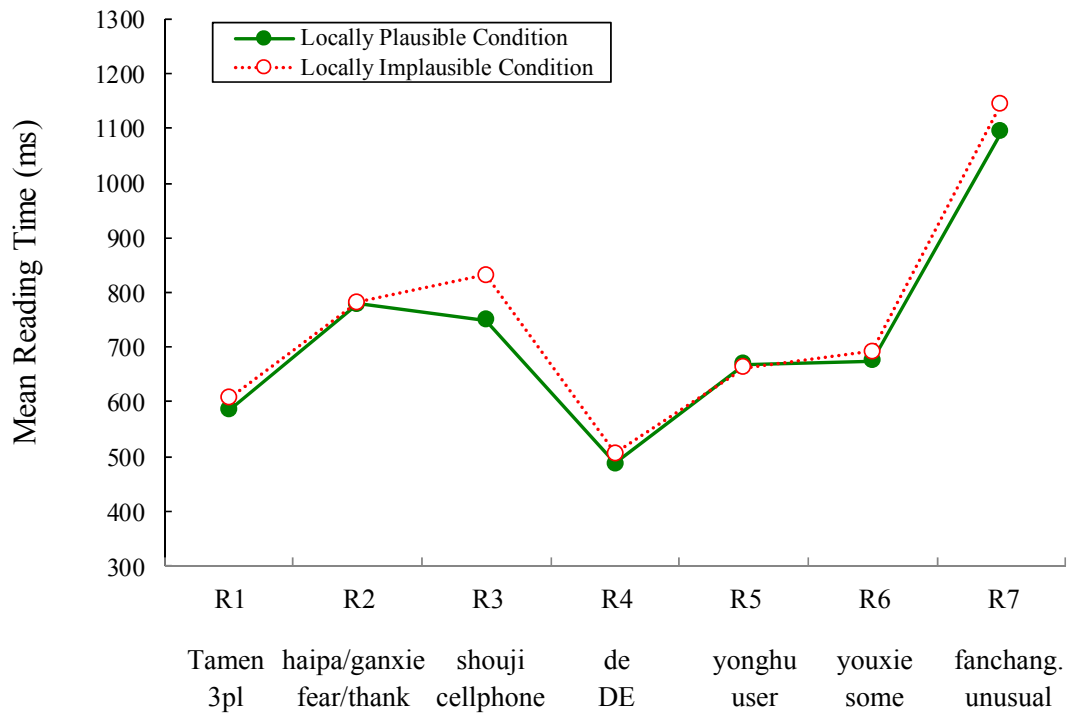


Figure 5.9. Advanced L2ers' Mean RT (msec) Profile of Experiment 5.2.

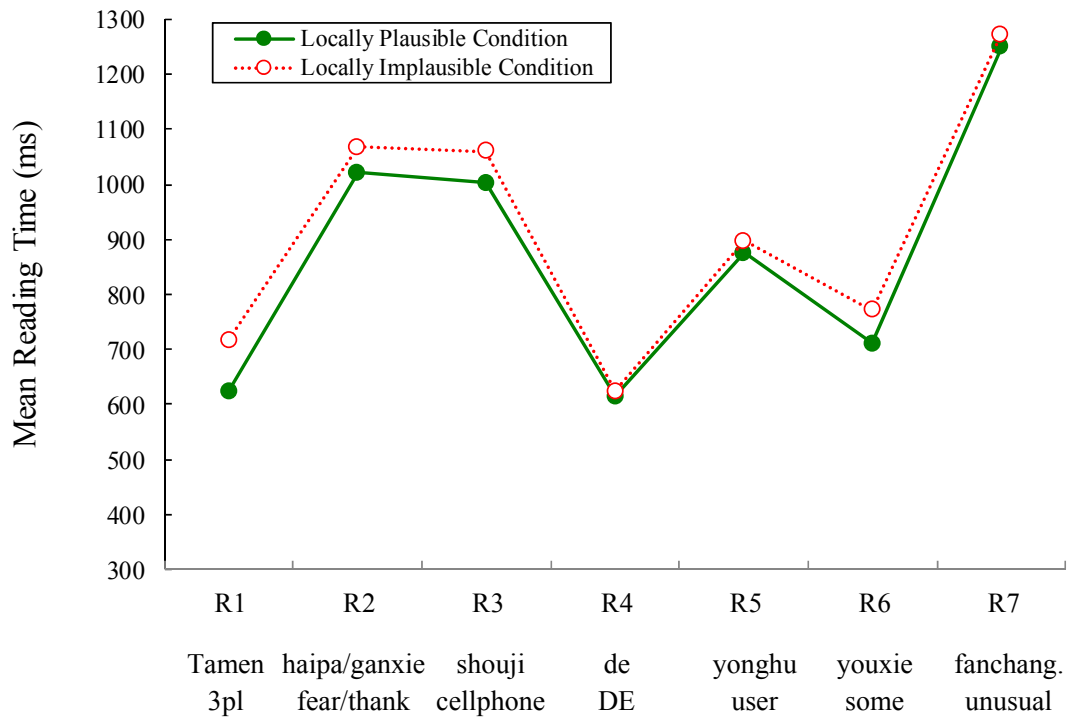


Figure 5.10. Intermediate L2ers' Mean RT (msec) Profile of Experiment 5.2.

By visually inspecting these figures, we can see that both the Natives and Advanced L2ers read Region 3 faster in the locally plausible condition than in the locally implausible condition. The Intermediate L2ers also displayed such a tendency at Region 3, but the tendency started earlier, from Region 1, for which there is no explanation. At Region 4, only the Natives showed a relatively large RT difference between the conditions. At Region 5, the Natives alone reversed their earlier RT pattern, spending slightly longer in the locally plausible condition than the locally implausible condition. No such a reversed RT pattern was observable at Region 6 for any group.

To test the statistical significance of these RT patterns, we ran separate paired-samples *t* tests (two-tailed) for each participant group on both participant and item means, as in Experiment 5.1.

The Native group's RT difference between the conditions at Region 3 showed a trend in the predicted direction but did not reach statistical significance, $t_1(29) = 1.44$, $p = .162$; $t_2(15) = 1.21$, $p = .246$; but the difference became significant at Region 4, $t_1(29) = 3.42$, $p = .002$; $t_2(15) = 2.87$, $p = .012$, yielding a reliable local plausibility effect. At Regions 5 and 6, the RT differences between the conditions were not significant ($ts < 1.36$, $ps > .19$). The Natives did not reveal a reanalysis effect at or after the disambiguating region.

For the Advanced L2ers, the local plausibility effect was significant at Region 3 in the participant analysis, $t_1(17) = 2.81$, $p = .012$, but not in the item analysis, $t_2(15) = 1.65$, $p = .120$. The effect was nonsignificant at Region 4 ($ts < 1.01$, $ps > .30$). At Regions 5 and 6, there was no reanalysis effect ($ts < 0.40$, $ps > .70$). Briefly, the Advanced L2ers showed some local plausibility effect at the critical region, but the effect was not robust; and they did not show any reanalysis difficulty at or after the disambiguating region.

As for the Intermediate L2 group, there was no local plausibility effect at Regions 3 or 4 ($ts < 0.96$, $ps > .35$), nor was there a reanalysis effect at Regions 5 or 6 ($ts < 1.22$, $ps > .24$).

To sum up, the Native and Advanced L2 groups were sensitive to the experimental manipulation of local (im)plausibilities, with the Native group showing more robust effects than the Advanced L2 group. The Intermediate L2ers did not indicate

any reliable effect of local plausibility. None of the groups demonstrated reanalysis effects.

5.3.3 Discussion

Using a structure-focused AJSPR task, Experiment 5.2 produced local plausibility effects for both the Native and Advanced L2 groups, confirming once again that natives and proficient L2ers of Chinese can make rapid use of lexical-semantic information in Chinese sentence processing. In addition, the experiment also demonstrates a proficiency effect in L2 semantic processing since only the Advanced L2ers, but not the Intermediate L2ers, revealed some sensitivity to the local plausibility manipulation. Furthermore, none of the participant groups exhibited effects of reanalysis at the disambiguating regions.

The results of Experiment 5.2 did not provide evidence of more “pronounced” local plausibility effects for L2ers than for Natives. The Natives’ local plausibility effects appeared numerically large at Region 3 and reached significance at Region 4 both by participants and by items. In contrast, the Advanced L2ers only showed some local plausibility effect at Region 3, which was significant only in the participant analysis. The intermediate L2ers did not show any plausibility effect at either Region 3 or Region 4. It is obvious that L2ers use no more of lexical-semantic information than natives do even when the processing task is structure-focused. These results can hardly be reconciled with the Roberts & Felser (2011) predictions stemming from the SSH.

As in Experiment 5.1, the (Advanced) L2ers in this experiment did not experience increased reanalysis difficulty. From the RT data we can see that local plausibilities at Region 3 affected the Native and Advanced L2 groups’ processing, but the 2 groups showed no reanalysis difficulty at Regions 5 or 6. Rather, they appeared similar in being able to recover from initial misanalysis. This was corroborated by the acceptability-judgment data. From Figure 5.7, we can see that the Native and Advanced L2 groups’ mean accuracy rates are comparable to each other as well as between conditions (Natives: 60% vs. 60%; Advanced L2ers: 63% vs. 65%). If L2ers have increased reanalysis difficulty, they should be trapped by locally plausible DOs and should consequently have significantly lower accuracy rates in the locally plausible condition than in the locally implausible condition—contra the actual results.

Overall, Experiment 5.2 validates the findings of Experiment 5.1 and confirms that L2ers do not necessarily make more use of lexical-semantic information than natives do in online sentence processing.

5.4 General Discussion

In two experiments, this study investigated L2 processing of lexical-semantic information in Chinese sentence interpretation in order to assess whether adult L2ers, as the SSH suggests, over-rely on lexical-semantic information in processing L2 sentences. The experiments took advantage of the selectional requirements of Chinese verbs to manipulate local semantic plausibilities so as to observe whether participants were sensitive to violations of selectional constraints. The results demonstrate that both the Natives and the Advanced L2ers, though not yet the Intermediate L2ers, were sensitive to the local plausibility manipulation, suggesting that natives and L2ers alike can make rapid use of lexical-semantic information online and that their semantic processing is incremental. Furthermore, the present study provides no evidence for L2ers' over-dependence on lexical-semantic information in L2 sentence processing. Against what Roberts and Felser (2011) predicted about L2 processing, in neither experiment do L2ers yield more robust or prolonged local plausibility effects than natives do, and neither experiment observes increased reanalysis difficulty for L2ers (although the null effects do not allow us to make a strong claim about this issue). Instead, both experiments demonstrate that L2 lexical-semantic processing is shaped by L2ers' target language proficiency. Since the Advanced L2ers in both experiments were able to recover from initial misanalysis as well as the natives did, the results suggest that adult L2ers can (come to) use available structural information in L2 sentence processing.

The current study used two self-paced reading tasks (CFSPR in Experiment 5.1 and AJSPR in Experiment 5.2). This practice proves to be informative. One interesting finding related to the use of the two tasks is that the local plausibility effects started earlier in AJSPR than in CFSPR. For both the Natives and the Advanced L2ers, the local plausibility effects appeared at Region 4 in Experiment 5.1 but at Region 3 in Experiment 5.2 (although for the Natives, the effects only became reliable at Region 4). The earlier semantic effects in AJSPR than in CFSPR suggest that participants do not

ignore semantic information in the structure-focused AJSPR task; rather, they compute structural and semantic details in an immediate and incremental fashion.

One issue that can be raised about this study is whether the Intermediate L2 participants, who failed to show any online effect of local plausibility, actually knew the selectional restrictions of the verbs. This is particularly worrisome in Experiment 5.2, where their mean acceptability-judgment accuracy rate was 50% in the locally implausible condition even dropped (numerically) below chance level in the locally plausible condition (47%; see Figure 5.7). Since this study did not include an independent offline task to assess participants' knowledge of the target semantic relationships, it is difficult to rule out the possibility that the Intermediate L2ers' lack of sensitivity to the selectional restrictions was due to a lack of the relevant knowledge. In spite of this possibility, we contend that the Intermediate L2ers actually had the relevant knowledge.

Recall that in Experiment 5.1, the Intermediate L2ers' accuracy rate for the non-shallow-targeted questions was 80%, which is quite high. If they did not have the relevant knowledge, they could not have understood the sentences and hence could not have achieved such high comprehension accuracy. In addition, our analysis of the online data in Experiment 5.1 was only based on participants' correctly-answered trials, and in this way we made certain that the online processing patterns derived only from trials for which participants demonstrated both the relevant knowledge (i.e., knowledge of selectional restrictions) and successful recovery from misanalyses. Since the Intermediate L2 groups in the two experiments showed very similar processing profiles and also since the Intermediate L2ers in the two experiments came from the same participant pool with comparable Chinese proficiency (see Section 5.3.1.1), we can safely assume that they were also comparable in terms of having the relevant knowledge. As for the Intermediate L2 group's low acceptability-judgment accuracy in Experiment 5.2, it is not too unexpected. As discussed in Section 5.3.2.1, online acceptability judgments tend to elicit low accuracy rates when sentences are complex in structure and/or contain GPs. The Intermediate L2ers' 47%/50% acceptance rates do not seem so low in comparison with the Natives' 60%/60% acceptance rates. Still, the worry about the Intermediate L2ers' knowledge at stake (selection constraints of the study's verbs) cannot be totally dispelled without an independent task assessing it. This should be remedied in future studies.

5.5 Conclusion

This study examined how natives and L2ers apply lexical-semantic knowledge in online sentence processing in two self-paced reading tasks (CFSPR and AJSPR). By testing whether participants can detect online violations of selectional restrictions in Chinese, the study found that both the Natives and the Advanced L2ers can rapidly use lexical-semantic information online in an incremental fashion and that L2ers do not over-use lexical-semantic information (at the expense of structural information) any more than natives do in online sentence processing. The results of this study pose a challenge to the SSH, which claims that L2 processing over-relies on lexical-semantic information. Moreover, this study does not provide evidence for a necessarily qualitative difference between L1 and L2 semantic processing in that L2 lexical-semantic processing is found to be modulated by target language proficiency. Although selectional restrictions may involve subtle lexical-semantic information and can be a potential area of acquisition/processing difficulty, L2ers can eventually acquire the knowledge and use it in online processing in a native-like fashion as they become sufficiently proficient in the target language.

CHAPTER 6

PROCESSING CONTEXTUAL INFORMATION IN CHINESE

As pointed out by Phillips and Ehrenhofer (in press),

the evidence for or against SSH generally involves demonstrations of L2ers' sensitivity or insensitivity to structural cues that native speakers attend to in parsing, rather than evidence that L2ers are more strongly guided by lexical and conceptual associations.

To further explore the latter type of evidence, this chapter looks at native ("L1") and nonnative ("L2") use of contextual information in Chinese sentence processing. It reports a study that was conducted to investigate how extrasentential discourse context influences L1 and L2 processing of null objects (object omissions) in Chinese. The main purpose of this study is to address the question of whether L2 learners ("L2ers") necessarily depend more on contextual information in sentence processing than L1 speakers ("natives") do, as the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, 2006b, 2006c) maintains.

The following sections of this chapter are arranged as follows: Section 6.1 reviews studies on L1 processing of discourse context in general, and Section 6.2 in Chinese sentence processing in particular. Section 6.3 turns to previous work on L2 processing of discourse-context information. Sections 6.4 and 6.5 report on two experiments that were designed to investigate how discourse context mediates L1 and L2 processing of null objects in Chinese. Section 6.6 offers a general discussion of the study. Section 6.7 presents the conclusion.

6.1 L1 Processing of Contextual Information

L1 research indicates that contextual information influences parsing decisions (e.g., Altmann, 1998; Gibson & Pearlmutter, 1998). In ambiguity resolutions, for example, contextual information can make one analysis more plausible than others or eliminate alternative analyses (for a review, see MacDonald et al., 1994). Researchers

generally agree that parsing processes are subject to the influence of contextual information, but they disagree on the nature and timing of the influence. Some researchers view contextual information as secondary compared to syntactic information (e.g., Clifton et al., 2003; Ferreira & Clifton, 1986; Frazier, 1987; Frazier & Fodor, 1978). Adopting a serial, modular approach to sentence processing, they argue that syntactic information guides the initial process of sentence interpretation and that nonsyntactic sources of information such as discourse context affect later stages of sentence processing such as reanalysis and semantic integration. One extreme version of the modular approach, known as *depth-first* models or the *garden-path* theory (Frazier, 1995), proposes that the parser constructs only a single syntactic analysis initially and subsequently revises it if it conflicts with other sources of information. Other researchers reject the privileged status of syntax in sentence processing. They view syntax as one of multiple constraints that interact with one another in the process of sentence interpretation (e.g., Bates & MacWhinney, 1989; MacDonald et al., 1994; Marslen-Wilson, 1975; Tanenhaus & Trueswell, 1995). This constraint-based approach contends that various sources of information (e.g., morphosyntactic, semantic, and contextual information) all come into play at the initial stage of sentence processing. Syntax, in this approach, can carry a heavy weight in the sentence-interpretation process, but other sources of information such as discourse context can override syntactic biases and can also guide the initial phase of sentence processing.

The issue of whether contextual information affects the initial stage of sentence processing is still unresolved and perhaps unresolvable with available methods, because depth-first models do not specify the length of the lag between the first-stage context-free parsing process and the later context-sensitive reanalysis and integration process and hence “all early context effects can be attributed to second-stage reanalysis at an undetectable delay” (MacDonald et al., 1994, p. 696). Nevertheless, empirical evidence has been accumulating, showing that nonsyntactic information such as discourse context can override syntactic biases and can even guide sentence processing (for a review, see, e.g., Pickering & van Gompel, 2006). For example, research within the framework of Referential Theory (Altmann & Steedman, 1988; Crain & Steedman, 1985) has shown that processing difficulties caused by temporary structural ambiguities in a null or

unsupportive discourse context can be reduced or eliminated when the same ambiguities are presented in a supportive discourse context (e.g., Altmann, Garnham, & Dennis, 1992; Altmann, van Nice, Garnham, & Henstra, 1998; van Berkum, Brown, & Hagoort, 1999) or in a visually-presented referential context (e.g., Chambers, Tanenhaus, & Magnuson, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Discourse context can also eliminate processing difficulties that stem from complex structures. For example, Hoeks, Vonk, and Schriefers (2002) carried out a study that employed Dutch translations of such sentences as *The model embraced the designer and the photographer smilingly opened a bottle of champagne*, in which the NP *the photographer* is temporarily ambiguous between the object of *embraced* (NP-coordination) and the subject of a conjoined sentence (S-coordination). The NP-coordination analysis is structurally simpler than the S-coordination analysis, because in the latter case, the parser has to project another S-node rather than an NP node after *and*. Dutch readers showed difficulty processing such sentences, compared to control sentences which were disambiguated by a comma following *designer*. However, the difficulty with an S-coordination analysis disappeared when the same sentences were presented in a context that was biased towards such an analysis (e.g., when the example target sentence followed this context sentence: *When they met the fashion designer at the party, the model and the photographer were very enthusiastic*). Similar results were obtained in other studies as well: Supportive discourse contexts can reduce or even eradicate the processing difficulties associated, for instance, with noncanonical vs. canonical word orders (e.g., Finnish noncanonical OVS vs. canonical SVO, Kaiser & Trueswell, 2004), with object vs. subject relative clauses (e.g., Dutch, Mak, Vonk, & Schriefers, 2008; English, Fedorenko, Piantadosi, & Gibson, 2012; F. Yang, Mo, & Louwerse, 2013), and even with unambiguous sentences (e.g., restrictive vs. non-restrictive relative clauses, Grodner, Gibson, & Watson, 2005).

Although the influence of context on parsing is fairly well established, it should be noted that context effects are not always strong enough to override syntactic biases in ambiguity resolutions. For example, Britt, Perfetti, Garrod, and Rayner (1992) and Traxler and Tooley (2007) observed that context effects are much more difficult to obtain for main clause vs. reduced relative clause ambiguities (e.g., Binder, Duffy, & Rayner,

2001; Ferreira & Clifton, 1986) than for VP- vs. NP-attachment ambiguities (e.g., Altmann & Steedman, 1988). It seems that context override is of a selective nature, depending on the specific types of sentence, the types of context, and perhaps even the specific languages involved.

6.2 Processing of Contextual Information in Chinese

As mentioned above, context effects on parsing may have to do with the specific language at issue. For example, the Chinese language may require more contextual involvement in sentence processing than morphosyntactically more constrained languages like English. Several distinct properties of Chinese highlight the role of context in sentence interpretation. First, Chinese is a null-argument language, allowing both null subjects and null objects, and identifying the referents of those null arguments typically depends on the discourse context. Moreover, Chinese does not have inflectional morphology for marking verb tense, subject-verb agreement, case, etc., and has to rely on word order, functional words (e.g., *ba* for indicating preverbal objects), a small set of particles (e.g., aspect markers) to express various grammatical relations and notions, which are usually constrained by intrasentential and/or extrasentential context. In addition, Chinese does not mark word boundaries in the written form, and readers must rely on the meaning of each character and the surrounding context to determine whether the character is a word or part of a word. These properties of Chinese grant context a logically more important role than is the case in other languages like English (for further discussions about the distinct properties of Chinese and the role of context in Chinese, see Hsu & Hsieh, 2013; C. N. Li & Thompson, 1981; P. Li, 1998; Lu et al., 2000).

Interestingly, evidence from experimental studies of Chinese also seems to point to context as having a primary role. For example, two offline sentence-interpretation studies by Su (2001, 2004) suggest that sentence comprehension depends on discourse context to a much greater degree in Chinese than in English. The studies were designed to test whether word order, animacy, and discourse context affect one another in sentence interpretation in Chinese and in English. The experimental manipulations concern word

order (V + NP₁ + NP₂; NP₁ + V + NP₂; NP₁ + NP₂ + V),³⁰ animacy (Animate vs. Inanimate), and discourse context (NP₁-supported vs. NP₂-supported). Example (1), taken from the 2001 study, illustrates how discourse context was manipulated in the experiments (for sentences with the NP₁ + V + NP₂ order).

(1) NP₁ context

Tuzi shenqi-le. (Chinese version)

The rabbit is angry. (English version)

NP₂ context

Laohu duzi e. (Chinese version)

The tiger is hungry. (English version)

Target sentence

Tuzi yao laohu. (Chinese version)

The rabbit bites the tiger. (English version)

The context sentences used in the 2001 study were shorter than in the 2004 study (one short clause vs. two to three clauses). The Chinese version was given to Chinese native speakers (and also to L1-English L2ers of Chinese in the 2001 study); the English version, which was translated from the Chinese version, was given to English native speakers (and also to L1-Chinese L2ers of English in the 2001 study). Participants were asked to decide which NP is the doer of the action after they listened to a target sentence,³¹ which was preceded by a context sentence. Results from both studies showed that discourse context influenced L1 Chinese-speaking participants' choice of the doer of each action to a much greater degree than it did to L1 English-speaking participants (a difference of about 20%–42% of the experimental variance was accounted for by context), regardless of whether they interpreted the Chinese version or the English version of the sentences.

³⁰ Of these three word-order types, NP₁ + V + NP₂ is grammatical in Chinese (and English), corresponding to S + V + O; the other types (V + NP₁ + NP₂ and NP₁ + NP₂ + V), in which NP₁ is O and NP₂ is S, are ungrammatical in Chinese (and English) if spoken without any pause (see, e.g., Sun, 2006, p. 148).

³¹ The 2001 study did not report exactly how the experimental stimuli were presented to participants, but as a standard practice of such tasks, participants usually are asked to listen to recorded stimuli.

These results suggest that discourse context plays a more important role in (offline) sentence interpretation in Chinese than in English. Of course, the results can also be interpreted as indicating that word order is stricter in English than in Chinese (since verb-initial and verb-final orders in English transitive sentences are just ungrammatical).

Online studies of sentence comprehension in Chinese also reveal an important role of context. For example, some studies (e.g., S. Wang, Chen, Yang, & Mo, 2008; Zhang, Shu, Zhang, & Zhou, 2002) suggest that extrasentential context exerts an immediate influence on Chinese sentence processing. Zhang et al. (2002), for instance, examined how discourse context influences structural ambiguity resolutions in Chinese. They made use of an ambiguous structure (i.e., V + NP₁ + de + NP₂) in sentences like (2).

- (2) 嘱咐 患者 的 家属 认为 多吃 水果 有益 康复。
Zhufu huanzhe de jiashu renwei duochi shuiguo youyi kangfu.
urge patient DE relative think more-eat fruit benefit recovery
'The relative who urged the patient thought that eating more fruits benefits the recovery.'

Zhufu huanzhe de jiashu 'urge patient DE relative' in (2) is temporarily ambiguous between two constructions: modifier + noun (meaning "the relative who urged the patient") and verb + object (meaning "[I/you/someone] urged the patient's relative"). In both a self-paced reading experiment and an eye-tracking experiment, Chinese natives were asked to read target sentences containing such ambiguities. The results showed, first, that participants were reliably faster (at and following the *huanzhe* 'patient' region) when the discourse context preceding the target sentence had two referents rather than just one and, second, that a two-referent context clearly led to a modifier-noun reading. (This was the case even when in a pilot study, an isolated V + NP₁ + de + NP₂ biases towards a verb-object reading.) The effects of discourse context started very early on from the verb region (*zhufu* 'urge'), indicating that a two-referent context creates a strong expectation for a modifier structure so that the referents can be identified.

The study by S. Wang et al. (2008) examined the time course of discourse integration in Chinese. They tracked participants' eye movements as they read short

passages, each of which was manipulated to include a critical word that fits well within the sentential context but is either congruent or incongruent with the discourse context. This manipulation produced significantly longer first-pass reading times (RTs) at the critical word in the incongruent condition than in the congruent condition, suggesting that discourse-context information was employed immediately in Chinese sentence processing.

Briefly, the few available offline and online studies we discussed above confirm that context appears to be a primary factor that influences Chinese sentence processing.

6.3 L2 Processing of Discourse-context information

Research on L2 processing of contextual information is still in its infancy. Only a very small number of relevant studies exist (e.g., Hopp, 2009; Pan & Felser, 2011; Pan, Schimke, & Felser, 2015; Roberts, Gullberg, & Indefrey, 2008). The studies by Hopp (2009) and Roberts et al. (2008) essentially assumed that context would matter to L2 (and L1) processing, while the study by Pan and Felser (2011) and Pan et al. (2015) directly addressed how discourse context affects L2 sentence processing. These studies generally show that L2ers are able to integrate extrasentential contextual information in online sentence processing.

In the study by Hopp (2009), discourse contexts produced a reliable effect on the online processing of noncanonical OS vs. canonical SO (embedded) word orders in German by advanced and near-native L2ers from L1-English and L1-Russian backgrounds (but not those from an L1-Dutch background). In particular, felicitous discourse contexts facilitated the L1-English and L1-Russian L2ers' processing of scrambled OS orders, in comparison with canonical SO orders, indicating that the (near-native) L2ers' processing pattern converged on that of native speakers of German. Note that whereas Russian and German allow scrambled OS order and, moreover, the interpretation of this word order in these languages involves similar discourse-to-syntax mappings, no such correspondence exists between English and German because English does not have scrambling. The results thus demonstrate that discourse constraints on syntax are acquirable even for L2ers whose L1 lacks such constraints.

The study by Roberts et al. (2008) also looked at L2 processing of discourse contexts, but it does not directly address how discourse context influences L2 parsing decisions. Roberts et al. compared L1 and L2 pronoun resolutions in Dutch, a nonnull-subject language, in two offline tasks (acceptability judgment and sentence completion) and one online task (eye tracking while reading). The L2 participants included 2 advanced groups: L1-German (a nonnull-subject language, similar to Dutch) and L1-Turkish (a null-subject language). The experiments involved three conditions, as in (3).

(3) a. Local resolution

De werknemers zitten in het kantoor. Terwijl Peter aan het werk is, eet **hij** een boterham. Het is een rustige dag.

‘The workers are in the office. While Peter is working, **he** is eating a sandwich. It is a quiet day.’

b. Disjoint resolution

De werknemers zitten in het kantoor. Terwijl Peter aan het werk is, eten **zij** een boterham. Het is een rustige dag.

‘The workers are in the office. While Peter is working, **they** are eating a sandwich. It is a quiet day.’

c. Optional resolution

Peter en Hans zitten in het kantoor. Terwijl Peter aan het werk is, eet **hij** een boterham. Het is een rustige dag.

‘Peter and Hans are in the office. While Peter is working, **he** is eating a sandwich. It is a quiet day.’

The offline and online tasks produced different L2 results. In offline acceptability judgments, L2ers did not differ from Dutch natives. In offline sentence completion, there was an L1 transfer effect in that only the L1-German group, not the L1-Turkish group, showed a native-like local preference—interpreting the subject pronoun as coreferential

with the sentence-internal referent (i.e., interpreting *Peter* as the antecedent for *hij* ‘he’ in [3a] and [3c]). Notice that in (3c), the preceding discourse creates an ambiguity for the potential antecedents of *hij* ‘he’ (i.e., either *Peter* or *Hans*). The ambiguity posed an online problem for L2ers but not for natives. Both L2 groups in the eye-tracking experiment failed to show a native-like local preference (viz., for the sentence-internal referent). In contrast to natives who spent the shortest time reading the subject pronoun in the optional-resolution condition (3c), both groups of L2ers spent longer time reading the subject pronoun in that condition than in the other two conditions. Roberts et al. (2008) interpreted the online results as indicating a general L2 processing effect caused by the ambiguity of the subject pronoun in the optional-resolution condition.

Although L2ers in the Roberts et al. (2008) study failed to show a native-like online preference, they nevertheless did show online sensitivity to the manipulations of discourse context.

The study by Pan and Felser (2011) investigated how discourse contexts influence L2 processing of PP-attachment ambiguities. They used experimental sentences, such as (4), that contained two types of target sentences (one with VP-modifying PP attachment vs. one with NP-modifying PP attachment) and two types of contexts (a VP-supporting one-referent context vs. an NP-supporting two-referent context).

(4) a. VP-supporting (one-referent) context and VP-modifying PP attachment

Context: Bill walked into a shop that he knew the police were keeping an eye on.

There was only **one other customer** in the shop. The customer was wearing old and filthy clothes, whereas the sales assistant was dressed very smartly.

Target: Bill glanced at the customer **with strong suspicion** and then walked away.

b. VP-supporting (one referent) context and NP-modifying PP attachment

Context: Bill walked into a shop that he knew the police were keeping an eye on.

There was only **one other customer** in the shop. The customer was wearing old and filthy clothes, whereas the sales assistant was dressed very smartly.

Target: Bill glanced at the customer **with ripped jeans** and then walked away.

- c. NP-supporting (two referents) context and VP-modifying PP attachment
Context: Bill walked into a shop that he knew the police were keeping an eye on. There were **two other customers** in the shop. One customer was wearing old and filthy clothes, whereas the other one was dressed very smartly.
Target: Bill glanced at the customer **with strong suspicion** and then walked away.
- d. NP-supporting (two referents) context and NP-modifying PP attachment
Context: Bill walked into a shop that he knew the police were keeping an eye on. There were **two other customers** in the shop. One customer was wearing old and filthy clothes, whereas the other one was dressed very smartly.
Target: Bill glanced at the customer **with ripped jeans** and then walked away.

The results from the online self-paced reading task (using a noncumulative, segment-by-segment, moving-window paradigm) showed that in VP-supporting contexts, L1-Chinese L2ers of English were significantly faster reading a VP-modifying PP than an NP-modifying PP, while in NP-supporting contexts they read an NP-modifying PP significantly faster than a VP-modifying PP. In other words, L2ers read the critical PPs faster in congruent contexts than in incongruent contexts. In contrast, natives did not exhibit such an RT pattern, although their RTs at the critical PPs were numerically larger in congruent contexts than incongruent contexts. The online results contrasted with the results from an offline questionnaire task, in which both natives and L2ers were significantly affected by the discourse-context information in their choices of VP vs. NP modification (e.g., *with strong suspicion* vs. *with ripped jeans*). Pan and Felser (2011) interpreted the results as demonstrating that L2ers, in comparison to natives, over-rely on discourse-level contextual information in resolving structural ambiguities online. They argued that the study provides evidence for the SSH, which maintains that in contrast to L1 processing, L2 processing over-depends on meaning-based information such as discourse context (“L2 processing is generally guided more by discourse-level information than is the case in monolingual processing” [p. 224]). As for the results concerning natives, Pan and Felser contended that natives rely primarily on syntactic

information and delay the use of contextual information to the sentence-final interpretation stage, which accounts for why they only showed context effects offline.

However, the online native English results from Pan and Felser (2011) conflict with those from the L1 study by Altmann and Steedman (1988), which used a similar (but cumulative) phrase-by-phrase self-paced reading task with similar experimental sentences, such as (5).

(5) a. VP-supporting (one-referent) context and VP-modifying PP attachment

Context: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was **a safe which had a new lock** and a strongbox which had an old lock.

Target: The burglar blew open the safe **with the dynamite** and made off with the loot.

b. VP-supporting (one referent) context and NP-modifying PP attachment

Context: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was **a safe which had a new lock** and a strongbox which had an old lock.

Target: The burglar blew open the safe **with the new lock** and made off with the loot.

c. NP-supporting (two referents) context and VP-modifying PP attachment

Context: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was **a safe which had a new lock** and **a safe which had an old lock**.

Target: The burglar blew open the safe **with the dynamite** and made off with the loot.

d. NP-supporting (two referents) context and NP-modifying PP attachment

Context: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was **a safe which had a new lock and a safe which had an old lock.**

Target: The burglar blew open the safe **with the new lock** and made off with the loot.

Natives in the Altmann and Steedman study showed immediate use of discourse context information in their online processing of the PP attachments. What then caused the discrepancy between the Pan and Felser study and the Altmann and Steedman study? A comparison of the native participants' mean reading speeds at the critical PP region reveals that the natives were faster in Pan and Felser (641.75 milliseconds) than in Altmann and Steedman (659.25 milliseconds). It is likely that the faster speed caused context effects to show up in post-critical regions rather than in the critical region. Unfortunately, though, the sole post-critical region in Pan and Felser was the last region. This can be problematic since in sentence reading-time experiments, there is a typical sentence-final wrap-up effect, which reflects such processes as global integration (Just & Carpenter, 1980) and oculomotor hesitation (Warren, White, & Reichle, 2009). If both context effects and wrap-up effects are packed into the final region, the aggregate reading time may reach ceiling. And even if it does not reach ceiling, a standard practice in psycholinguistic data analysis—data trimming—may also likely wipe away the differences between the different context conditions.

Nevertheless, Pan and Felser's (2011) argument that L2ers over-rely on contextual information hinges on a null effect by natives. It therefore provides no strong evidence for divergent L1 and L2 processing mechanisms. The reason why Pan and Felser failed to obtain context effects from natives may simply be due to a Type II error (i.e., failing to reject the null hypothesis). It may well be that the experimental design as well as the data-trimming procedure obscured the effects. These problems undermine their L2 vs. L1 findings and, therefore, their L2 processing vs. L1 processing conclusions.

Moreover, even in L1 research, contextual information has been established as an important factor that influences parsing. The premise on which Pan and Felser's (2011) argument is based, namely, natives delay the use of contextual information, is a highly

controversial issue. As Traxler and Tooley (2007) discussed, contextual override most likely has to do with the specific types of sentences and contexts involved, and PP-attachment ambiguity, for example, is precisely one of the phenomena that often exhibit robust, immediate context effects. Obviously, contending that natives delay the use of contextual information cannot give solid support to Pan and Felser's claim that L2ers over-rely on meaning-based cues such as contextual information.

The study by Pan et al. (2015) also investigated how discourse context affects L1 and L2 ambiguity resolutions. It is very similar to Pan and Felser (2011) in terms of the design of the study, the results, and the interpretations. The only difference is that this study focused on the discourse-context effects in the processing of another kind of ambiguity resolution, namely, the resolution of relative-clause-attachment ambiguities rather than PP-attachment ambiguities. Again, the argument of an L1–L2 processing difference in this study was also based on natives' null effects.

The four studies discussed above have observed some differences as well as similarities between L1 and L2 processing of contextual information. All of them in effect demonstrate that L2ers are able to use discourse-context information in sentence processing (as their proficiency rises). Nevertheless, the results of these studies were construed from different perspectives. Pan and Felser (2011) and Pan et al. (2015) viewed their L2ers as having enhanced abilities in using discourse-context information, in comparison with natives, to the point of over-relying on it so as to compensate for their inabilities in using morphosyntactic information. Hopp (2009) and Roberts et al. (2008), on the other hand, note that processing at the discourse-syntax interface requires coordinating and integrating information from the domains of both discourse and syntax, and suggest that up to a certain proficiency level in L2 acquisition, doing so may exceed L2ers' cognitive abilities and therefore cause L2 processing difficulties (also discussed in Sorace, 2011). In addition, these four studies interpreted their L2 results in rather different ways: While Pan and colleagues argued that their L2 results support the SSH, Hopp (2009) and Roberts et al. (2008) did not, attributing their L2 results to, respectively, cognitive limitations in L2 processing and general L2 processing effects. Evidently, the unsettled issues surrounding the L2 processing of contextual information call for more systematic investigations.

The present study deals with the L2 processing of discourse-context information to examine whether L2ers, as the SSH (Clahsen & Felser, 2006a, 2006b, 2006c) claims, (necessarily) over-depend on meaning-based information in L2 sentence processing, relative to natives in L1 sentence processing. Noting that discourse-context information is a type of meaning-based information, Pan and Felser (2011) and Pan et al. (2015), in line with the SSH, hypothesized, as mentioned above, that “L2 processing is generally guided more by discourse-level information than is the case in monolingual processing” (Pan & Felser, 2011, p. 224). The current study challenges this hypothesis. We approach the issue of contextual influence on L2 processing by taking advantage of Chinese, a highly context-dependent language. If L2ers are good at processing discourse-context information as the SSH suggests, and, at the same time, if the Chinese language encourages the practice of using contextual information, we have good reason to expect L2ers to be able to use contextual information from very early on in the course of L2 development. If L2ers, relative to natives, indeed show enhanced abilities in using discourse contexts to facilitate sentence processing, then this would support the SSH and, simultaneously, challenge approaches (e.g., Hopp, 2009; Roberts et al., 2008; Sorace, 2011) which propose that the integration of discourse information and syntactic information causes L2 difficulty. If, on the other hand, L2 results from Chinese, a language that typologically favors the SSH, disconfirm that L2ers, relative to natives, are over-reliant on contextual information, then this would be counterevidence to the SSH.

The present study chose to use null objects (object drop) as the target for investigating discourse-context effects in Chinese.³² Null objects are a type of reduced referring expression for achieving discourse coherence (e.g., C. N. Li & Thompson, 1981; C. L. Yang, Gordon, Hendrick, & Wu, 1999). As a null-argument language, Chinese does not always require an overt subject or (with transitive verbs) an overt object.

Null objects occur frequently in Chinese, but not randomly. Discourse contexts establish discourse topics, which license null objects (Huang, 1984, 1991). Since null objects are constrained by discourse context, they should be sensitive to manipulations of contextual appropriateness. If a null object is not supported by the discourse context, as in (6a), this should cause processing difficulties at the region where it occurs (i.e., the

³² Chinese does not have PP-attachment ambiguities.

critical *kan* ‘watch’ region), relative to (6b), in which an appropriate discourse context does support the null object.

(6) a. Yijing bu-zao-le; ta zheng-yao **kan** *e_i*, **dianying_i** jiu fang-wan -le.
already not-early-LE 3sg just-will **watch** **movie** then play-finish -LE
‘It was already a little late; just as s/he was about to watch [it], the movie was over.’

b. **Dianying_i** hen-youming; ta zheng-yao **kan** *e_i*, **dianying_i** jiu fang-wan -le.
movie very-famous 3sg just-will **watch** **movie** then play-finish -LE
‘The movie was very famous; just as s/he was about to watch [it], the movie was over.’

Note that the use of a comma after a transitive verb, as in (6), *orthographically forces* the reader to posit a null object. In (6b) the object of the verb *kan* ‘watch’ is null, but it is coindexed with the discourse topic *dianying* ‘movie; hence, the referential requirement of the null object is satisfied. In (6a), however, the prior discourse context does not provide a relevant discourse topic and thus cannot fulfill the referential requirement of the null object. If the preceding discourse does not establish the referent of the null object, it should be difficult for readers to process the null object. Readers have to either make inferences about it in order to make up for the missing information or keep searching for a potential referent. These efforts have a cost, e.g., taking up extra memory resources, and should lead to longer processing times at and/or following the critical verb region (*kan* ‘watch’ in [6a] relative to [6b]).

Since null objects occur frequently in simple everyday use of Chinese, L2ers should have ample null-object input used in appropriate contexts.

The current study includes two experiments with critical sentences like (6). Experiment 6.1 employed comprehension-focused self-paced reading (CFSPR) as the online task. It was intended to see how participants make use of contextual information in meaning-focused processing. Experiment 6.2 tested participants’ online use of contextual information in an acceptability-judgment self-paced reading (AJSPR) task. This task was

intended to be form-focused—but as we shall see later, the participants in this study also based their acceptability judgments on the propositional content and pragmatic appropriateness of the experimental stimuli (see the discussions in Section 6.5.2.2). In any case, this AJSPR task can engage participants in detailed, incremental processing, which makes it possible for us to see whether a task that requires more fully specified processing also promotes participants’ sensitivity to contextual information at the discourse-syntax interface.

In addition, each experiment also included an offline comprehension task, designed to assess whether participants could understand the target sentences correctly when they read them without time pressure and with the possibility to “go back.” This design allows us to compare participants’ online and offline performances. In addition, since half of the target sentences involved null objects (see below), the offline data can tell us whether participants were able to comprehend sentences with null objects and hence whether they had knowledge of null objects in Chinese, without which comprehension of the target sentences should be hampered. In brief, the offline data can help us make better sense of participants’ online responses.

The main research question of the present study is formulated as follows:

1. Does L2 processing rely on contextual information more than L1 processing does, as indicated by L2ers’ enhanced or early discourse-context effects on null-object processing at the critical region vs. Chinese natives’ “reduced or delayed” (Pan & Felser, 2011, p. 226) discourse-context effects at the post-critical region(s)?

6.4 Experiment 6.1: Comprehension-Focused Self-Paced Reading

This experiment used CFSPR, a more meaning-focused online task, in comparing how discourse contexts influence L1 and L2 sentence processing in Chinese, in addition to an offline comprehension task. As discussed above, the SSH (Clahsen & Felser, 2006a, 2006b, 2006c) predicts that discourse context should affect L2ers’ online processing of null objects in Chinese to a (much) larger extent than it does for natives, in the sense that intermediate-to-advanced L2ers of Chinese should show sensitivity to discourse-context

manipulations earlier (i.e., at the critical region), while Chinese natives should show sensitivity to such manipulations later (i.e., at post-critical regions). If the SSH is correct, processing a highly context-dependent language in a meaning-focused task should maximize the opportunity of observing the predicted L2 context effects. On the other hand, approaches that hypothesize L2 difficulty with the integration of discourse information and syntactic information (e.g., Hopp, 2009; Roberts et al., 2008) would predict delayed and/or weak discourse-context effects for L2ers rather than for natives.

6.4.1 Method

6.4.1.1 Participants

Participants of this experiment were 30 adult L1-English learners, 30 adult L1-Japanese learners, and 25 adult L1-Chinese controls. They were all compensated with a small payment for taking part. Due to unexpected interruptions during testing sessions, one Chinese participant and two Japanese participants were excluded.

Having L2 participants from two different language backgrounds permits us to examine possible L1 effects. English is a nonnull-object language while Japanese is a null-object language. Japanese is very similar to Chinese in terms of how null objects operate, in terms of both licensing and identification (e.g., Abe, 2009; Huang, 1984). Because of this similarity between the two languages, we predict that owing to L1 transfer, L1-Japanese L2ers of Chinese should pattern more like Chinese natives than L1-English L2ers of Chinese do. By including participants of null-object and nonnull-object languages alike, we can assess the extent to which the way context and syntax interact in the L1 shapes the processing of a similar/distinct phenomenon in the target language (TL).

To examine L2 proficiency effects, L2 participants from each language background were divided into 2 proficiency levels (Advanced and Intermediate) according to their Chinese proficiency scores on a 50-point quasi-C-test, developed for this dissertation research (see Chapter 3: General Methods). For conducting balanced statistical analyses, this experiment used the L1-English L2ers' median score (21) as the proficiency cutoff point. The same cutoff point was applied to grouping the L1-Japanese participants so that the L2 groups could be compared with reference to their proficiency

levels. There were thus 2 Advanced L2 groups (English: $n = 15$; Japanese: $n = 18$) and 2 Intermediate L2 groups (English: $n = 15$; Japanese: $n = 10$). All the participants in the Advanced L2 groups had a proficiency score of 21 and above, and those in the Intermediate L2 groups had a score of less than 21.

Table 6.1 gives the Chinese background information of all 5 participant groups.

Table 6.1. Chinese Background Information of the Participants in Experiment 6.1

Group	Years of learning Chinese		Years of residence in China/Taiwan		Chinese proficiency scores (Max = 50)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Natives ($n = 24$)	N/A	N/A	N/A	N/A	46.2	2.0
English Advanced L2ers ($n = 15$)	6.2	7.3	3.9	6.6	29.8	7.8
Japanese Advanced L2ers ($n = 18$)	2.6	2.1	1.8	2.4	29.6	7.5
English Intermediate L2ers ($n = 15$)	2.7	1.4	1.1	1.1	14.1	5.7
Japanese Intermediate L2ers ($n = 10$)	1.9	1.1	1.2	1.2	18.3	1.9

A one-way ANOVA showed that the 5 groups' Chinese proficiency levels differed from each other significantly, $F(4, 77) = 92.56, p < .001$. SPSS Bonferroni corrected post hoc tests showed that the L1 group had significantly higher Chinese proficiency than all the L2 groups ($ps < .001$) and the Advanced L2 groups had significantly higher Chinese proficiency than the Intermediate L2 groups ($ps < .001$). No significant difference emerged between the 2 Advanced L2 groups ($p = 1.000$) or between the 2 Intermediate L2 groups ($p = .693$), although the L1-English L2ers, on average, had spent (many) more years learning Chinese and had resided longer in Chinese-speaking countries than the L1-Japanese L2ers had.

6.4.1.2 Materials

The online task included 16 experimental items like (7), using simplified Chinese characters when presented to participants (see Appendix AA for a complete list of the experimental items).

(7) a. Irrelevant context and null object (“Irrelevant-Null” Condition)

Yijing bu-zao-le; ta zheng-yao **kan, dianying** jiu fang-wan -le.
already not-early-LE 3sg just-will **watch movie** then play-finish-LE
‘It was already a little late; just as s/he was about to watch [it], the movie was over.’

b. Relevant context and null object (“Relevant-Null” Condition)

Dianying hen-youming; ta zheng-yao **kan, dianying** jiu fang-wan -le.
movie very-famous 3sg just-will **watch movie** then play-finish-LE
‘The movie was very famous; just as s/he was about to watch [it], the movie was over.’

c. Irrelevant context and nonnull object (“Irrelevant-Nonnull” Condition)

Yijing bu-zao-le; ta zheng-yao **kan dianying** jiu fang-wan -le.
already not-early-LE 3sg just-will **watch movie** then play-finish-LE
i) ‘It was already a little late; just as s/he was about to watch [it], the movie was over.’
ii) ‘It was already a little late; just as s/he was about to watch the movie, [it] was over.’

d. Relevant context and nonnull object (“Relevant-Nonnull” Condition)

Dianying hen-youming; ta zheng-yao **kan dianying** jiu fang-wan -le.
movie very-famous 3sg just-will **watch movie** then play-finish-LE
i) ‘The movie was very famous; just as s/he was about to watch [it], the movie was over.’
ii) ‘The movie was very famous; just as s/he was about to watch the movie, [it] was over.’

Each item has four conditions, as shown in (7). The *irrelevant-null* condition (7a) contained an irrelevant context sentence and a target sentence with a null object. The *relevant-null* condition (7b) contained a relevant context sentence and the same target

sentence. Note that the null object in the target sentence of these two conditions was created by the use of a comma after the critical verb (i.e., *kan* ‘watch’). These two conditions were used to test whether there is an effect of discourse context on participants’ parsing of null objects in Chinese. If preceding discourses influence parsing, (7b) should be read faster than (7a) at the critical verb region and/or at the post-critical (i.e., *dianying* ‘movie’) region (if context effects spill over to this region), because the referent of the null object has been supplied by the preceding discourse context in (7b) but not in (7a).

One problem arises if we look at spillover effects at the post-critical region, because the discourse topic NP (*dianying* ‘movie’) in (7b) reappeared (as the subject) in this region; and the repetition of this NP might produce a (receptive) lexical priming effect. In other words, the NP (*dianying* ‘movie’) at the post-critical region may be processed faster simply because it is primed in the context. If this is the case, then a priming effect and a discourse-context effect are confounded at this region. To check whether a priming effect played a role at this region, the experiment also included two nonnull-object conditions: the *irrelevant-nonnnull* condition (7c) and the *relevant-nonnnull* condition (7d). These two conditions corresponded exactly to the two null-object conditions except that there was no comma after the critical verb. Notice that the NP *dianying* ‘movie’ in the target sentence of (7c) and (7d) is temporarily ambiguous between an object and a subject. In the psycholinguistic literature, sentences such as (7c) and (7d), which involve a temporary structural ambiguity, are called “garden-path” (GP) sentences (e.g., Bever, 1970; Frazier & Fodor, 1978; Pritchett, 1992; Weinberg, 1999). The NP *dianying* in (7c) should be temporarily analyzed as a direct object in online reading before the parser reaches the disambiguating *fang-wan* ‘play-finish’ region. This is because the verb *kan* ‘watch’ is transitive and hence requires an object. The object analysis is also in line with the “late closure” principle (Frazier, 1987), because *dianying* ‘movie’ is the phrase currently being processed. In (7d), however, the NP *dianying* in the target sentence can, in principle, also be locally analyzed as a subject, because the preceding discourse can license a null object. If this is the case, (7d) will not be a GP sentence anymore and should be processed in the same way as (7b) where *dianying* is the subject. However, we argue that this is not very likely to happen. Although a discourse topic in Chinese *can* license null objects, it does not obligatorily force a null object for

every transitive verb if that verb is followed by an NP that can serve as the potential object. In other words, without a comma to force positing a null object (as in [7b]), *dianying* in both (7c) and (7d) should be temporarily analyzed as an object. If (7c) and (7d) do not involve nonlocal referential dependencies at the critical verb *kan* region, then the following NP region should not be sensitive to discourse manipulations; in this case, the subject of the second part of the target sentence in (7c) and (7d) becomes null, and this null subject should be sensitive to the discourse-context manipulations. Nevertheless, regardless of whether the parser analyzes the NP *dianying* after the verb *kan* in (7d) as an object or a subject, this region involves no immediate reanalysis or unsatisfied referential requirement, and therefore the RT patterns of (7c) and (7d) should not differ at this region.³³ In short, any observable effect at this NP region from the two nonnull-object conditions should simply be due to lexical priming.

As in (7), the context sentence always appeared before the target sentence. They were always separated by a semicolon. Irrelevant contexts (e.g., in [7a] and [7c]) were created by using simple sentences about time, weather, and the like, which did not provide the appropriate discourse topics for a null object in the target sentences. Relevant contexts were created by using simple sentences whose subject NP (e.g., *dianying* ‘movie’ in [7b] and [7d]) served as the discourse topic.

The Chinese characters used in the experimental items were all familiar to the intermediate-level (and hence advanced-level) L2ers (for the detailed procedures that were followed in developing the experimental materials, see Chapter 3: General Methods).

For the purpose of the self-paced reading task, each experimental sentence was split into 9 regions, as show in Table 6.2.

³³ This assumes that the processing required of null object to overt subject is equivalent to that of overt object to null subject (B. D. Schwartz, personal communication, February 6, 2015).

Table 6.2. *Sample Set of Experimental Stimuli in Experiment 6.1*

Conditions	Regions								
	1	2	3	4	5	6	7	8	9
Irrelevant-null	已经	不早了;	他	正要	看,	电影	就	放完	了。
	Yijing already	bu-zao-le; not-early-LE	ta 3sg	zheng-yao just-will	kan, watch	dianying movie	jiu then	fang-wan play-finish	-le. -LE
Relevant-null	电影	很有名;	他	正要	看,	电影	就	放完	了。
	Dianying movie	hen-youming; very-famous	ta 3sg	zheng-yao just-will	kan, watch	dianying movie	jiu then	fang-wan play-finish	-le. -LE
Irrelevant-nonnull	已经	不早了;	他	正要	看	电影	就	放完	了。
	Yijing already	bu-zao-le; not-early-LE	ta 3sg	zheng-yao just-will	kan watch	dianying movie	jiu then	fang-wan play-finish	-le. -LE
Relevant-nonnull	电影	很有名;	他	正要	看	电影	就	放完	了。
	Dianying movie	hen-youming; very-famous	ta 3sg	zheng-yao just-will	kan watch	dianying movie	jiu then	fang-wan play-finish	-le. -LE

For the two null-object conditions (i.e., the irrelevant-null condition and the relevant-null condition), the regions of interest were Region 5 (the critical verb region) and Regions 6 and 7 (the post-critical regions). We examined the RTs of these regions in order to find out whether there is any immediate, delayed, or spillover effect of discourse context on participants' parsing of null objects in Chinese. For the two nonnull-object conditions (i.e., the irrelevant-nonnull condition and the relevant-nonnull condition), we examined the RTs of only Region 6, because the purpose of these two conditions was to check whether a discourse topic NP causes a lexical priming effect at this region (7d).

To focus participants on processing the meaning, each trial was followed by a comprehension question. The questions also serve to check whether participants were able to understand the test materials accurately as well as retain the meaning after they read each sentence. The questions were like this: *Ta kan dianying le, dui ma?* 'He watched the movie, right?' (for [7]). The number of *dui* 'right' and *bu dui* 'not right' answers was kept counterbalanced. One half of the comprehension questions verified the meaning of the first part of the target sentence, and the other half verified the meaning of the second part of the target sentence. Note that the comprehension questions never asked about the context sentences. This was done for these three reasons: (a) to encourage participants to focus on the target sentences; (b) to avoid drawing their attention to the contextual manipulations; (c) to permit us to assess whether they were able to comprehend the target sentences correctly.

Since the discourse topic NP of each experimental sentence is coindexed with the null object of the critical verb, it is necessary to check whether that NP goes well with the verb. To check this, a norming study was administered to 20 Chinese natives, all from Beijing, who did not participate in the experimental tasks. They were asked to judge the acceptability of 16 simple sentences such as (11), in which the 16 verb-object combinations were embedded.

- (11) Ta zai he niunai.
 3sg ZAI drink milk
 ‘He/She is drinking milk.’

All the norming sentences followed exactly the same pattern as in (11). They were mixed with 128 filler sentences of a variety of syntactic kinds. Participants made acceptability decisions by circling one of the two answer options following each sentence: *tongshun* ‘well-formed, acceptable’ and *bu tongshu* ‘not well-formed, not acceptable’. Each *tongshun* answer was assigned a score of 1, and each *bu tongshu* answer was assigned 0. The means and the standard deviations of the acceptability rates are given in Table 6.3.

Table 6.3. *Participants’ (N = 20) Acceptance Rates of the Norming Sentences (k = 16) in Experiment 6.1*

Verb-Object Combination	<i>M</i>	<i>SD</i>
1. <i>ca</i> ‘wipe’ + <i>xiangzi</i> ‘box’	1.00	0.20
2. <i>canguan</i> ‘visit’ + <i>gongchang</i> ‘factory’	1.00	0.20
3. <i>canjia</i> ‘take part in’ + <i>bisai</i> ‘competition’	1.00	0.00
4. <i>chi</i> ‘eat’ + <i>fan</i> ‘meal’	0.90	0.30
5. <i>chuan</i> ‘put on’ + <i>yifu</i> ‘clothes’	0.90	0.30
6. <i>da</i> ‘call’ + <i>dianhuan</i> ‘phone’	1.00	0.00
7. <i>dasao</i> ‘clean’ + <i>fangjian</i> ‘room’	0.90	0.30
8. <i>guan</i> ‘switch off’ + <i>diandeng</i> ‘light’	0.90	0.30
9. <i>he</i> ‘drink’ + <i>niunai</i> ‘milk’	1.00	0.00
10. <i>kai</i> ‘drive’ + <i>che</i> ‘car’	1.00	0.00
11. <i>kan</i> ‘watch’ + <i>dianying</i> ‘movie’	1.00	0.00
12. <i>mai</i> ‘buy’ + <i>shouji</i> ‘cellphone’	0.90	0.30
13. <i>na</i> ‘take’ + <i>beizi</i> ‘cup’	0.80	0.40
14. <i>ting</i> ‘listen to’ + <i>yinyue</i> ‘music’	1.00	0.20
15. <i>wan</i> ‘play’ + <i>youxi</i> ‘game’	1.00	0.20
16. <i>yong</i> ‘use’ + <i>diannaoh</i> ‘computer’	1.00	0.00
Overall	0.9	0.2

The acceptance rates here are extremely high, suggesting that the verbs and their topic/object NPs are appropriate for the experiments.

6.4.1.3 Procedure

The experiment used a CFSPR task. Participants were tested individually. The DMDX software (Forster & Forster, 2003) was used to run the experiment. The design ensured that the participants were unaware of the purposes of the experiment. Before starting, participants were given specific instructions on the experimental procedure. They were asked to read the Chinese sentences on a computer monitor in a segment-by-segment fashion at their own pace, and at the end of each sentence, they answered a *dui* ‘right’/*bu dui* ‘not right’ comprehension question by pressing one of the two designated keys for *Dui* (Yes) and *Bu dui* (No) on the computer keyboard. The instructions (given in Appendix I) also gave a specific example. The experimental materials were presented in simplified Chinese characters (font: STSong; font size: 14). Each test sentence appeared in a single line on the computer screen, going from left to right. The self-paced reading task used a noncumulative moving-window paradigm (Just et al., 1982). The session commenced with 8 practice items (see Appendix J). After the self-paced reading experiment, participants were asked to fill out a language background questionnaire first (Appendix E) and then take the Chinese proficiency test (Appendix F). Finally, participants completed the offline comprehension task testing the same sentences read in the online task (but with a different *right/not right* comprehension question—see above). The entire experimental session lasted about 40–60 minutes.

6.4.2 Results

6.4.2.1 Offline Task Results

The 5 participant groups’ comprehension accuracy rates for the experimental sentences in the offline task are shown in Figure 6.1 (for the descriptive statistics, see Appendix AB).

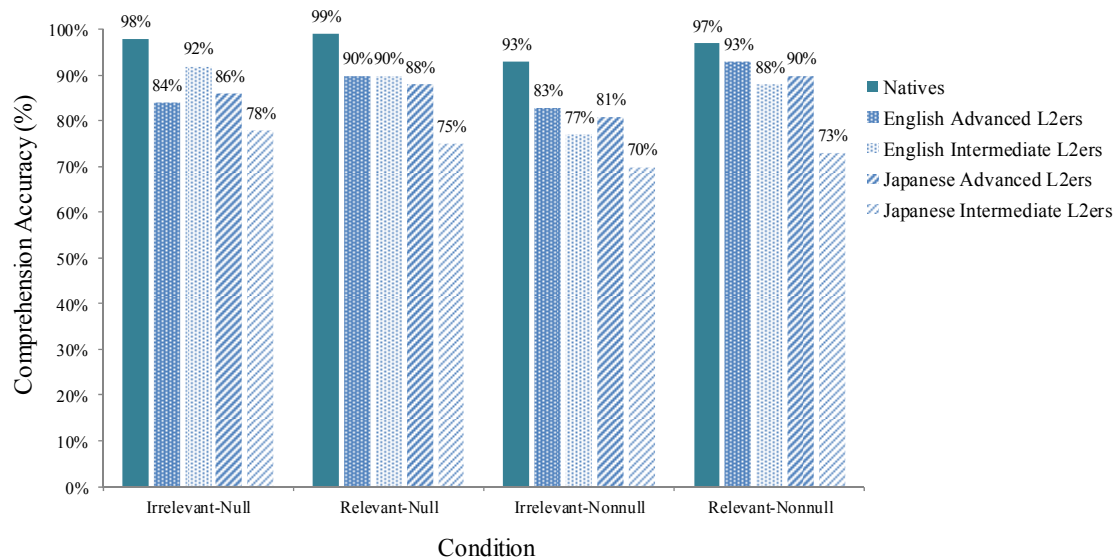


Figure 6.1. Mean Comprehension Accuracy Rates for the Experimental Sentences in the Offline Task of Experiment 6.1.

We can see from Figure 6.1 that understanding of the target sentences in all the groups was good (all mean accuracy rates $\geq 70\%$). For the comprehension questions in the null-object conditions, all L2 groups' accuracy was equal to or above 75%, suggesting that null objects were not an obstacle to their comprehending the target sentences.

A $5 \times 2 \times 2$ (Group [Native, English Advanced L2, English Intermediate L2, Japanese Advanced L2, Japanese Intermediate L2] \times Context [irrelevant, relevant] \times Object [null, nonnull]) repeated measures ANOVA performed on the arcsine-transformed accuracy data yielded a main effect of group, $F(4, 77) = 7.42, p < .001, \eta_p^2 = .28$, a main effect of context, $F(1, 77) = 7.20, p = .009, \eta_p^2 < .09$, a marginal effect of object, $F(1, 77) = 3.14, p = .080, \eta_p^2 < .03$, and a significant Context \times Object interaction, $F(1, 77) = 7.60, p = .007, \eta_p^2 = .09$. These effects indicate that the 5 participant groups responded to the comprehension questions differently for different context-object conditions. SPSS Bonferroni corrected post hoc tests showed that regardless of condition, the Native group was more accurate than the English Intermediate group ($p = .035$) and the 2 Japanese groups ($p \leq .030$) but not the English Advanced group ($p = .142$). Of the 5 participant groups, the Native group was the most accurate in answering the comprehension questions ($M_{\text{accuracy}} = 97\%, SD = .07$) and the Japanese Intermediate group was the least accurate ($M_{\text{accuracy}} = 74\%, SD = .13$). In addition, the participants

generally responded more accurately in relevant-context conditions than in irrelevant-context conditions ($p = .009$) and in null-object conditions than in nonnull-object conditions ($p = .080$), suggesting that relevant discourse context succeeds in facilitating comprehension of target sentences and that nonnull-object conditions are harder to understand than null-object conditions (probably due to the temporary object-subject ambiguities in those sentences).

To see how the experimental conditions influenced each individual group's offline accuracy, we performed separate 2×2 (Context [irrelevant, relevant] \times Object [null, nonnull]) repeated measures ANOVAs on the arcsine-transformed accuracy data. The Natives showed a main effect of context, $F(1, 23) = 4.66, p = .042, \eta_p^2 = .17$, a main effect of object, $F(1, 23) = 5.53, p = .028, \eta_p^2 = .19$, and a nonsignificant Context \times Object interaction, $F(1, 23) = 2.49, p = .12, \eta_p^2 = .10$. As for the L2ers, a marginal main effect of context was found in the English Advanced L2 group, $F(1, 14) = 4.07, p = .063, \eta_p^2 = .23$, and in the Japanese Advanced L2 group, $F(1, 17) = 3.78, p = .069, \eta_p^2 = .18$. No other main effect or interaction was statistically significant ($F_s < 3.03, p_s > .10$). These data patterns further indicate the effects of proficiency and context on the participants' comprehension accuracy.

In brief, the offline data demonstrate that the participants generally responded accurately to the comprehension questions of the experimental sentences ($M_{\text{accuracy}} = 86\%$) and their accuracy rates reflected influences of Chinese proficiency and discourse context.

6.4.2.2 Online Task Results

Data analysis of the online task covered both participants' comprehension accuracy and RTs.

As mentioned above, one special type of filler ($k = 8$) was used for checking participants' attentiveness in performing the online task as well as for assessing participants' normal online comprehension accuracy (see Appendix H). Participants' mean accuracy rates for this type of filler were extremely high (see Table 6.4). The results indicated high attentiveness and high comprehension accuracy for all the groups. As can be seen from Table 6.4, the accuracy rates of the other fillers were also high.

Table 6.4. Mean Comprehension Accuracy Rates of the Fillers in Experiment 6.1

Group	Type	
	Special ($k = 8$)	Other ($k = 64$)
Natives ($n = 24$)	.98 (.04)	.91 (.04)
English Advanced L2ers ($n = 15$)	.99 (.03)	.89 (.04)
English Intermediate L2ers ($n = 15$)	.91 (.09)	.88 (.04)
Japanese Advanced L2ers ($n = 18$)	.93 (.08)	.88 (.04)
Japanese Intermediate L2ers ($n = 10$)	.94 (.09)	.87 (.08)

Note. Standard deviations are in parentheses.

As for the experimental sentences, the 5 participant groups' comprehension accuracy rates are shown in Figure 6.2 (for the descriptive statistics, see Appendix AC).

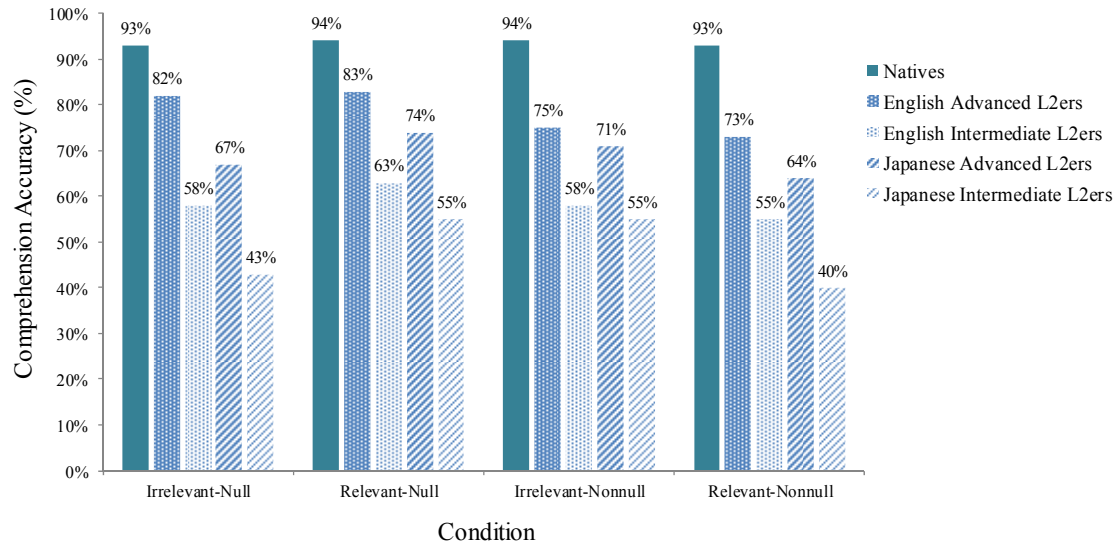


Figure 6.2. Mean Comprehension Accuracy Rates for the Experimental Sentences in the Online Task of Experiment 6.1.

The Native participants responded accurately to 93% or more of the comprehension questions. The Advanced L2ers' comprehension accuracy rates were between 64% and 83%, better than those of the 2 Intermediate L2 groups, which were between 40% and 63%. The Natives' online accuracy was not too different from their offline accuracy, but the L2 participants' online accuracy was worse than their offline accuracy. Not surprisingly, the L2ers found the online task much more demanding than the offline task, and the online time pressure affected their comprehension accuracy.

A $5 \times 2 \times 2$ (Group [Native, English Advanced L2, English Intermediate L2, Japanese Advanced L2, Japanese Intermediate L2] \times Context [irrelevant, relevant] \times Object [null, nonnull]) repeated measures ANOVA performed on the arcsine-transformed accuracy data showed a significant main effect of group, $F(4, 77) = 28.24, p < .001, \eta_p^2 = .60$, and a marginal Context \times Object interaction, $F(1, 77) = 3.14, p = .080, \eta_p^2 = .04$. No other main effect or interaction reached statistical significance ($F_s < 1.36, p_s > .24$). SPSS Bonferroni corrected post hoc tests showed that the Native group outperformed all the L2 groups ($p \leq .008$). The English Advanced group outperformed the English Intermediate group ($p = .002$) and the Japanese Intermediate group ($p < .001$), and the Japanese Advanced group outperformed the Japanese Intermediate group ($p = .003$) but not the English Intermediate group ($p = .841$). The 2 Advanced L2 groups did not differ from each other significantly ($p = .263$), and neither did the 2 Intermediate L2 groups ($p = .353$). These comparisons attest to the influence of proficiency on participants' online task performance. The low accuracy of the Intermediate L2 groups might indicate that the experimental sentences were somewhat difficult for them to comprehend under the online reading pressure. Another possibility is that the meanings of the experimental sentences were difficult for the participants to retain—even though they had no problem comprehending them—because the sentences were long. This latter interpretation is especially pertinent if we take into consideration the participants' much higher accuracy rates in the offline task. Nevertheless, failing to respond accurately to a comprehension question in the online task cannot be taken as directly pointing to a failure in processing the experimental sentence or, more specifically, the null object involved.

Note that numerically, the 2 English groups achieved higher accuracy than their 2 Japanese counterparts. This is probably due to a speed vs. accuracy tradeoff. Scrutinizing these groups' mean reading speeds (in terms of milliseconds, calculated over the RTs of all regions), we found that the Japanese participants read the experimental stimuli much faster than the English participants did: (a) The Japanese Advanced group ($M_{\text{speed}} = 520.3$) was 302.4 milliseconds faster than the English Advanced group ($M_{\text{speed}} = 822.7$); (b) the Japanese Intermediate group ($M_{\text{speed}} = 531.4$) was 347.5 milliseconds faster than the English Intermediate group ($M_{\text{speed}} = 878.9$).

Overall, the online accuracy data, in tandem with the offline data, indicate that the participants' response accuracy in the online task was affected by their proficiency in Chinese, context-object condition, online reading pressure, and reading speed.

The above-mentioned factors work together to sway the outcome of participants' online comprehension accuracy. Because of this, our RT data analyses reported here did not exclude the trials on which participants answered the comprehension questions incorrectly.³⁴ As discussed in Chapter 4, one primary reason to exclude online data for incorrect trials should be that the errors are caused by inattentiveness on the part of participants. If we exclude incorrect trials caused by the factors under investigation rather than by inattentiveness, we are in danger of losing crucial information. Since our participants demonstrated a very high degree of attentiveness in the online task, assessed through the special type of filler, we contend that our participants' inaccurate responses to the comprehension questions are not caused by inattentiveness. The other reason to exclude incorrect trials is that they actually result from participants' lack of the relevant knowledge. Our participants read the same sentences in the offline and online tasks, but they understood the sentences very well in the offline task, which indicates that they did not lack knowledge of null objects in Chinese; hence, their incorrect responses are unlikely to have arisen from ignorance. In brief, if the incorrect trials are not caused by carelessness or ignorance on the part of participants, we have no good reason to remove those trials. For the above considerations, our participants' RT data from incorrect trials were all kept in the data analysis.

To evaluate the effects of discourse context on participants' online processing of null objects in Chinese, the RT data analysis for the two null-object conditions focused on the critical and two post-critical regions (i.e., Regions 5, 6, and 7). For the two nonnull-object conditions, the focus was on Region 6. As discussed in Section 6.4.1.2, these two conditions are crucial for checking whether any *apparent* discourse-context effect observed at Region 6 (from the two null-object conditions) is *actually* due to a lexical priming effect.

³⁴ Our data analyses based on correct trials only (see Appendix AD for the trimmed raw RT means) indicate that the results became very noisy and sometimes spurious when a large proportion of the data was removed (see Appendix AE for the consequent changes in the significance levels). This is especially the case with the 2 Japanese groups, whose mean accuracy rates were rather low (ranging from 40% to 63%).

Prior to the RT data analysis, the raw RTs were trimmed to adjust outlier responses. As described in Chapter 3, the data-trimming followed two steps. First, all RTs longer than 2,500 ms were substituted with that cutoff point. Second, each participant's RTs exceeding two standard deviations from that participant's mean were replaced with the corresponding cutoff. These data-trimming steps affected 4.5% of the Natives' data, 9.6% of the English L2ers' data, and 5.7% of the Japanese L2ers' data. (For the trimmed raw RT means, see Appendix AD.)

In the following, we examine the 5 participant groups' RT profiles one by one. For the statistical analyses, we perform separate paired-samples *t* tests (two-tailed) on the RTs at each region of interest (Regions 5–7 in the null-object conditions and Region 6 in the nonnull-object conditions). The statistical tests ran on both participant (t_1) and item (t_2) means. Following the standard convention of reading-time studies, an alpha level of .05 was used for all the statistical tests.

The Native group. Figure 6.3 presents the Native group's region-by-region mean RTs for the null-object conditions. As can be seen, the Natives were faster at Regions 5–7 in the relevant-null condition than in the irrelevant-null condition, which indicates that discourse context influenced this group's processing of the null objects. Statistically, the difference between the irrelevant-null and relevant-null conditions was significant at Region 5, $t_1(23) = 2.20, p = .038$; $t_2(15) = 3.01, p = .009$. The effect also spilled over to the post-critical regions: Although nonsignificant (but numerically large) at Region 6 ($ts < 1.26, ps > .22$), the effect at Region 7 became marginal in the participant analysis, $t_1(23) = 1.92, p = .067$, and significant in the item analysis, $t_2(15) = 3.30, p = .005$. Moreover, the context effect found at Regions 5 and 7 could not be attributable to lexical priming, as no priming effect was observed at Region 6, as shown in Figure 6.4. No significant difference existed at this region between the irrelevant-nonnul condition and the relevant-nonnul condition ($ts < 0.39, ps > .70$). In sum, discourse context produced an immediate effect on the Natives' processing of null objects at Region 5, and the effect was stretched to Region 7.

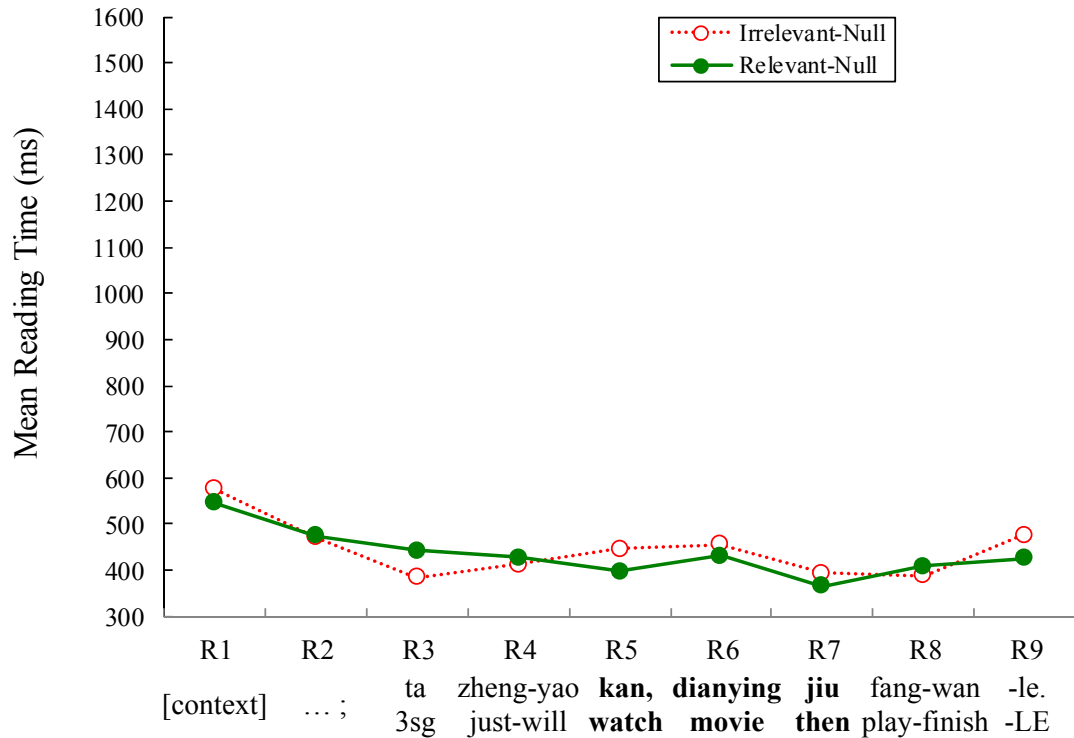


Figure 6.3. Natives' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.

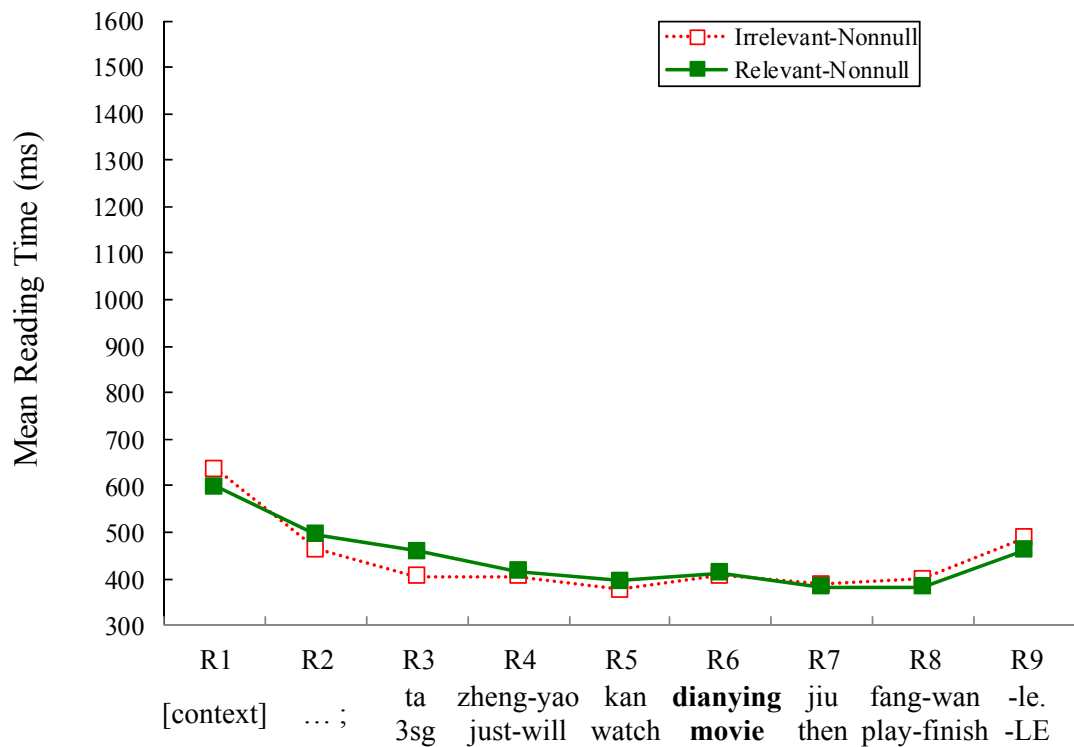


Figure 6.4. Natives' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.

The English Advanced L2 group. The English Advanced L2ers' mean RT profile for the null-object conditions are shown in Figure 6.5. No significant difference was found between the two context conditions at Region 5 ($t_s < 0.44$, $p_s > .66$). At Region 6, the difference became significant in the participant analysis, $t_1(14) = 2.65$, $p = .019$, but not in the item analysis, $t_2(15) = 1.37$, $p = .191$. At Region 7, the difference reached significance in both the participant analysis and the item analysis, $t_1(14) = 3.55$, $p = .003$; $t_2(15) = 3.25$, $p = .005$. The results demonstrate an effect of context on the English Advanced L2ers' processing of null objects in Chinese. Moreover, the context effect cannot be accounted for by a lexical priming effect, because the difference between the irrelevant-nonnull condition and the relevant-nonnull condition at Region 6—see Figure 6.6—was nonsignificant ($t_s < 0.43$, $p_s > .67$). In short, discourse context also influenced the English Advanced L2ers' null-object processing, but the influence was delayed to the post-critical regions (Regions 6 and 7).

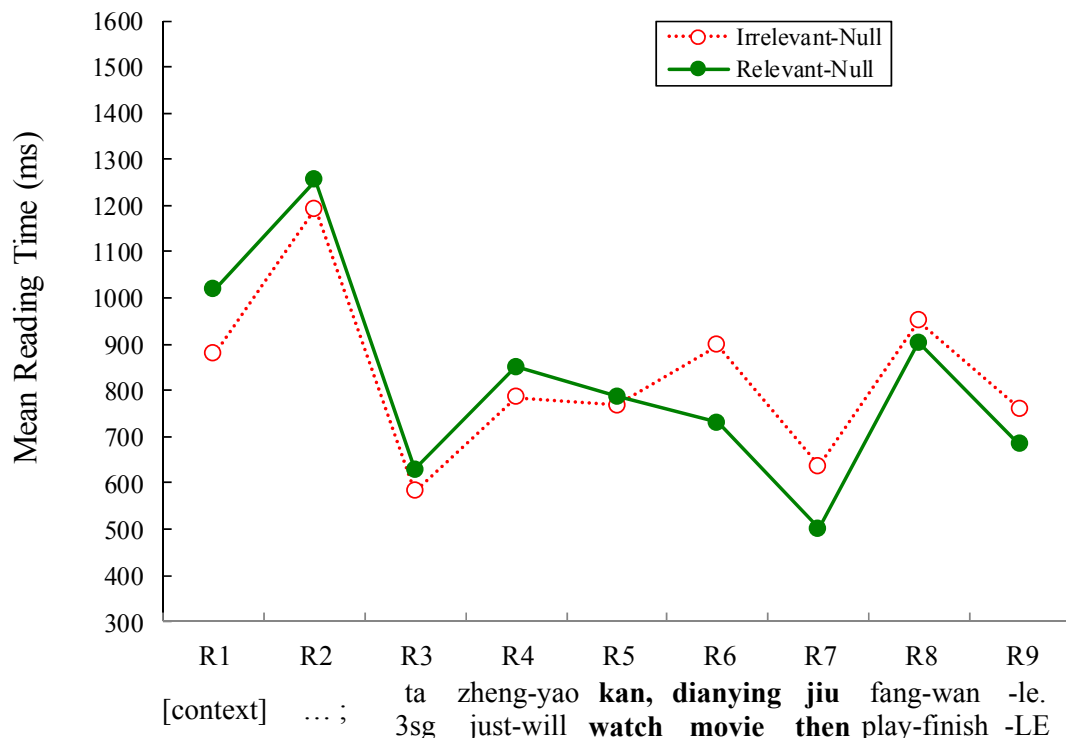


Figure 6.5. English Advanced L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.

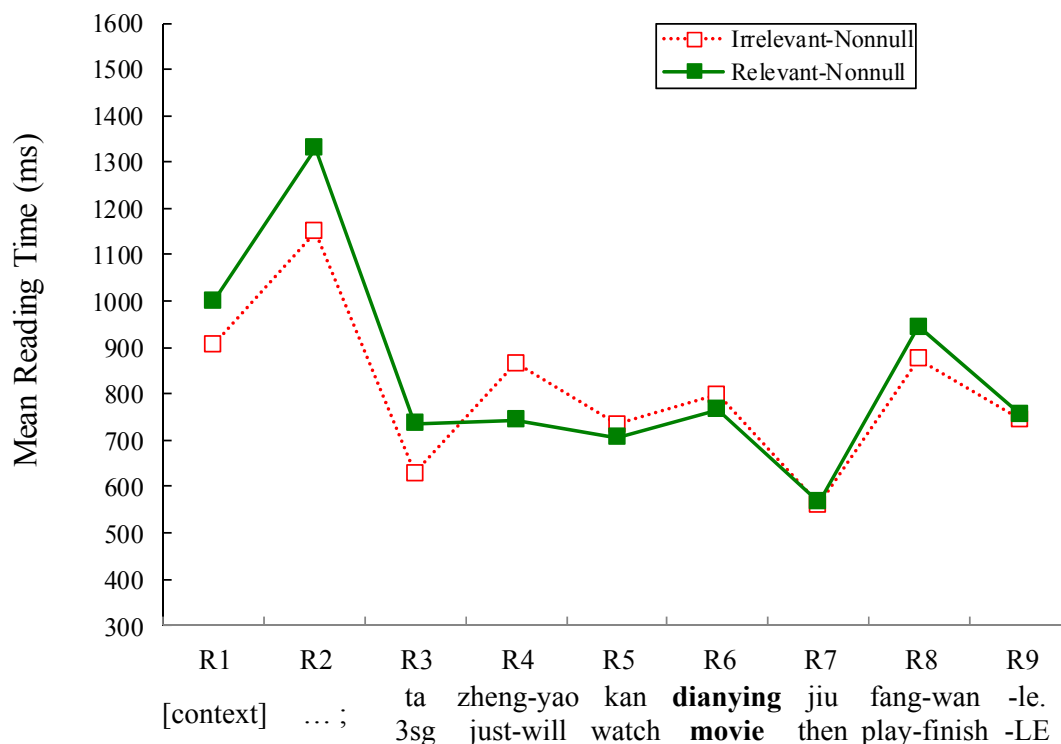


Figure 6.6. English Advanced L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.

The English Intermediate L2 group. Figure 6.7 presents the English Intermediate L2ers' mean RT profile for the null-object conditions. At Region 5, there was no significant difference between the irrelevant-null and relevant-null conditions ($t_s < 0.63$, $p_s > .54$). At Region 6, there was a marginally significant effect of context, $t_1(14) = 2.05$, $p = .060$; $t_2(15) = 1.88$, $p = .080$. At Region 7, the context effect was marginal in the participant analysis, $t_1(14) = 1.91$, $p = .077$, and significant in the item analysis, $t_2(15) = 2.96$, $p = .010$. However, the observed effect of context was confounded with a priming effect—see Figure 6.8—because the difference between the irrelevant-nonnul condition and the relevant-nonnul condition at Region 6 was significant in the participant analysis, $t_1(14) = 2.98$, $p = .010$, although not in the item analysis, $t_2(15) = 1.20$, $p = .247$. In other words, the effect found between the two null-object conditions could actually be due to lexical priming. To sum up, then, because of the confound between discourse-context effects and lexical priming effects, nothing can be concluded about context influencing the English Intermediate L2ers' processing of null objects.

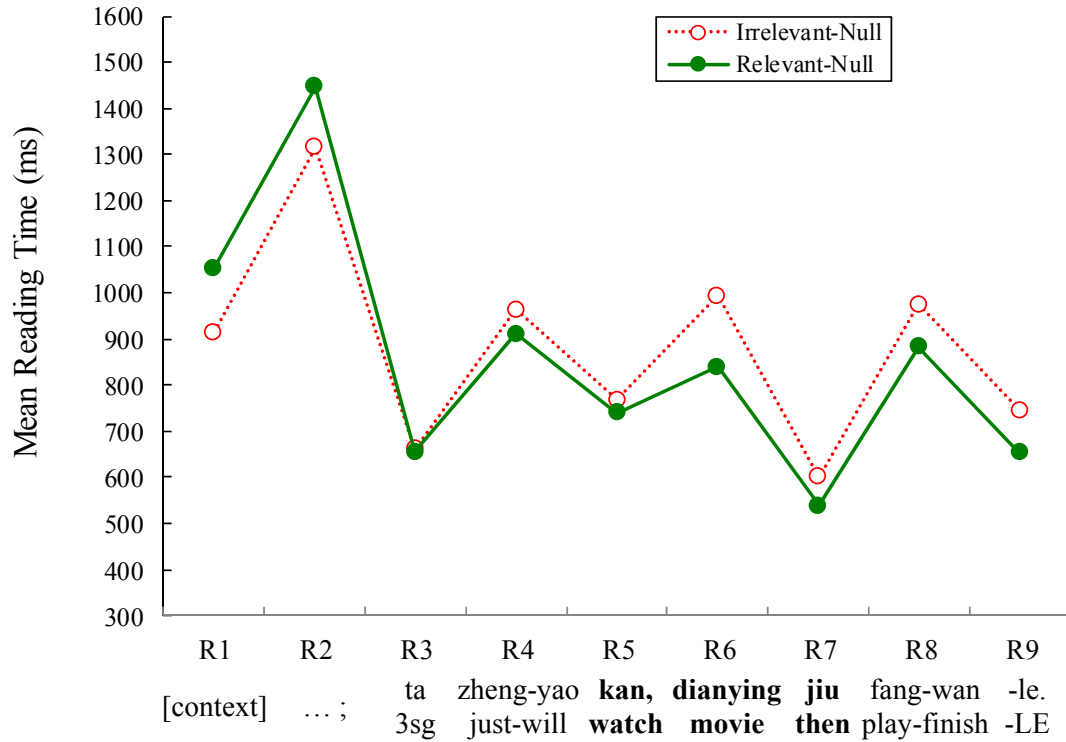


Figure 6.7. English Intermediate L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.

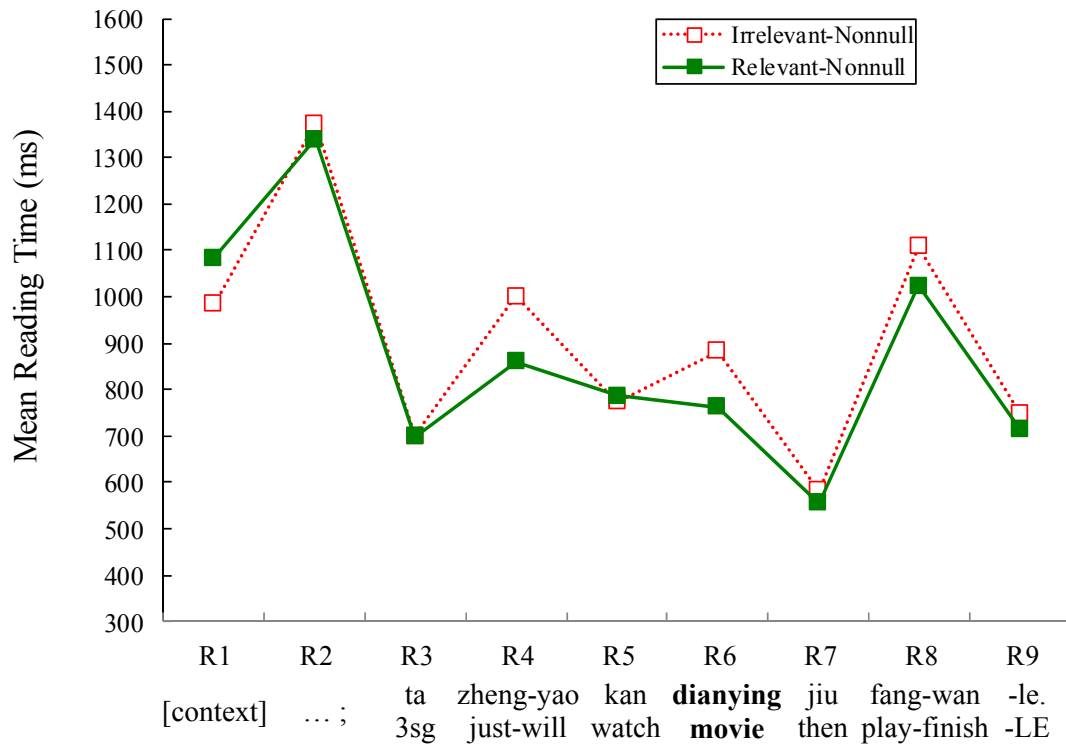


Figure 6.8. English Intermediate L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.

The Japanese Advanced L2 group. The Japanese Advanced L2ers' mean RT profile for the null-object conditions is shown in Figure 6.9. There was a marginal effect of context at Region 5, $t_1(17) = 2.03, p = .058$; $t_2(15) = 1.79, p = .093$ and Region 6, $t_1(17) = 2.06, p = .056$; $t_2(15) = 1.75, p = .100$. At Region 7, the context effect was significant in the participant analysis, $t_1(17) = 2.21, p = .041$, but not in the item analysis, $t_2(15) = 0.97, p = .350$. As indicated in Figure 6.10, this context effect is not attributable to lexical priming: The difference between the irrelevant-nonnull and relevant-nonnull conditions at Region 6 was nonsignificant ($ts < 0.95, ps > .35$). In summary, discourse context showed immediate influence on the Japanese Advanced L2ers' processing of null objects and the influence was protracted to the post-critical regions as well.

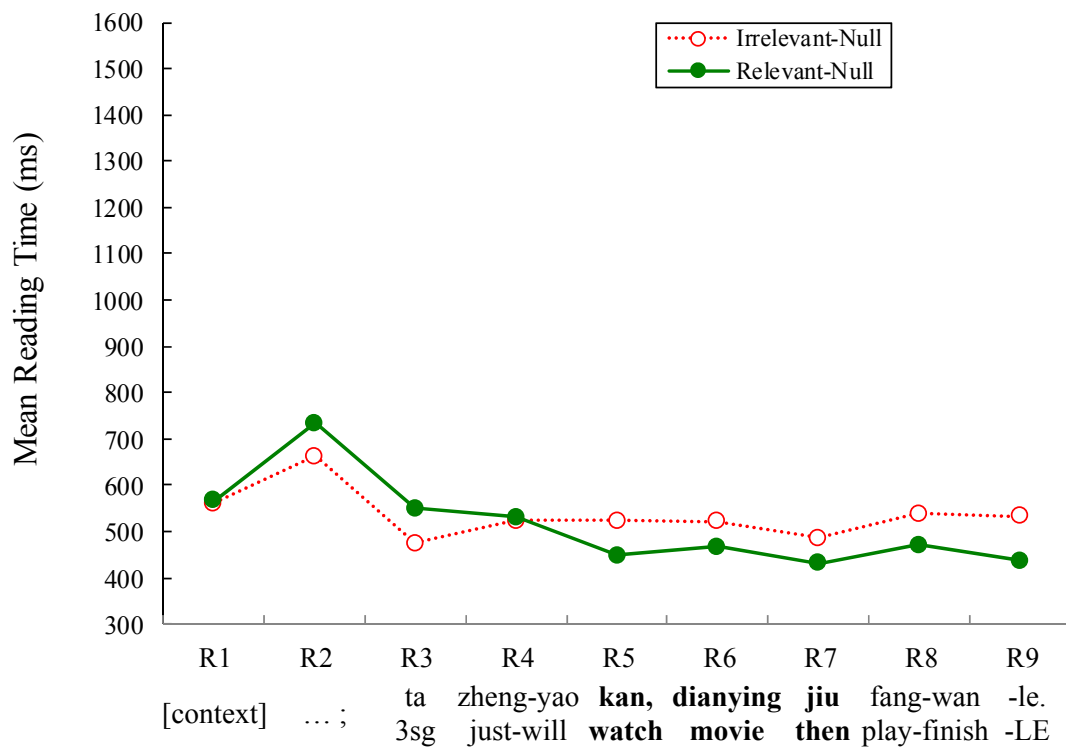


Figure 6.9. Japanese Advanced L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.

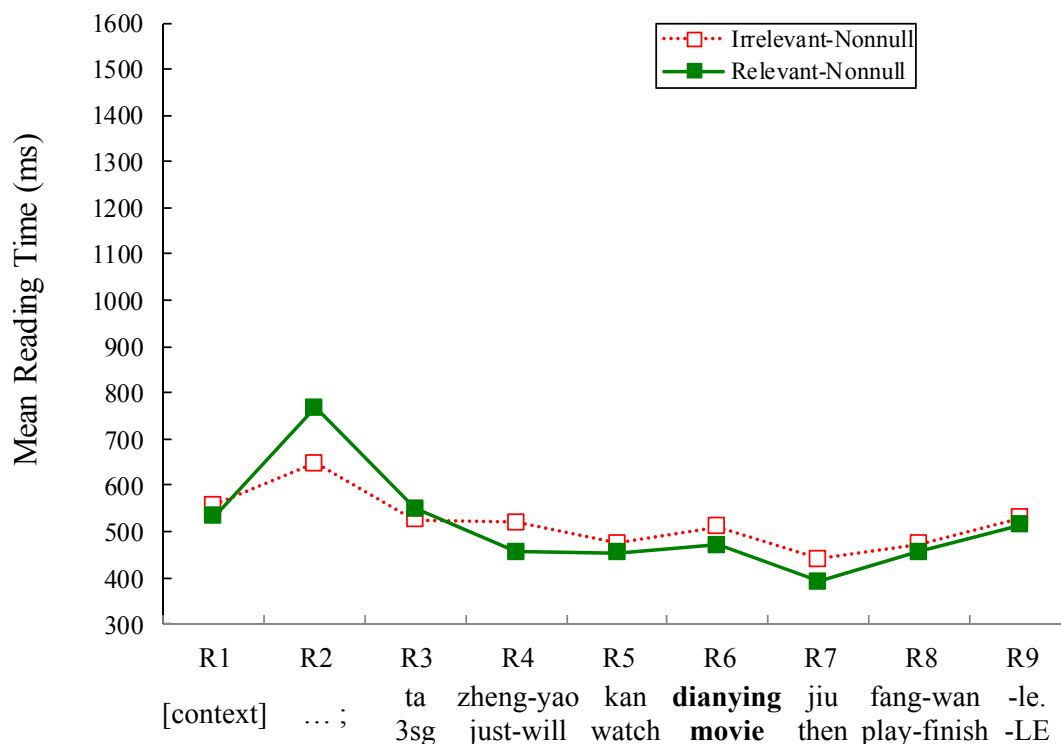


Figure 6.10. Japanese Advanced L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.

The Japanese Intermediate L2 group. The Japanese Intermediate L2ers' mean RT profile for the null-object conditions is shown in Figure 6.11. There was no significant effect of context at Region 5 ($t_s < 1.63$, $p_s > .13$), Region 6 ($t_s < 0.70$, $p_s > .50$), or Region 7 ($t_s < 0.45$, $p_s > .66$). For this group, only some weak priming effect—see Figure 6.12—was observed: The difference between the irrelevant-nonnul and relevant-nonnul conditions at Region 6 was marginal in the participant analysis, $t_1(9) = 2.13$, $p = .062$, and nonsignificant in the item analysis, $t_2(15) = 1.40$, $p = .183$. In brief, the Japanese Intermediate L2ers did not show any discourse-context effect.

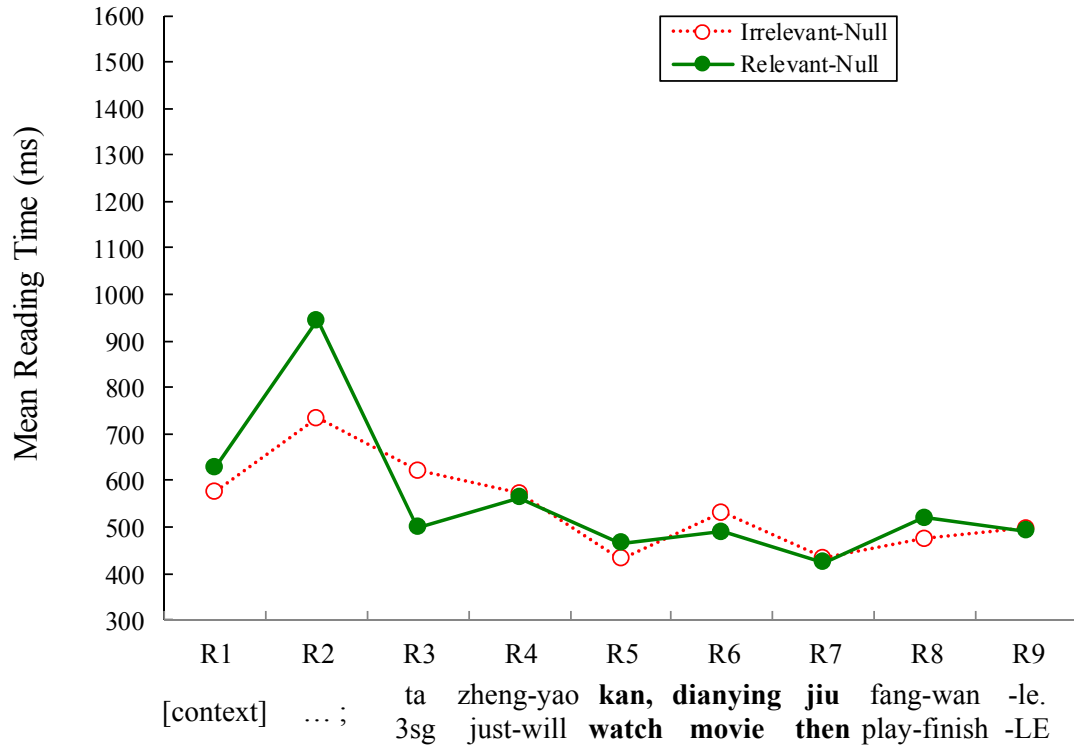


Figure 6.11. Japanese Intermediate L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.1.

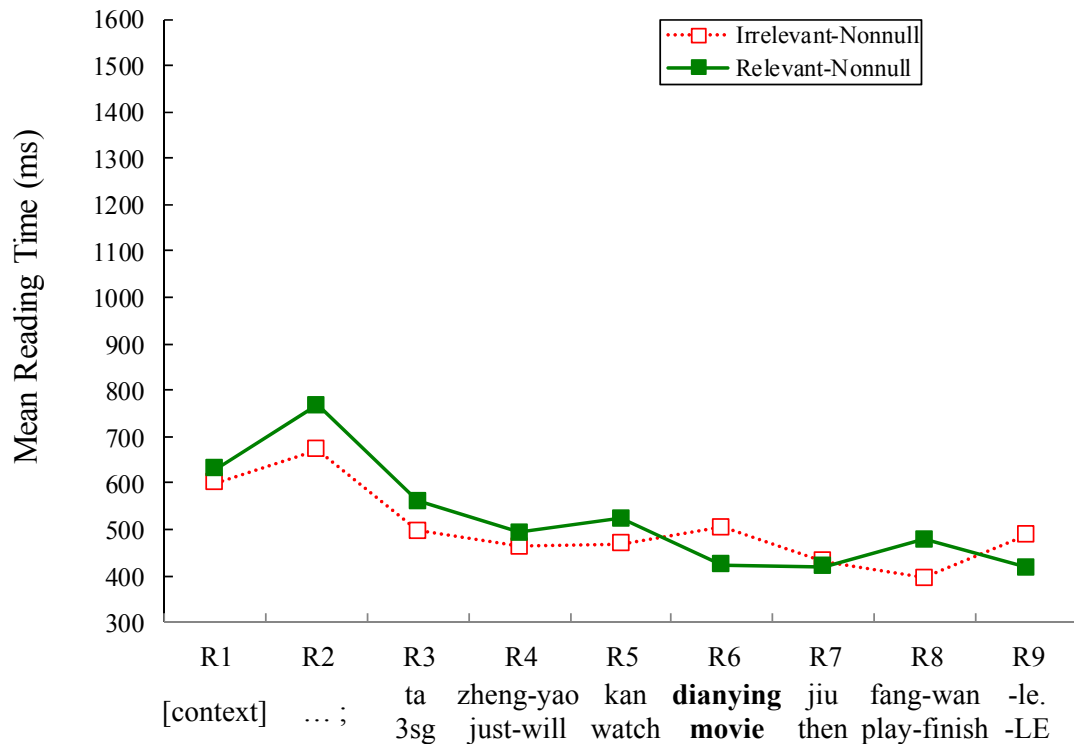


Figure 6.12. Japanese Intermediate L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.1.

A summary of contextual influence on the 5 groups' processing of null objects is presented in Table 6.5.

Table 6.5. Summary of the Main Online RT Results of Experiment 6.1

Group	Influence of discourse context on null-object processing		
	Region 5	Region 6	Region 7
Natives	Yes	No	Some
English Advanced L2ers	No	Some	Yes
English Intermediate L2ers	No	No	No
Japanese Advanced L2ers	Some	Some	Some
Japanese Intermediate L2ers	No	No	No

As can be seen from Table 6.5, at the critical region (Region 5), only the Native group and the Japanese Advanced L2 group showed contextual influence; at the post-critical regions (Regions 6 and 7), both the Native group and the 2 Advanced L2 groups showed some context effects. The Intermediate L2 groups did not show any contextual influence in their null-object processing.

6.4.3 Discussion

The results of Experiment 6.1 showed clear context effects in L1 processing of null objects in Chinese. Felicitous discourse contexts facilitated natives' processing of null objects. Moreover, context effects appeared at the critical region, suggesting that they make use of discourse-context information immediately rather than delaying use of it till the sentence-final interpretation stage. The results lend support to parsing theories that propose an immediate effect of discourse context on early parsing processes (e.g., Altmann et al., 1992; Altmann & Steedman, 1988). The results also underscore the role of discourse-context information in Chinese sentence processing, in line with the findings of studies such as S. Wang et al. (2008) and Zhang et al. (2002). On the other hand, it was not the case that the results revealed the SSH-predicted early context effect for all of the L2 groups, suggesting that integrating discourse-context information with null objects in Chinese was not always an easy task for L2ers; in fact, it was only the Advanced Japanese L2 group that showed a little sensitivity to discourse-context information at the critical region. The English Advanced L2 group as well as the Japanese Advanced group showed some sensitivity at the post-critical regions. Intermediate L2ers did not show any

sensitivity to discourse, although they had no problem understanding the experimental sentences offline. As is clear, L2 processing of discourse-context information is not as easy as the SSH predicts, even in a test condition that is maximally favorable to inducing sensitivity to contextual information. Rather, it is modulated by L2 proficiency. L2ers become better at integrating contextual information with syntactic information as their proficiency rises. The Advanced L2ers' results contrast with those of Pan and Felser (2011) and Pan et al. (2015) and challenge the SSH claim that L2ers necessarily rely more on contextual information in sentence processing than natives do. By contrast, the Intermediate L2ers' results can easily be accommodated by proposals about L2 processing difficulties with coordinating and integrating discourse with syntax at the discourse-syntax interface (Hopp, 2009; Roberts et al., 2008).

In addition to discourse-context effects and proficiency effects, Experiment 6.1 also observed an L1 effect. As was mentioned above, Japanese is a null-argument language like Chinese, and the way discourse context constrains how null objects operate in Japanese is very similar to that in Chinese (Abe, 2004; Huang, 1984). Experiment 6.1 indicates that the Japanese Advanced L2ers had a more native-like processing profile than the English Advanced L2ers did. They showed context effects at both the critical and post-critical regions, whereas the English Advanced L2ers showed context effects at only the post-critical regions. The difference between the 2 Advanced L2 groups cannot be explained by inattentiveness or ignorance on the part of the English Advanced L2ers. Recall that the English Advanced L2ers actually achieved much higher online comprehension accuracy than their Advanced Japanese counterparts (see Figure 6.2). However, the English Advanced L2ers' high comprehension accuracy did not translate into a native-like RT response pattern. Rather, it was the Advanced Japanese L2ers who demonstrated a more native-like RT pattern. The results suggest that structural properties of the L1 of L2ers influence their processing of similar properties in the TL.

One question that arises here is why the Japanese Intermediate L2ers did not show context effects online if they have an L1 advantage. One possible explanation, as the Roberts et al. (2008) study suggested, may be that L2 online processing is cognitively effortful and presupposes a certain level of TL proficiency for L2ers to be able to benefit from the similarities between their L1 and the TL. Another possibility relates to online

task demands. Experiment 6.1 used a CFSPR online task. As Wen and Schwartz (2012) argued (see Chapter 2), online tasks requiring participants to focus on holistic comprehension may encourage L2ers to rely on shallow processing. Since a null object is a syntactic phenomenon, a shallow processing strategy may lead L2ers to become less sensitive to its structural requirements as well as to its connection with the discourse context, and hence a CFSPR task may inhibit Japanese Intermediate L2ers from showing a context effect. This task-demand account also explains why both Intermediate L2 groups failed to show a context effect in processing null objects.³⁵ If this account is on the right track, L2ers might be expected to show better results in an AJSPR task, which encourages fully specified, incremental processing. Notice that this account conforms to L2 proposals appealing to the cost of integrating discourse information with syntactic analysis (Hopp, 2009; Roberts et al., 2008). Since the SSH hypothesizes that L2 processing of discourse-context information belongs to semantics-based comprehension strategies (Pan & Felser, 2011), it predicts that regardless of whether or not an online task encourages structural processing, this should not matter to L2ers, whose processing supposedly lacks structural details and hence cannot benefit from tasks, such as AJSPR, that encourage fully specified, incremental processing. If L2ers show stronger context effects in an AJSPR task, this would substantiate the task-demand account here and confirm that L2 processing of discourse information is not merely meaning-based but rather involves the simultaneous processing of both discourse-level information and syntactic information. Experiment 6.2 seeks to verify the task-demand account of discourse-context effects on (L1 and) L2 processing of null objects in Chinese.

6.5 Experiment 6.2: Acceptability-Judgment Self-Paced Reading

This experiment examines contextual effects on L1 and L2 processing in an AJSPR task. Since AJSPR is hypothesized to encourage fully specified, incremental processing, we predict that in processing null objects in Chinese, both L1 and L2 groups will show stronger context effects in this experiment than in Experiment 6.1. For L2ers in particular, both of the 2 advanced-level groups should show sensitivity to discourse-context information at the critical region, like the Native group.

³⁵ We momentarily ignore the English Intermediate L2ers' priming effect in the nonnull-object conditions.

6.5.1 Method

6.5.1.1 Participants

Participants for this experiment included 32 adult L1-English learners, 32 adult L1-Japanese learners, and 25 adult L1-Chinese controls. Interruptions during testing sessions resulted in the exclusion of 2 participants: one Chinese and one Japanese. Furthermore, due to relative poor performance for the special type of filler (designed for assessing attentiveness—see Chapter 3), 2 L1-English participants ($M_{\text{error rate}} = 50\%$) and one L1-Japanese participant ($M_{\text{error rate}} = 62.5\%$) were also excluded. No participant of Experiment 6.1 took part in this experiment, and all participants received a small payment for taking part.

As in Experiment 6.1, the L2 participants were recruited from two different language backgrounds (i.e., L1-English and L1-Japanese) for the purpose of examining possible L1 effects. L2ers were further divided according to proficiency level based on the quasi-C-test. To ensure that the proficiency groups in this experiment are comparable to those in Experiment 6.1, the cutoff point is set at the same level as in Experiment 6.1 (i.e., 21). Those L2ers who scored 21 and above were grouped as Advanced (English: $n = 16$; Japanese: $n = 22$). Those who scored less than 21 were grouped as Intermediate (English: $n = 14$; Japanese: $n = 8$). Table 6.6 gives the participants' Chinese background information.

Table 6.6. *Chinese Background Information of the Participants in Experiment 6.2*

Group	Years of learning Chinese		Years of residence in China/Taiwan		Chinese proficiency scores (Max = 50)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Natives ($n = 24$)	N/A	N/A	N/A	N/A	45.9	1.4
English Advanced L2ers ($n = 16$)	4.2	3.7	1.2	1.4	26.1	5.4
Japanese Advanced L2ers ($n = 22$)	3.1	2.5	2.6	3.3	30.4	8.3
English Intermediate L2ers ($n = 14$)	2.9	2.6	1.2	1.3	15.7	3.7
Japanese Intermediate L2ers ($n = 8$)	2.4	1.0	0.7	0.3	17.5	3.2

The 5 groups' Chinese proficiency levels differed from each other significantly, $F(4, 79) = 94.54, p < .001$. SPSS Bonferroni corrected post hoc tests showed that the L1 group differed from all the L2 groups ($ps < .001$) and that the Advanced L2 groups had significantly higher Chinese proficiency than the Intermediate L2 groups ($ps \leq .003$). The

2 Advanced L2 groups did not significantly differ from each other ($p = .144$), and neither did the 2 Intermediate L2 groups ($p = 1.000$).

6.5.1.2 Materials

This experiment used exactly the same 16 sets of experimental sentences from Experiment 6.1. Using the same sentences makes it possible for us to compare the results from the two experiments. Since all the experimental sentences are grammatical, it is necessary to include ungrammatical fillers; otherwise, the AJSPR task cannot work. The fillers of this experiment comprised 32 ungrammatical and 40 grammatical sentences, representing a variety of syntactic forms and sentence types: Some were simple sentences, and some were complex sentences; some were preceded by a context sentence, and some were not. One type of the fillers ($k = 8$; ungrammatical) was specifically devoted to assessing participants' attentiveness in performing the online task (see Chapter 3). Note that the other fillers included experimental sentences for some other studies, and thus the results of them will not be reported in detail here.

The offline task used in this experiment was exactly the same as that in Experiment 6.1. Using the same offline task makes it possible for us to assess whether the participants in the two experiments had comparable knowledge of null objects.

6.5.1.3 Procedure

This experiment used an AJSPR task (for a description of the AJSPR task, see Chapter 3; for the specific instructions given in this task, see Appendix I). After reading each sentence, participants were prompted to judge whether the sentence is acceptable in terms of its well-formedness. Other than this, the procedure is identical to that of Experiment 6.1. Participants signaled their acceptability-judgment decisions by pressing the designated *Tongshun* ('well-formed, acceptable') or *Bu tongshun* ('not well-formed, not acceptable') key on the computer keyboard. The self-paced reading task was followed by a language background questionnaire (see Appendix E) and the Chinese proficiency test (see Appendix F).

Note that depending on the specific experimental stimuli involved, AJSPR can focus participants not only on the linguistic form of the stimuli but also on their

propositional content and the pragmatic appropriateness. This is because participants in this task are not explicitly told to judge the *grammaticality* of the stimuli. This practice avoids participants' using metalinguistic knowledge, but on the other hand, they have to detect everything unusual in a test sentence (including, e.g., structural, propositional, and pragmatic aspects of the sentence). In short, AJSPR encourages fully specified, incremental processing of all aspects of experimental stimuli.

6.5.2 Results

6.5.2.1 Offline Task Results

Figure 6.13 shows the 5 participant groups' comprehension accuracy rates for the experimental sentences in the offline task (see Appendix AF for the descriptive statistics).

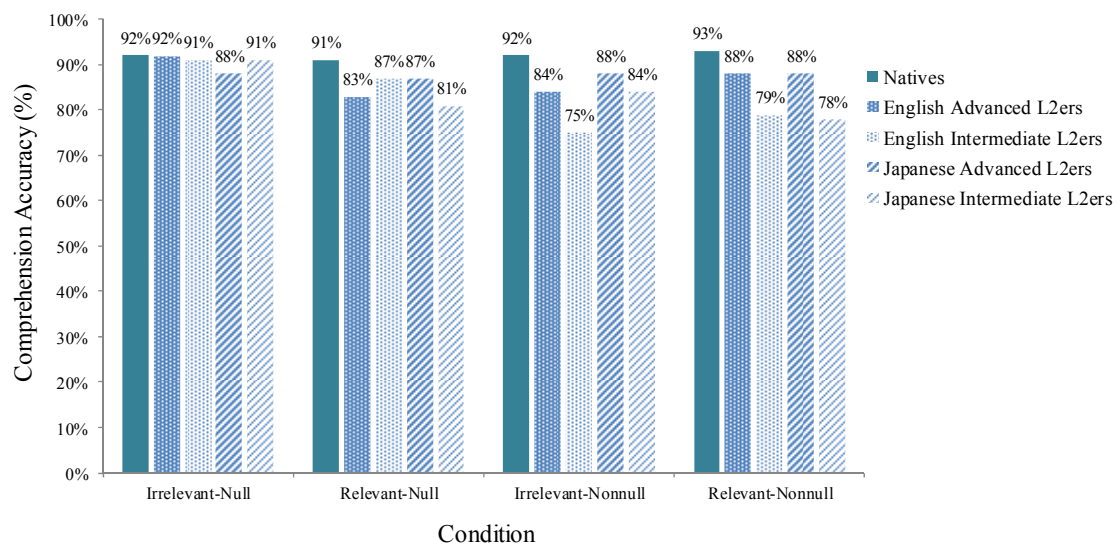


Figure 6.13. Mean Comprehension Accuracy Rates for the Experimental Sentences in the Offline Task of Experiment 6.2.

As is clear from Figure 6.13, all the participant groups' offline comprehension accuracy was high (all mean accuracy rates $\geq 75\%$), suggesting that participants could understand the target sentences well. All L2 groups accurately responded to at least 81% of the comprehension questions in the null-object conditions, which indicates that null objects did not pose a problem for their comprehension.

A $5 \times 2 \times 2$ (Group [Native, English Advanced L2, English Intermediate L2, Japanese Advanced L2, Japanese Intermediate L2] \times Context [irrelevant, relevant] \times Object [null, nonnull]) repeated measures ANOVA performed on the arcsine-transformed accuracy data showed a main effect of object, $F(1, 79) = 3.99, p = .049, \eta_p^2 = .05$, and a marginal Context \times Object interaction, $F(1, 79) = 3.63, p = .060, \eta_p^2 = .04$. No other main effect or interaction reached statistical significance ($F_s < 1.53, p_s > .22$). SPSS Bonferroni corrected post hoc tests showed that, overall, the participants did better in the null-object conditions than in the nonnull-object conditions ($p = .049$), again most likely caused by the temporary object-subject ambiguity (see Section 6.4.2.1).

In order to make comparisons between the online results of Experiment 6.1 and Experiment 6.2 later on, we first need to establish that the participants in the two experiments are comparable in terms of their knowledge of null objects in Chinese. This can be done by inspecting the offline accuracy rates in the two experiments. For this we conducted—for every 2 groups matched in L1 and Chinese proficiency—a 2×4 (Task [CFSPR, AJSPR] \times Condition [irrelevant-null, relevant-null, irrelevant-nonnul, relevant-nonnul]) repeated measures ANOVA on the arcsine-transformed accuracy data. Across the two experiments, no main effect or interaction was found for the 2 Native groups ($F_s < 1.89, p_s > .14$), for the 2 English Advanced L2 groups ($F_s < 1.63, p_s \geq .20$), for the 2 Japanese Advanced L2 groups ($F_s < 0.95, p_s > .41$), or for the 2 Japanese Intermediate L2 groups ($F_s < 2.22, p_s > .15$). The 2 English Intermediate L2 groups, however, showed a main effect of condition, $F(3, 81) = 5.02, p = .006, \eta_p^2 = .16$, but no main effect of task or an interaction of Task \times Condition ($F_s < 0.37, p_s \geq .55$). Crucially, none of the matched-group comparisons showed a main effect of task or a Task \times Condition interaction. Since the offline accuracy rates in the two experiments were not significantly different from each other, we conclude that the participants in the two experiments are comparable in their knowledge of null objects in Chinese.

6.5.2.2 Online Task Results

As in Experiment 6.1, data analysis was performed on both participants' accuracy and online RTs.

Table 6.7 presents the 5 participant groups' mean acceptability-judgment accuracy rates of the filler sentences.

Table 6.7. Mean Acceptability-Judgment Accuracy Rates of the Fillers in Experiment 6.2

Group	Type		
	Special (ungramm; <i>k</i> = 8)	Other (gramm; <i>k</i> = 40)	Other (ungramm; <i>k</i> = 24)
Natives (<i>n</i> = 24)	.91 (.12)	.77 (.08)	.76 (.14)
English Advanced L2ers (<i>n</i> = 16)	.94 (.09)	.69 (.10)	.64 (.16)
English Intermediate L2ers (<i>n</i> = 14)	.92 (.09)	.73 (.11)	.51 (.16)
Japanese Advanced L2ers (<i>n</i> = 22)	.89 (.13)	.80 (.10)	.48 (.17)
Japanese Intermediate L2ers (<i>n</i> = 8)	.83 (.15)	.75 (.06)	.45 (.26)

Note. Standard deviations are in parentheses. ungramm = ungrammatical; gramm = grammatical.

For the special type of filler (see Chapter 3), the participants' acceptability judgments on average were accurate at least 83% of the times, indicating that all groups were attentive in performing the reading task. The data here also serve to demonstrate that the AJSPR task worked well in this experiment. As for the other fillers, the participants generally did fairly well for the grammatical ones but not quite so well for the ungrammatical ones.

For the experimental sentences, the acceptability-judgment accuracy rates are shown in Figure 6.14 (for the descriptive statistics, see Appendix AG). Note that we will refer to the acceptability-judgment accuracy in this experiment as *acceptance*. This is because all the experimental sentences are grammatical, and the acceptability-judgment accuracy actually indicates participants' accepting a sentence as well-formed, coherent, and appropriate.

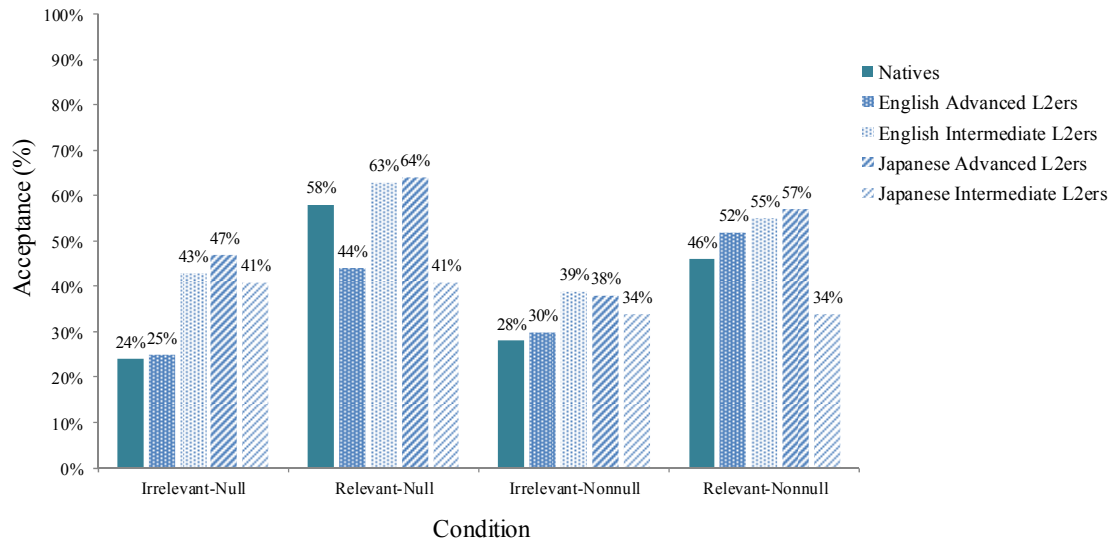


Figure 6.14. Mean Acceptance Rates for the Experimental Sentences in Online Task of Experiment 6.2.

A $5 \times 2 \times 2$ (Group [Native, English Advanced L2, English Intermediate L2, Japanese Advanced L2, Japanese Intermediate L2] \times Context [irrelevant, relevant] \times Object [null, nonnull]) repeated measures ANOVA performed on the arcsine-transformed acceptance data yielded a marginal main effect of group, $F(4, 79) = 2.05, p = .095, \eta_p^2 = .09$, a main effect of context, $F(1, 79) = 36.40, p < .001, \eta_p^2 = .32$, and a marginal Group \times Context interaction, $F(4, 79) = 2.08, p = .091, \eta_p^2 = .10$. No other main effect or interaction reached statistical significance ($F_s < 1.45, p_s > .23$). SPSS Bonferroni corrected post hoc tests did not reveal any significant difference between groups ($p_s \geq .338$). The overall acceptance rates for the relevant-context conditions were significantly higher than for the irrelevant-context conditions ($p < .001$) (with the exception of the Japanese Intermediate group, which showed no difference), suggesting that supportive contexts raise both natives' and L2ers' acceptance rates of the experimental sentences. The results here are important: On the one hand, they reveal that discourse context influences participants' parsing of null objects in Chinese; on the other hand, they serve to demonstrate that the AJSPR task of this experiment is sensitive to the experimental manipulations and thus is a valid and reliable online task.

Notice that, overall, the participants' online acceptance rates for the experimental sentences were not high, even for the relevant-context conditions. Given that all of the

experimental sentences are grammatical and the participants achieved overall high offline comprehension accuracy, the relatively low online acceptance rates are rather surprising. We will return to this issue in the Discussion section (Section 6.5.3).

We turn now to the RT data, where the analyses focused, as in Experiment 6.1, on Regions 5, 6, and 7 for the two null-object conditions and Region 6 for the two nonnull-object conditions. Recall that the null-object conditions were used for examining discourse-context effects, and the nonnull-object conditions were used for checking lexical priming effects (so as to see whether any ostensible discourse-context effect from the null-object conditions could instead be due to lexical priming; see Section 6.4.1.2).

Raw RTs were trimmed before they were submitted to statistical tests. We applied the same data-trimming steps as in Experiment 6.1. The trimming affected 6.0% of the Natives' data, 14.2% of the English L2ers' data, and 8.8% of the Japanese L2ers' data. (The trimmed raw RT means are given in Appendix AH.)

As in Experiment 6.1, the data analysis included all the trials rather than only those to which the participants responded correctly, because we cannot establish that the incorrect trials were caused by participants' inattentiveness in performing the online task or their ignorance of null objects in Chinese (see the arguments in Section 6.4.2.2). Furthermore, since all participant groups' acceptance rates for the experimental items were relatively low, removing such a large proportion of the data may lead to noisy or bizarre results (see the notes in Sections 4.6.2.2 and 6.4.2.2 about the consequences of removing a large proportion of data).

The 5 groups' mean RT profiles are examined separately. Paired-samples *t* tests (two-tailed) were conducted on the mean RTs at each region of interest (Regions 5–7 for the null-object conditions; Region 6 for the nonnull-object conditions). For the *t* tests, we report both by-participant (t_1) and by-item (t_2) analyses. Following the convention of reading-time studies, the alpha level is set at .05 for the tests.

The Native group. This group's mean RT profile for the null-object conditions is shown in Figure 6.15. As can be seen, the Natives spent much less time at Regions 5–7 for the relevant-null condition than for the irrelevant-null condition. The difference between the two conditions was statistically significant: at Region 5, $t_1(23) = 2.60$, $p = .016$; $t_2(15) = 2.27$, $p = .038$, Region 6, $t_1(23) = 2.72$, $p = .012$; $t_2(15) = 2.70$, $p = .016$,

and Region 7, $t_1(23) = 2.99, p = .007$; $t_2(15) = 3.31, p = .005$. These results indicate that discourse context significantly affected the Natives' null-object processing. The context effect is not attributable to a priming effect—see Figure 6.16—because the difference between the irrelevant-nonnull and relevant-nonnull conditions at Region 6 was nonsignificant ($t_s < 0.98, p_s > .34$). Taken together, the Native group's RT data thus indicate an immediate and lasting effect of discourse context on Natives' null-object processing.

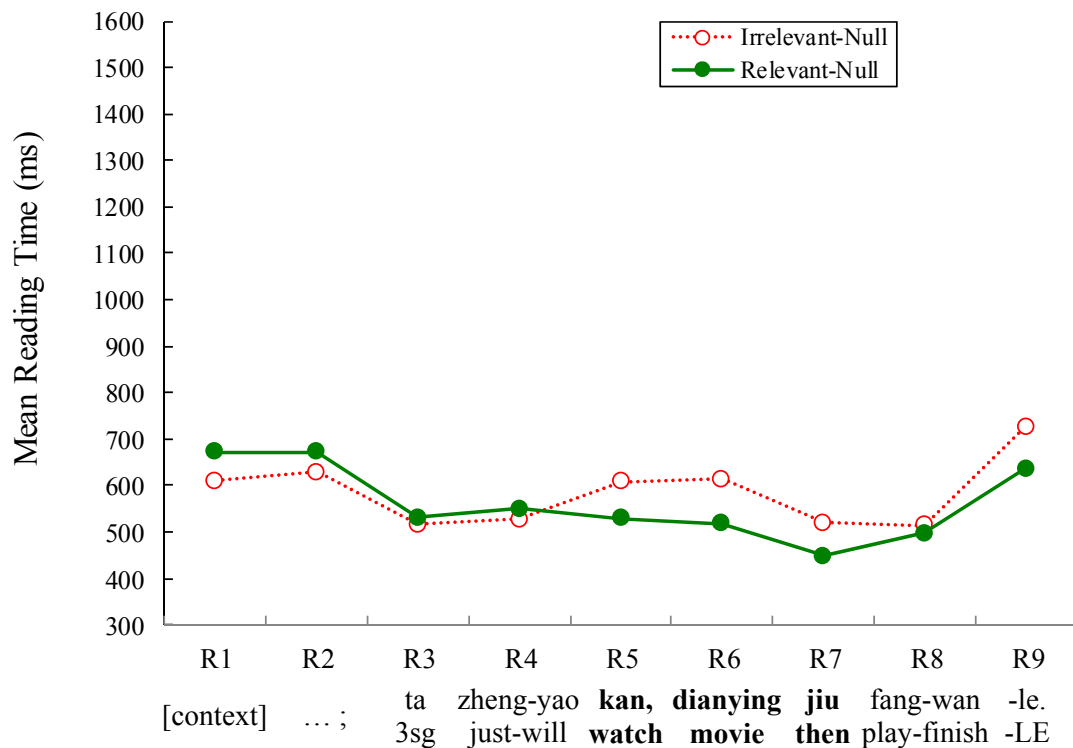


Figure 6.15. Natives' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.2.

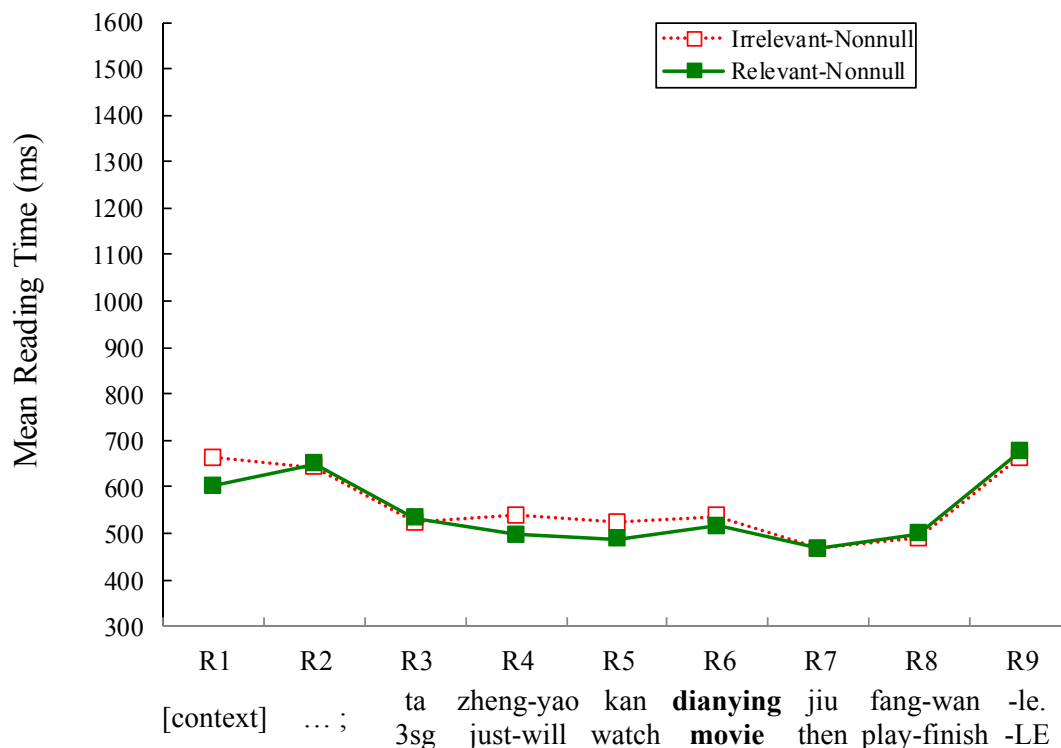


Figure 6.16. Natives' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.2.

The English Advanced L2 group. Figure 6.17 presents this group's mean RT profile for the null-object conditions. There was a significant difference between the irrelevant-null and relevant-null conditions at Region 5, $t_1(15) = 2.20, p = .044$; $t_2(15) = 2.16, p = .048$, Region 6, $t_1(15) = 4.55, p < .001$; $t_2(15) = 3.46, p = .004$, and Region 7 $t_1(15) = 2.62, p = .019$; $t_2(15) = 2.47, p = .026$, indicating that supportive discourse contexts significantly reduced the English Advanced L2ers' processing time of the null-object sentences. The influence of discourse context on this group's null-object processing was not caused by priming—see Figure 6.18—as no significant difference between the irrelevant-nonnul and relevant-nonnul conditions was found at Region 6 ($ts < 0.27, ps > .79$). In brief, discourse context has an immediate and lasting effect for English Advanced L2ers' null-object processing.

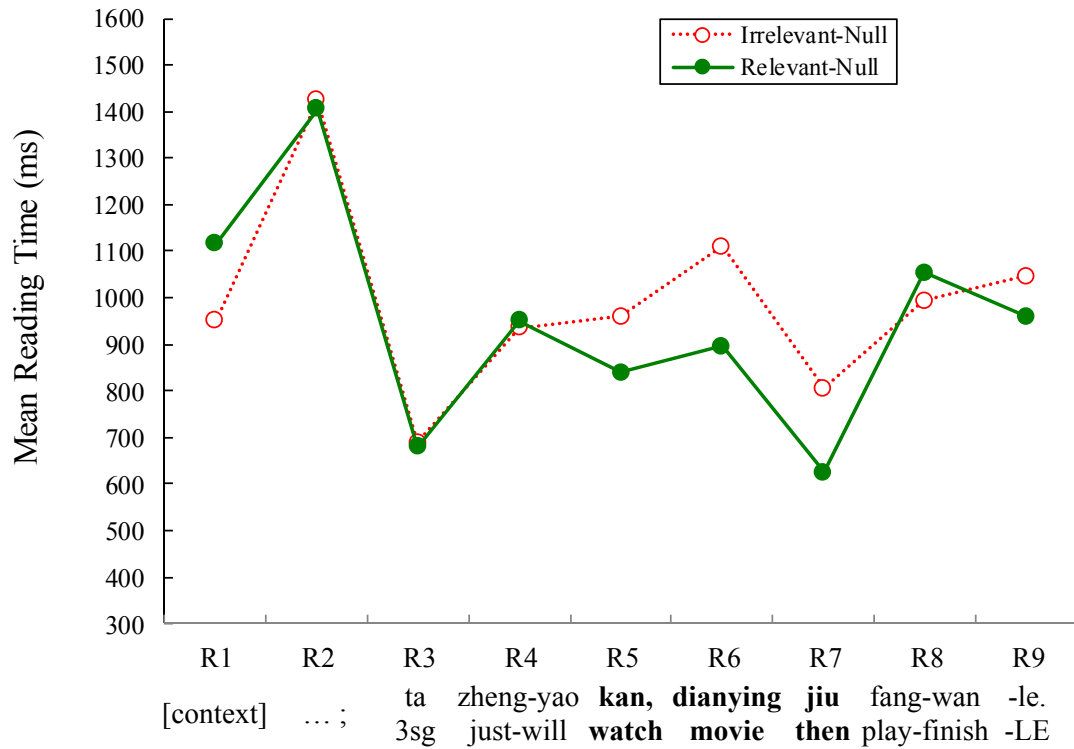


Figure 6.17. English Advanced L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.2.

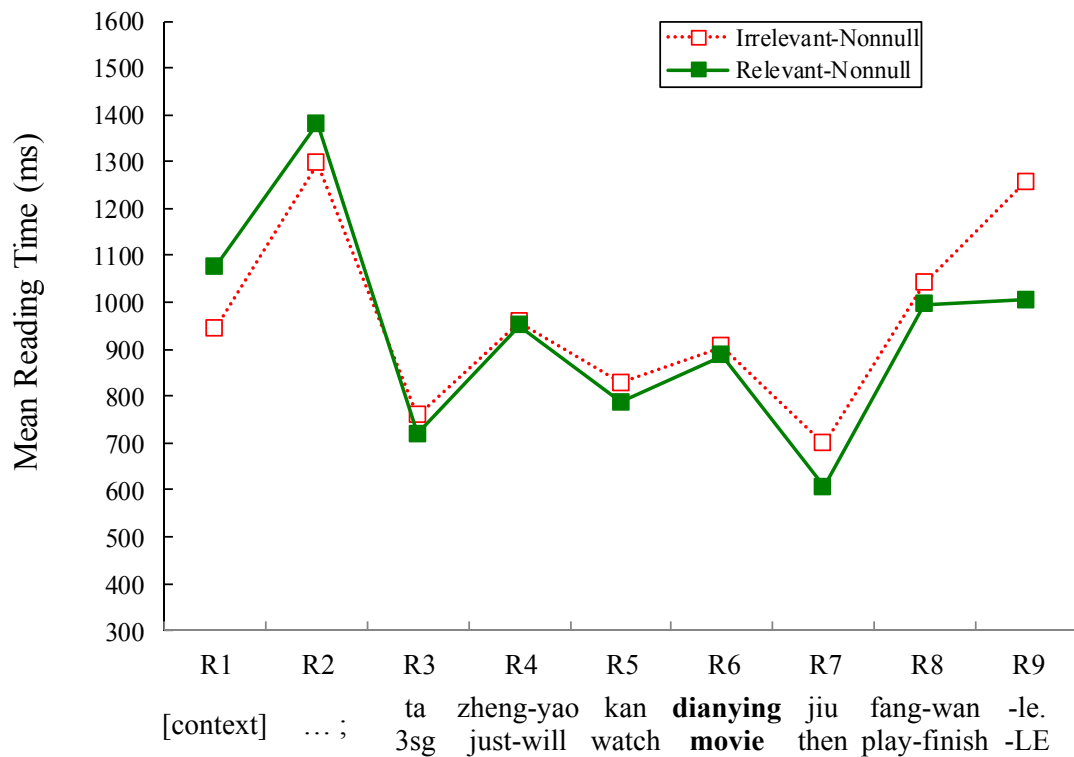


Figure 6.18. English Advanced L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.2.

The English Intermediate L2 group. Figure 6.19 displays the English Intermediate L2 group's mean RT profile for the null-object conditions. The difference between the irrelevant-null and relevant-null conditions was not significant at Region 5, but it reached marginal significance at Region 6, $t_1(13) = 1.89, p = .082$; $t_2(15) = 1.87, p = .081$; the effect did not last to Region 7 ($ts < 0.43, ps > .67$). This group's mean RT profile for the nonnull-object conditions is presented in Figure 6.20. No significant difference between the irrelevant-nonnul and relevant-nonnul conditions was observed at Region 6 ($ts < 1.02, ps > .32$), which means that the observed context effect (as in Figure 6.19) is not attributable to a priming effect. In sum, discourse context exerts some influence on English Intermediate L2ers' processing of null-object sentences.

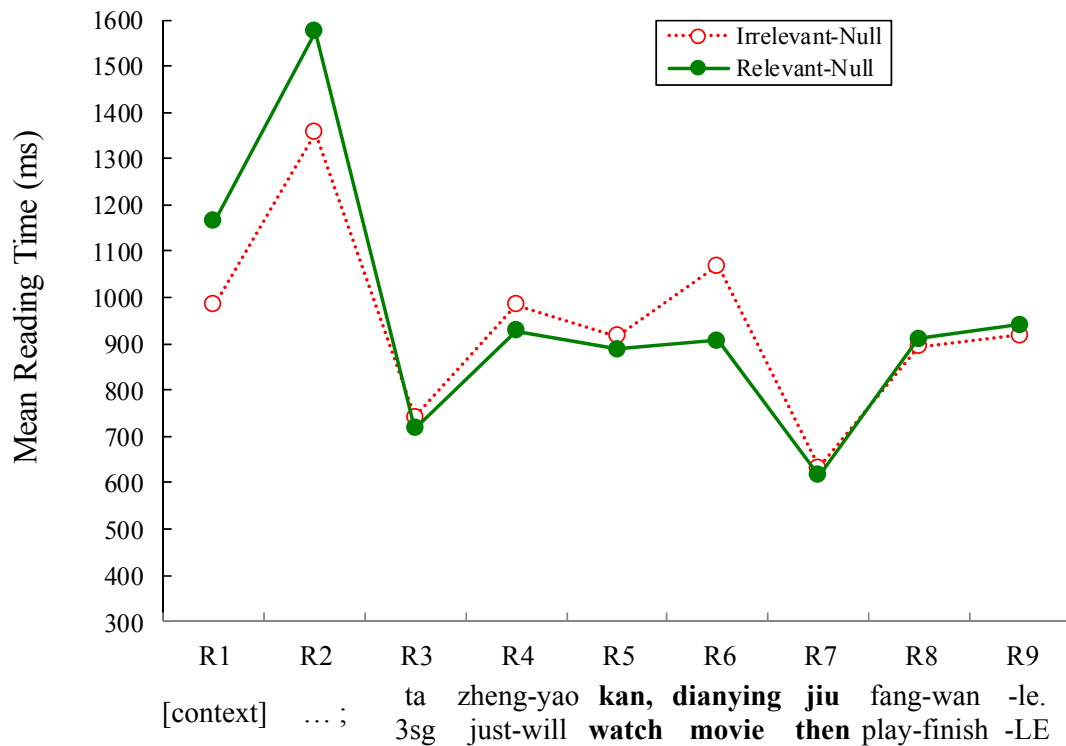


Figure 6.19. English Intermediate L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.2.

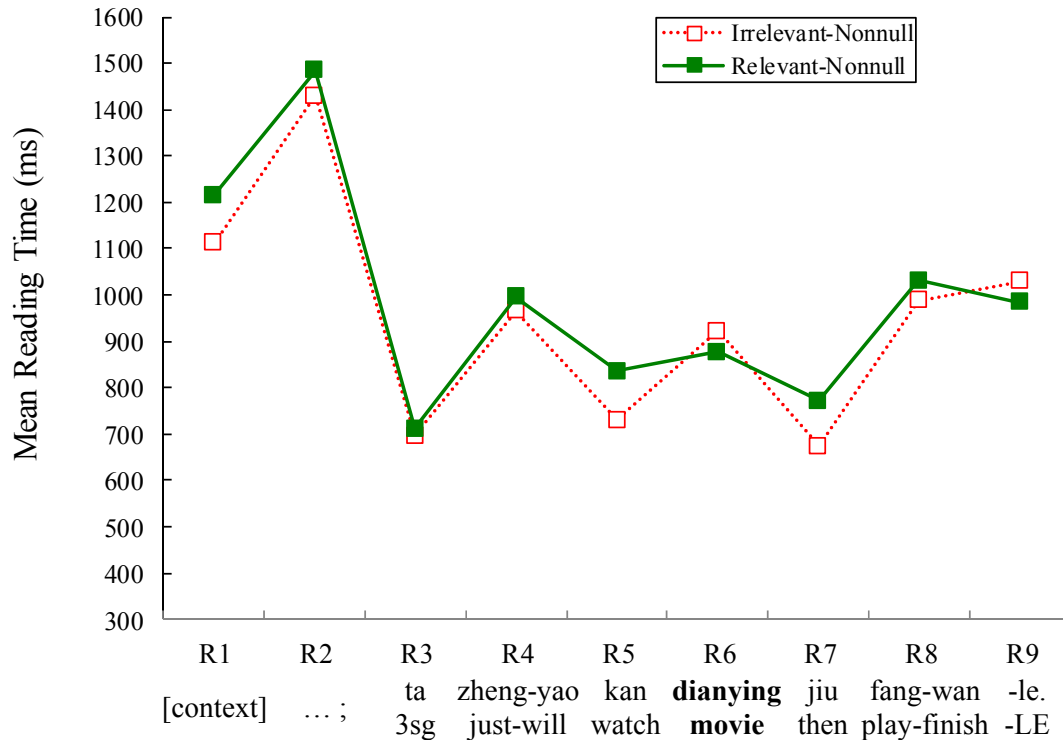


Figure 6.20. English Intermediate L2ers’ Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.2.

The Japanese Advanced L2 group. Figure 6.21 presents this group’s mean RT profile for the null-object conditions. The difference between the irrelevant-null and relevant-null conditions at Region 5 was significant in the participant analysis, $t_1(21) = 3.56, p = .002$, and marginal in the item analysis, $t_2(15) = 2.10, p = .053$, which indicates that relevant discourse context facilitated the Japanese Advanced L2ers’ processing of null-object sentences. At Region 6, the difference between the two conditions was not significant in the participant analysis, $t_1(21) = 1.51, p = .147$, but significant in the item analysis, $t_2(15) = 3.02, p = .009$. At Region 7, the difference, although numerically large, was not significant in either the participant analysis or the item analysis, $t_1(21) = 1.32, p = .201; t_2(15) = 1.08, p = .296$. Figure 6.22 presents the results of the irrelevant-nonnul and relevant-nonnul conditions; since the difference between the two conditions at Region 6 was nonsignificant ($ts < 1.16, ps \geq .26$), the observed context effect (as in Figure 6.21) was not due to lexical priming. In sum, discourse context affects Japanese Advanced L2ers’ processing of null objects, although the context effect is not so robust.

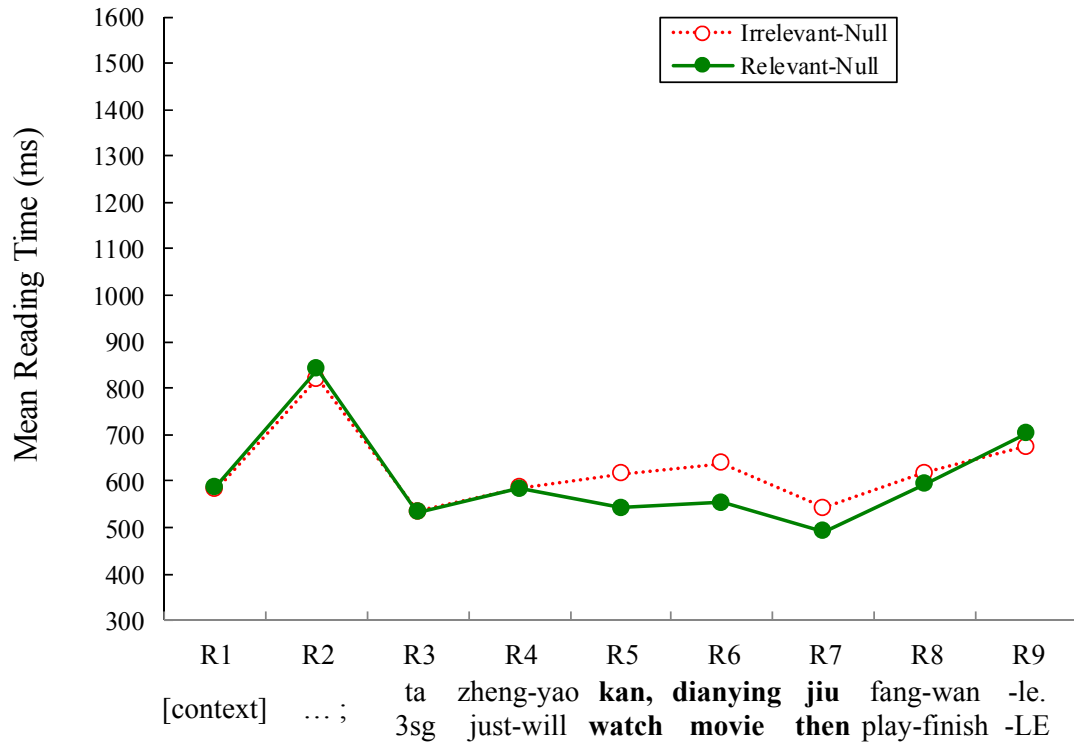


Figure 6.21. Japanese Advanced L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.2.

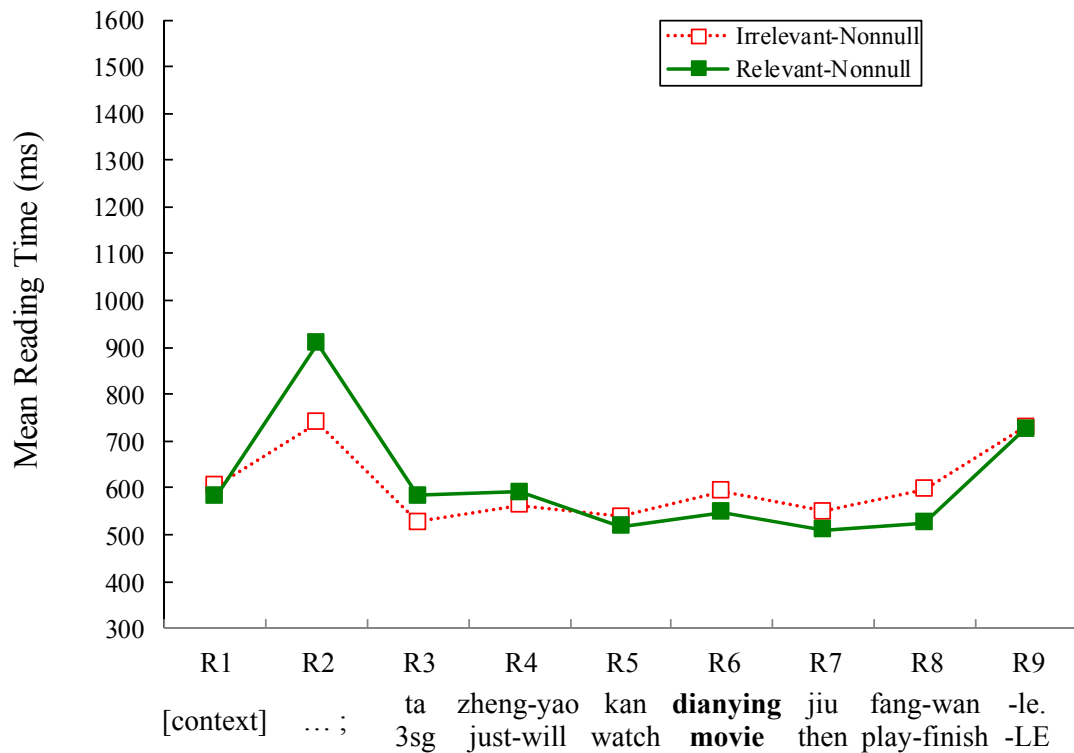


Figure 6.22. Japanese Advanced L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.2.

The Japanese Intermediate L2 group. The Japanese Intermediate L2 group's mean RT profile for the null-object conditions is presented in Figure 6.23. The difference between the irrelevant-null and relevant-null conditions was not significant at Region 5 ($ts < 0.83, ps > .42$) or Region 6 ($ts < 0.77, ps > .45$); but at Region 7, it became marginal in the participant analysis, $t_1(7) = 2.07, p = .078$, and significant in the item analysis, $t_2(15) = 2.44, p = .028$. The difference between the irrelevant-nonnul and relevant-nonnul conditions at Region 6—see Figure 6.24—was not significant, $t_1(7) = 1.75, p = .124$; $t_2(15) = 1.58, p = .134$, although numerically large. This nonsignificant difference might be taken to indicate that the observed context effect (see Figure 6.23) was not attributable to lexical priming. Yet, we can see in Figure 6.24 that the difference between the irrelevant-nonnul and relevant-nonnul conditions lingers and seems to become larger at Region 7; statistically, this difference was marginal in the participant analysis, $t_1(7) = 2.21, p = .063$, and significant in the item analysis, $t_2(15) = 3.00, p = .009$. It is a little difficult to explain why the difference occurred at Region 7. It may be a spillover effect of priming from Region 6; however, receptive lexical priming typically does not spill over to subsequent regions. Since the Japanese Intermediate L2 group has only a very small number of participants ($n = 8$), the difference at Region 7 may simply be a spurious effect. Nevertheless, this limits us from being able to make an unequivocal claim about the observed “context” effect from Region 7 of the null-object conditions. In short, the Japanese Intermediate L2ers did not reveal a clear context effect.

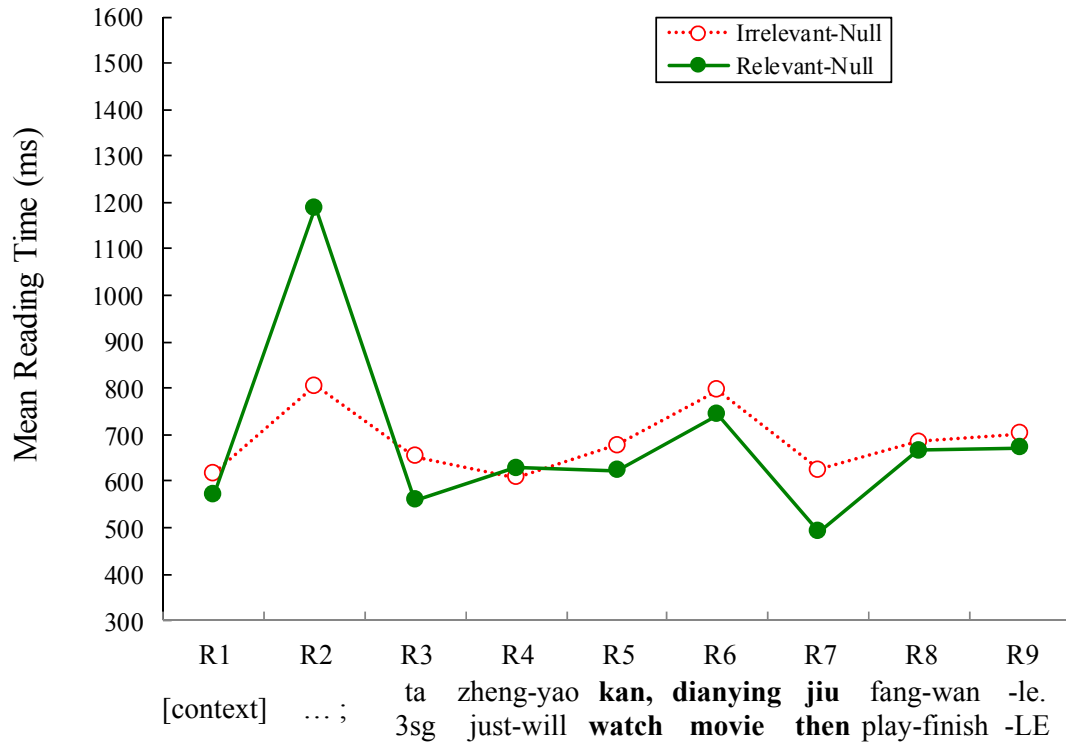


Figure 6.23. Japanese Intermediate L2ers' Mean RT (msec) Profile for the Null-Object Conditions of Experiment 6.2.

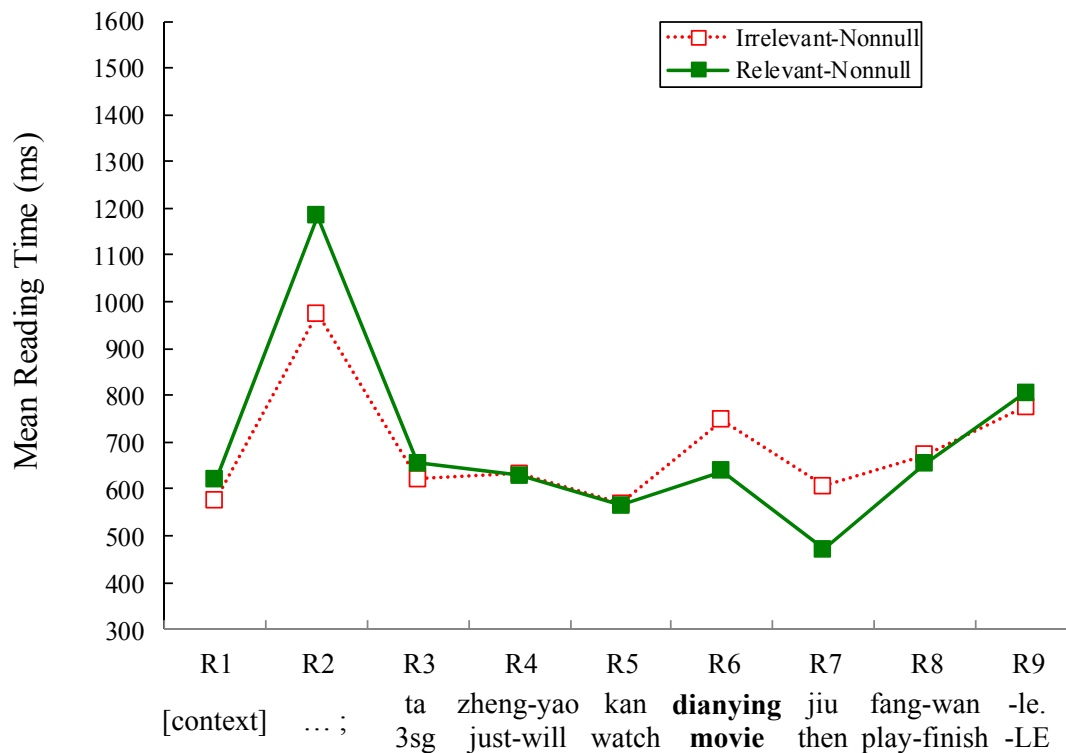


Figure 6.24. Japanese Intermediate L2ers' Mean RT (msec) Profile for the Nonnull-Object Conditions of Experiment 6.2.

The contextual influence on the 5 groups' null-object processing is summarized in Table 6.8. Both the Natives and the Advanced L2ers showed immediate and lasting effects of discourse context on their processing of null objects in Chinese. The English Intermediate L2ers also showed some delayed context effects.

Table 6.8. *Summary of the Main Online RT Results of Experiment 6.2*

Group	Influence of discourse context on null-object processing		
	Region 5	Region 6	Region 7
Natives	Yes	Yes	Yes
English Advanced L2ers	Yes	Yes	Yes
English Intermediate L2ers	No	Some	No
Japanese Advanced L2ers	Yes	Some	No
Japanese Intermediate L2ers	No	No	No (?)

6.5.3 Discussion

The predictions of the task-demand account of discourse-context effects on L1 and L2 null-object processing were borne out. First, both the Japanese Advanced L2ers and the English Advanced L2ers of Experiment 6.2 showed reliable discourse-context effects at the critical region, just like the Natives. Second, the context effects of the Native group and the 2 Advanced L2 groups in this experiment were generally more robust than those in Experiment 6.1 in that the effects were more reliable at the critical region and extended to the post-critical region(s). Third, Experiment 6.2 also observed a marginal context effect for English Intermediate L2ers at a post-critical region (i.e., Region 6). The only group that did not show a clear context effect in this experiment is the Japanese Intermediate group, although their mean RTs for the irrelevant-null sentences were also longer than for relevant-null sentences at Regions 5–7 (see Figure 6.23). Considering the small sample size of this group ($n = 8$), in contrast with other groups (Natives: $n = 24$; English Advanced L2ers: $n = 16$; Japanese Advanced L2ers: $n = 22$; English Intermediate L2ers: $n = 14$), one should not be too surprised at this group's failure to show a clear context effect. Besides, this group read the experimental items much faster ($M_{\text{speed}} = 686.6$) than the English Intermediate group ($M_{\text{speed}} = 961.2$) did, which might have compromised this group's online performance to some extent. Overall, all groups in Experiment 6.2, except the Japanese Intermediate group, showed

stronger context effects than the corresponding groups in Experiment 6.1 did, indicating that AJSPR can indeed induce fully specified processing (Wen & Schwartz, 2012).

To sum up, context manipulations affected natives' processing of null-object sentences at the critical region, suggesting that natives make use of discourse-context information at the very first opportunity. As for the L2ers, their results did not reveal an enhanced context effect as the SSH predicts in that Intermediate L2ers did not show context effects at the critical region. These results again indicate an L2 proficiency effect, as in Experiment 6.1.

We now address the more important unresolved issues pertaining to the results of Experiment 6.2.

What is a little surprising about Experiment 6.2 is that it did not find a clear L1 effect for the Japanese Intermediate group. One likely reason, as touched on earlier, is that the Japanese L2ers generally read the experimental sentences much faster than the English L2ers did, and this might have somewhat compromised their online performance. As revealed by the mean RTs, the Japanese Advanced group ($M_{\text{speed}} = 610.3$) was 349.1 milliseconds faster on average than the English Advanced group ($M_{\text{speed}} = 959.3$), and the Japanese Intermediate group ($M_{\text{speed}} = 686.6$) was 274.6 milliseconds faster on average than the English Intermediate group ($M_{\text{speed}} = 961.2$). Although the Japanese participants' fast reading speeds might indicate that they were less committed to fully specified, incremental online processing, this explanation is not satisfactory once we consider the L1 effect found in Experiment 6.1 for the Japanese Advanced L2ers, who also read much faster than their English counterparts but nevertheless displayed a native-like RT pattern. The most likely cause of no clear L1 effect for the Japanese Intermediate group is its small sample size, which prevented the data from yielding statistical significance.

Let us now consider why participants' acceptance of the experimental sentences was rather low. For the relevant-context conditions, one possible source of the low acceptance rates may have to do with the repetition of the topic NP in the target sentence. Generally, when an entity referred to by a lexical NP in (closely) preceding discourse is referred to again, a reduced referring expression, such as an overt or null pronoun, is used; in such cases, if a repeated NP is used instead, this requires additional mental processes and thus may introduce processing difficulties, dubbed the *Repeated Name Penalty* (RNP)

(Gordon, Grosz, & Gilliom, 1993; Gordon & Hendrick, 1998). Specifically, an RNP effect is evidenced by longer processing time for a repeated NP than for a reduced referring expression such as a null/overt pronoun. So, if this were the real cause of the low acceptance rates in the two relevant-context conditions, then there should be an RNP effect at Region 6, where the topic NP was repeated. The RT data of this experiment, however, did not show such a penalty: None of the groups spent longer time at Region 6 for the relevant-context conditions than for the corresponding irrelevant-context conditions. In fact, the RT patterns were the opposite (see Figures 6.15–6.24).

For the nonnull-object conditions, the low acceptance rates may be due to the garden path (GP) in these sentences. As discussed above (in Section 6.4.1.2), the NP following the target-sentence verb in these conditions involved a local structural ambiguity between an object and a subject, which might have affected acceptance of the sentences in those conditions. The RT data did not seem to give clear clues to this cause, however: For all participant groups, Region 8 of the nonnull-object conditions (i.e., the disambiguating region) did not show a significantly longer processing time than Region 8 of the null-object conditions ($F_s \leq 2.65$, $ps > .12$). In other words, the object-subject ambiguity did not produce a significant GP effect. Since Region 8 immediately precedes the final region, it is quite possible that the GP effect was delayed to the last region, in which case it would be difficult to detect. A possible (undetectable) GP effect notwithstanding, the GP sentences nevertheless did seem troublesome for the L2 participants, as their offline comprehension accuracy data show that they had more difficulty understanding the target sentences in the nonnull-object conditions than in the null-object conditions. This was also the case in Experiment 6.1 (see Section 6.4.2.1).

One very likely source of the participants' overall low acceptance is their trouble with (the pragmatics of) the propositional content of the experimental sentences. This was revealed by an informal interview conducted later, in which 2 Chinese native speakers (Interviewee A and Interviewee B) were asked to do the online task while the researcher watched how they did it.

When Interviewee A judged sentences like (12) as unacceptable, the researcher immediately stopped her and asked why she had made that judgment.

- (12) Xiangzi hen-zang; ta gang-yao ca, xiangzi jiu hua-po le.
box very-dirty 3sg just-will wipe, box then scratch LE
'The box is very dirty; just as s/he was about to wipe [it], the box was scratched.'

For (12) she said, “这个句子让人感觉挺奇怪挺意外的。怎么可能她还没有擦箱子就划破了呢？这样说不自然，所以我觉得这个句子不好。” ('The sentence sounds strange and unexpected. How can it be possible for the box to be scratched before she wiped it? It's not natural to say that, so I think it's not a good sentence.')

Interviewee B judged (13) as unacceptable.

- (13) Niunai bu-tai-re; ta zheng-yao he, niunai jiu liu-chulai le.
milk not-too-hot 3sg just-will drink milk then leak-out LE
'The milk was not too hot; just as s/he was about to drink [it], the milk leaked out.'

When asked why she made that judgment, she said, “我觉得这句话不太自然。牛奶是从哪儿流出来的？那跟牛奶不太热有关系吗？” ('I feel that this sentence is not quite natural to me. Where did the milk leak out from? Does it have anything to do with it being not too hot?')

From the answers provided by the 2 interviewees, it became clear that they gave a lot of consideration to the propositional content and pragmatic appropriateness of the sentences when judging acceptability, and they were obviously troubled by what they perceived as the unexpectedness in the meaning of the sentences. Since the target sentences were all created to fit into the schema “Just as someone was about to do something, an unexpected thing happened to it,” it is very likely, then, that (the pragmatics of the propositional content constituted the major cause of the overall low acceptance rates. This idea is also supported by the acceptability-judgment data from the two pairs of sentences below: (14) received the lowest mean acceptance (32%), while (15) received the highest mean acceptance (61%).

- (14) a. Tian jiuyao-hei-le; tamen zheng-yao dasao(.) fangjian jiu zhengqi le.
 sky will-dark-LE 3pl just-will clean room then tidy LE
 ‘It was getting dark; just as they were about to clean [it], the room got tidy.’
- b. Fangjian hen-zang-le; tamen zheng-yao dasao(.) fangjian jiu zhengqi le.
 room very-dirty-LE 3pl just-will clean room then tidy LE
 ‘The room got very dirty; just as they were about to clean [it], the room got tidy.’
- (15) a. Kuaidao wanshang-le; ta hai-mei canguan(.) gongchang jiu guanmen le.
 almost evening-LE 3sg still-not visit factory then close LE
 ‘It was almost evening; although s/he hadn’t visited [it], the factory was closed.’
- b. Gongchang hen-tebie; ta hai-mei canguan(.) gongchang jiu guanmen le.
 factory very-unusual 3sg still-not visit factory then close LE
 ‘The factory was very unusual; although s/he hadn’t visited [it], the factory was closed.’

The meaning of (14) is quite curious, a little like magic, because it seems to suggest that the room was able to clean itself. In contrast, the meaning of (15) is more natural: that because the factory was unusual, it was closed regardless of whether someone would visit it or not.

In short, the cause discussed above might have led to low acceptance rates of the experimental sentences. Although the AJSPR task was intended to be a form-focused task, it in effect became a coherence- and appropriateness-based acceptability-judgment task. In doing this task, the participants seemed to have paid more attention to the coherence of propositions and the pragmatic appropriateness than we had anticipated.

Still, the crucial and most meaningful take-home message from the online acceptance results of Experiment 6.2 is this: They showed reliable contextual effects, which suggest that discourse context influences both natives’ and L2ers’ parsing of null objects in Chinese.

6.6 General Discussion

Both experiments investigated the degree to which natives and L2ers are sensitive to discourse-context information in processing null objects in Chinese. In Experiment 6.1 (CFSPR), only the Natives and the Japanese Advanced L2ers showed sensitivity at the critical region to manipulations of discourse context. The English Advanced L2ers showed sensitivity only at the post-critical regions, while the 2 Intermediate L2 groups showed no sensitivity. In Experiment 6.2 (AJSPR), context effects were found for the Native and Advanced L2 groups at both the critical region and the post-critical region(s). The English Intermediate L2ers also showed some context effect at a post-critical region. The results of the two experiments confirm that discourse context exerts an immediate effect on L1 parsing and that L2 processing does not necessarily rely more on contextual information than L1 processing does. Contrary to what the SSH (Clahsen & Felser, 2006a, 2006b, 2006c) claims, processing discourse-context information is not always easy for L2ers. Rather, it is subject to influence from both L2 proficiency and L1 transfer.

According to the SSH, adult L2ers have (virtually) insurmountable difficulties in using morphosyntactic information in online sentence processing, as a result of which L2ers have to compensate for it by relying on nonsyntactic information, including lexical-semantic and discourse-level information. According to Pan and Felser (2011, p. 232), “the hypothesis that L2 processing is generally more meaning-based than L1 processing raises the question of whether L2ers’ grammatical processing abilities can ever become native-like.” Clahsen and Felser (2006c, p. 568) had in fact earlier suggested that “a high degree of proficiency in the L2 does not necessarily lead to native-like processing.” While we would of course agree with the latter statement, the SSH nevertheless seems to predict that increases in L2ers’ proficiency are not likely to bring about any change to their meaning-based *processing patterns*. The L2 proficiency effects in our study, as well as in other studies (e.g., Hopp, 2009), do not support this (apparent) SSH prediction. The L2ers in the current study showed clear proficiency effects in both experiments. The Advanced L2ers were better than the Intermediate L2ers in integrating discourse information in processing null objects, demonstrating that a higher degree of proficiency can lead L2ers to (come to) display native-like processing patterns.

The present study examined L1 transfer effects by recruiting L2 participants from two L1 backgrounds: English and Japanese. English is a nonnull-object language while Japanese is a null-object language. The results of Experiment 6.1 revealed that integrating discourse-context information and syntactic knowledge of null objects appears to be easier for L1-Japanese L2ers than for L1-English L2ers. The Advanced Japanese L2 group showed context effects at the critical region while the Advanced English L2 group showed it only later. These results challenge the SSH, which, to repeat, maintains that meaning-based strategies drive L2 processing, and even when the L1 and the target language instantiate the same structural properties, L2ers do not use that structural information. By contrast, these results support L2 theories that predict L1 transfer (of grammar/processing) in L2 processing such as the Fundamental Identity Hypothesis (Hopp, 2007) and the Full Transfer/Full Access/Full Parse model (Dekydtspotter et al., 2006; Schwartz & Sprouse, 1996).

The results of the current study are in sharp contrast to those found in Pan and Felser (2011) and Pan et al. (2015), in which L2ers, relative to natives, showed enhanced discourse-context effects in the online resolution of temporary structural ambiguities; the authors viewed these results as demonstrations of L2 over-reliance on discourse context as a cue to sentence interpretation. To test whether this SSH conclusion is tenable, we created a test design in Experiment 6.1 that *advantages* the SSH—an online task that is conducive to yielding context effects. First, the task focused on comprehension and thus was meaning-based. Second, the test language, Chinese, is context-dependent, at least in comparison to English. We assumed that processing a context-dependent language in a meaning-focused task should maximize the opportunity of showing a context effect for L2ers. However, the SSH prediction was not borne out. The English L2ers, even at an advanced level of proficiency, did not show early sensitivity (at the critical region) to discourse-context information in processing Chinese null objects (showing it at post-critical regions instead). In view of these results, it is difficult for the SSH to be maintained in its most extreme form.

As mentioned earlier, null objects are a type of reduced referring expressions, used for purposes of discourse coherence. Processing null objects requires the parser to coordinate and integrate information from both discourse and syntax, which is a

demanding task and can pose processing difficulty to L2ers (Hopp, 2009; Roberts et al., 2008). The results of the current study indeed indicate such a difficulty at the discourse-syntax interface for L1-English L2ers of Chinese which L2ers can (eventually) overcome, coming to pattern like native speakers (at least in Experiment 6.2).

Interestingly, comparing the more form-focused online task (AJSPR in Experiment 6.2) to the more meaning-focused task (CFSPR in Experiment 6.1) reveals it was the former that evinced greater L2 sensitivity to discourse-context information in null-object processing. This is unexpected under the SSH. If L2 processing lacks detailed morphosyntactic structure, a task that encourages structural processing should not help L2ers; rather, it should logically disadvantage them (if the task is to be used for comparing L1 and L2 structural processing), sharpening the contrast between L1 and L2 structural processing rather than diminishing it. Our AJSPR and CFSPR results taken together, then, pose a nontrivial problem for the SSH.

Overall, the results of the present study challenge the SSH but are consistent with proposals suggesting that L2 processing is epistemologically similar or identical to L1 processing (e.g., Dekydtspotter et al., 2006; Hopp, 2007, 2010).

6.7 Conclusion

Most of the studies on L2 sentence processing deal with L2ers' processing of narrow syntax, and often they present target sentences in isolation, without providing appropriate discourse context. A few recent studies (e.g., Hopp, 2009; Pan & Felser, 2011; Pan et al., 2015; Roberts et al., 2008) have begun to purposefully consider the role of discourse-context information in L2 sentence processing, of which the studies by Pan and Felser (2011) and Pan et al. (2015) directly tested whether manipulations of discourse context influence L2 processing. The present study has added to this line of inquiry by examining whether L2ers can use discourse-context information in processing Chinese sentences that contain null objects. In order to see whether—and if so, to what extent—the processing of null objects in Chinese is facilitated by an appropriate discourse context, this study employed two experiments: one focused on comprehension and one focused on judgments of acceptability. Both experiments were designed in such a way that the discourse context either did or did not satisfy the referential requirement of

the null objects. While the relevant-context condition provided a context that licenses null objects, the irrelevant-context condition did not. The results demonstrate that felicitous discourse contexts facilitate both L1 and L2 processing of Chinese null objects. This study also found (a) an L2 proficiency effect (Advanced L2ers pattern more like the natives than Intermediate L2ers), (b) an L1 effect (Advanced Japanese L2ers, whose L1 is very similar to Chinese in terms of how null objects operate, show earlier context effects than Advanced English L2ers do [at least in CFSPR]), and (c) a task effect (context effects are more robust in AJSPR than in CFSPR). The results indicate that L2ers are not necessarily over-reliant on contextual information or more sensitive to contextual influence than natives are, in contradiction to what the SSH claims.

The results of this study not only provide interesting insights into L2 research but also into L1 research on sentence processing. Both experiments show that the discourse-context information produces an immediate effect on L1 parsing. Processing difficulties associated with null objects can be eliminated by the presence of appropriate discourse contexts. The results of this study are consistent with those showing that discourse context can eliminate processing difficulties that stem from complex or noncanonical structures (e.g., restrictive relative clauses in Grodner et al., 2005; S-coordination in Hoeks et al., 2002; Finnish OVS order in Kaiser & Trueswell, 2004; Dutch object-initial relative clauses in Mak et al., 2008). The results are also in line with those suggesting immediate context effects in L1 Chinese sentence processing (e.g., in resolving the ambiguous $V + NP_1 + de + NP_2$ in S. Wang et al., 2008; in discourse integration in Zhang et al., 2002). Current theories of human sentence processing such as the depth-first modular approach (Frazier, 1987; Frazier & Fodor, 1978) and the constraint-based interactive approach (Bates & MacWhinney, 1989; MacDonald et al., 1994) are sharply divided over the role of (morpho)syntax and its relationships with other sources of knowledge such as (lexical) semantics and context. Depth-first models assign a primary role to syntax over other sources of knowledge (e.g., lexical semantics, discourse context, and world knowledge) and propose that syntax is applied at the very early stage of sentence processing while other sources of knowledge becomes available only at later stages. In contrast, constraint-based models propose that various sources of knowledge are present at the beginning stage of sentence processing and interact in

parallel to determine the outcome of comprehension. By showing that discourse context exerts immediate effects on both L1 and L2 processing of Chinese null objects, the present study thus informs the debates between depth-first and constraint-based theories and contributes to theories of human language processing in general. Since, as has been claimed, context effects on L1 parsing have to do with specific sentences types and specific types of contexts (Britt et al., 1992; Traxler & Tooley, 2007), they may be language-specific as well.

CHAPTER 7

CONCLUSION

In recent years, the field of second language (L2) studies has experienced an “online research” turn. A growing body of L2 research has turned to online psycholinguistic research methods such as self-paced reading (Just et al., 1982) or self-paced listening (Ferreira, Henderson, Anes, Weeks, & McFarlane, 1996), cross-model priming (Swinney, 1979; Swinney, Onifer, Prather, & Hirshkowitz, 1979), eye tracking (for a history, see Rayner, 1978, 1998), and event-related potentials (ERP; for reviews of using ERP in first language [L1] research, see Friederici, 2002; Kuperberg, 2007). Online methodologies provide reaction-time data that can reveal (or at least tell us something about) the underlying cognitive processes of language use. Online investigations of sentence-level L2 processing have demonstrated some observable differences between native speakers (“natives”) and nonnative speakers (“L2ers”), which led Clahsen and Felser (2006a, 2006b, 2006c) to propose the Shallow Structure Hypothesis (SSH). The central claim of the SSH is that L2 processing is fundamentally different from L1 processing, such that L2ers are restricted to shallow processing. The SSH assumes that normal L1 sentence processing in adults operates through two parallel routes: the *full parsing* route and the *shallow processing* route. The former is fed and constrained by the grammar and computes fully specified representations of the input, while the latter relies on nonsyntactic information and computes shallow representations of the input. The SSH proposes that L2 sentence processing, by contrast, mainly operates through the shallow processing route only; consequently, the representations that L2ers compute for the target language (TL) input during online sentence processing are shallow and contain little structural detail. The reason why L2ers cannot use the full parsing route, according to the SSH, is that the L2 grammar is deficient (relative to the TL grammar) and thus cannot supply the necessary morphosyntactic information for the full parsing route to function properly. Since the shallow processing route is not fed or constrained by morphosyntactic information, L2 processing has to rely on nonsyntactic information such as lexical-semantic information, contextual information, and real world knowledge.

The nature of L2 sentence processing is a critical issue in L2 acquisition research because it concerns the nature of L2 representation and L2 ultimate attainment. If L2ers cannot perform full (or even partial) parsing, then there is no way for L2 grammatical development (i.e., learning) to benefit from *analyzed input*—because the input cannot be structurally analyzed; consequently, the L2 “grammar” is necessarily deficient, and not only can it not converge on the TL grammar, it is not even of the same knowledge type.

To see whether L2ers are capable of full parsing (on the assumption that their Interlanguages do have grammars, albeit perhaps distinct from the TL grammar), we need to examine whether L2ers are able to access different sources of information in online sentence processing in the same way as natives do. The present dissertation research focuses on three major sources of information: morphosyntactic, lexical-semantic, and discourse-context. It comprises three studies in Chinese, testing whether L2ers are able to use these sources of information in processing Chinese sentences, in addition to a study in English, testing L2 processing strategies under different task conditions. In this way, we can find out whether L2ers are indeed restricted to shallow processing.

This dissertation employs two different online processing tasks: comprehension-focused self-paced reading (CFSPR) and acceptability-judgment self-paced reading (AJSPR). Comparing results from L2ers (matched for L1 and Chinese proficiency) across the two tasks allows us to inspect L2 processing from a new angle.

The results of this dissertation research demonstrate that adult L2ers, as their proficiency in the TL advances, are able to make use of these three sources of information in online TL sentence processing. This confirms that the L2 grammar is capable of supplying the information the parser needs in analyzing, integrating, and interpreting the TL input during online sentence processing. The differences between L1 processing and L2 processing found in previous research may thus not necessarily be of a qualitative nature; instead, they may indicate that L2 processing is subject to the influence of such factors as online task demands, L2 proficiency, and L1 transfer.

7.1 Summary of the Major Findings

This dissertation includes 4 major studies on L2 sentence processing: one study on English (Chapter 2) and three studies on Chinese (Chapters 4–6).

The main purpose of the study reported in Chapter 2 was methodological in nature: to establish AJSPR as a valid and useful tool for assessing L2 full parsing abilities. Since comprehension-focused online processing tasks such as CFSPR tend to be more difficult for adult L2ers than for natives, this methodological study questioned the “fairness” of using CFSPR-type tasks to assess L2 sentence processing. For adult L2ers whose access to the TL knowledge is not yet automatic (or established at all), sentence comprehension in the TL requires them to allocate their processing resources to both form and meaning at the same time (which is, of course, the typical way comprehension proceeds). But this means that comprehension-focused online reading tasks are more cognitively demanding for L2ers than for natives. For this reason, the present study maintains that studying L2 sentence processing in comparison to L1 sentence processing via CFSPR-type online tasks (for the purpose of determining the nature of L2 processing) is not fair practice for L2ers; if we are to compare natives and L2ers on equal terms, we should use online tasks such as AJSPR that can focus both natives and L2ers on processing the form and can thereby encourage full parsing. The study explored this issue by comparing L1 and L2 processing of English subject-verb number agreement in an AJSPR task in comparison to a CFSPR task in Wen (2007) that used exactly the same stimuli. The results indicate that AJSPR is more sensitive than CFSPR in detecting full parsing behavior and hence confirm that AJSPR is a useful tool for studying L2ers’ fully specified processing abilities. Since this methodological study demonstrated the usefulness of AJSPR, we applied this method in conjunction with CFSPR in all three studies of Chinese sentence processing.

The study reported in Chapter 4 deals with L2 processing of morphosyntactic information in Chinese. It looked at L1 and adult L2 processing of Chinese negation–aspect interactions; this was to tackle the question of whether adult L2ers of Chinese whose L1 (i.e., English) lacks negation–aspect interactions can ultimately acquire such morphosyntactic knowledge and access that knowledge in online sentence processing. Using experimental sentences manifesting nonadjacent structural relations, the study also evaluated whether adult L2ers can (come to) engage in fully-specified morphosyntactic processing. The results from both the CFSPR and AJSPR experiments reveal that L2ers whose L1 does not instantiate negation–aspect interactions are able to

acquire such knowledge and use it in online sentence processing as their proficiency rises. The study provided evidence for adult L2ers' fully specified morphosyntactic processing, because the Advanced L2 participants, like the Chinese Natives, showed successful processing of the negation–aspect interactions despite the fact that these interactions involve nonlocal structural relations. The Intermediate L2 participants did not pattern like the Natives in their processing profile; thus, the L2ers' data together indicate that there is an L2 proficiency effect in the L2 processing of negation–aspect interactions in Chinese.

Chapter 5 is concerned with L2 use of lexical-semantic information in Chinese sentence processing. The study reported in this chapter compared natives' and L2ers' sensitivity to temporary violations of the selectional restrictions of Chinese transitive verbs as well as natives' and L2ers' abilities to recover from initial misanalyses. The L2ers came from L1-English and L1-Japanese backgrounds. The results of the two experiments demonstrated Advanced L2ers' native-like online sensitivity to the selectional requirements of Chinese verbs in both CFSPR and AJSPR tasks, suggesting that both natives and Advanced L2ers make immediate use of lexical-semantic information in online sentence processing. Moreover, the study did not produce evidence for Roberts and Felser's (2011) predictions about L2ers' enhanced effects in meaning-based processing: L2ers did not show “more pronounced and/or prolonged” (p. 304) effects of local (im)plausibility than natives did. The findings therefore do not support the SSH claim that in comparison to natives, L2ers (necessarily) over-depend on lexical-semantic information in sentence processing at the expense of morphosyntax. Rather, L2 processing of lexical-semantic information is modulated by TL proficiency; more specifically, Advanced L2ers, but not Intermediate L2ers, demonstrate native-like use of lexical-semantic information in L2 sentence processing.

The third study on Chinese sentence processing, reported in Chapter 6, deals with L2 use of discourse-context information. It tested whether L2ers can make use of discourse-context information in processing Chinese sentences that contain null objects. The study manipulated the prior discourse context to make it either satisfy or dissatisfy the referential requirements of null objects in Chinese. The L2 participants of this study came from two L1 backgrounds: English (a nonnull-object language) and Japanese (a null-object language). This enabled us to test for L1 transfer effects. The results from the

CFSPR and AJSPR experiments demonstrated that Advanced L2ers, similar to natives, show discourse-context effects in processing null objects in Chinese and hence are able to use discourse-context information in online sentence processing. In contrast, this is not the case for Intermediate L2ers. The proficiency effect suggests that increases in L2ers' TL proficiency can lead to changes in their online TL processing routines. In addition, the study observed an L1 transfer effect: In Experiment 6.1 (CFSPR), the Advanced L1-Japanese L2ers, whose L1 (i.e., Japanese) is very similar to Chinese in terms of how null objects operate, showed earlier contextual effects in processing null objects than the Advanced L1-English L2ers did. The data suggest that structural properties in the L1 affect L2ers' processing of similar properties in the TL, and thus L2 processing is not necessarily predominantly meaning-based. Taking the results of the two experiments together, this study argues against the SSH in that adult L2ers do not necessarily overuse nonsyntactic information, and neither do they necessarily underuse syntactic information.

Briefly, the three studies on Chinese approach L2 sentence processing from three different angles: morphosyntactic, lexical-semantic, and discourse-context. The results of these studies reveal that L2ers are able to (come to) successfully access different sources of information in processing L2 sentences. Thus, L2 sentence processing need not be qualitatively different from L1 sentence processing, and the L2 parser is capable of full parsing.

7.2 Implications

L2 researchers generally agree that differences exist between L1 processing and adult L2 processing (at least for nonadvanced L2ers), but they disagree on the nature and sources of the differences. Advocates of the SSH view the differences as fundamental (i.e., of a qualitative nature). They attribute the differences to a deficient L2 grammar. Other researchers (e.g., Dekydtspotter et al., 2006; Hopp, 2007; Wen & Schwartz, 2012) believe that the differences are caused by L2ers' (developing Interlanguage grammar and) performance limitations and are therefore quantitative in nature.³⁶ The empirical evidence that proponents of the SSH have used to back up their arguments for L2 shallow

³⁶ "quantitative in nature" in so far as L2ers have knowledge of the grammatical phenomena/properties at issue—for without L2ers having that knowledge, there would be no point in testing L1 vs. L2 processing

processing chiefly depends on L2ers' nontarget-like online response patterns. However, as pointed out by several L2 researchers (e.g., Dekydtspotter et al., 2006; Miller, 2011, 2014, 2015; Omaki & Schulz, 2011), L2ers' nonnative-like online behavior is not necessarily indicative of shallow processing in the sense of the SSH. Factors that lead to L2ers' nontarget-like behavior may not be directly related to the *nature* of L2 representations (and hence of L2 processing). The present dissertation research demonstrates that other things can also cause divergent L2 online performance.

First, different task demands may affect the depth of L2 processing. L2ers who exhibit shallow processing in CFSPR tasks may show full parsing in AJSPR tasks (Chapter 2). Moreover, AJSPR tasks generally produce earlier and/or more robust effects than CFSPR tasks (Chapters 3–5). These results suggest that participants prioritize their cognitive resources in different ways under different task conditions. Task demands direct the allocation of cognitive resources and consequently influence (natives' and) L2ers' processing strategies. The task effects obtained in the current research suggest that L2 shallow processing is sometimes just a performance issue and does not necessarily attest to a deficient grammar.

Second, proficiency in the TL is a decisive factor in the development of L2 full parsing abilities. All three studies on Chinese sentence processing observed L2 proficiency effects (Chapters 4–6). As L2ers' proficiency in the TL improves, they can develop native-like (structural) processing abilities. L2 proficiency effects point to the quantitative rather than qualitative nature of the differences between L1 and L2 sentence processing.

Third, this research also observed an L1 transfer effect in L2 online processing (Chapter 6). Structural properties of the L1 (e.g., licensing and identifying null objects in Japanese) affect L2 processing of similar properties in the TL (e.g., licensing and identifying null objects in Chinese). The L1 effect likewise suggests that L2ers can make use of structural information online.

These things (i.e., task demands, L2 proficiency, and L1 transfer) work together in shaping L2 processing behavior, but none of them points to L2ers' processing as being necessarily qualitatively divergent from natives' processing.

In addition, it is worth noting that this dissertation research chose Chinese, an understudied language, as the target language of investigation and observed some interesting phenomena. First, morphosyntax plays an indispensable role in Chinese sentence processing (Chapter 4), although sentence interpretation in Chinese is said to be more context-dependent and more meaning-based than in English (e.g., C. N. Li & Thompson, 1981; Lu et al., 2000). Our data show that violations of morphosyntactic constraints in Chinese (i.e., negation–aspect interactions) lead to processing difficulties. Second, Chinese morphosyntax does not seem to be too difficult for L2ers to acquire and process, at least in terms of the morphosyntactic phenomena involved in this research, since Advanced L2ers do not seem to have much trouble with them (Chapter 4). Third, Chinese natives do not delay the use of lexical-semantic and discourse-context information in Chinese sentence processing (see Chapters 5 and 6).

Finally, the current research also has some implications for L2 teaching and learning. For example, form-focused practice may be helpful for successful L2 acquisition in that such practice can “force” (or at least encourage) L2ers to use the full parsing route (see Chapter 2), which in turn may facilitate the development of grammar as well as the formation of full parsing routines. Of course, the effectiveness of pedagogical intervention on L2 grammatical/processing development needs to be empirically established.

7.4 Concluding Remarks

This dissertation research aims to provide a systematic understanding of certain aspects of L2 sentence processing by revealing how shallow processing and full parsing operate in adult L2 processing. Findings from this research can illuminate L2 research and contribute to L2 theory-building. As Bley-Vroman (2009, p. 193) commented, “psycholinguistic research into online processing is required to explore further the interplay of rough-and-ready shallow processing and full grammar-driven processing.”

This research has examined the extent to which L2ers are able to make use of different sources of information in online sentence processing and what underlying causes (e.g., lack vs. possession of the TL knowledge, task demands, insufficient vs. sufficient proficiency in the TL, or absence vs. presence L1 transfer) there are for L2

shallow vs. deep processing. The specific sources of information that were focused on are morphosyntax, lexical semantics, and discourse context. They are the major sources of information the parser exploits in making parsing decisions. The results should be of interest and value to L2 researchers as well as to scholars of Chinese.

This research not only expands L2 research to locating specific causes for divergence between L1 processing and L2 processing but also provides a new methodology for investigating them—the use of AJSPR-type processing tasks to supplement CFSPR-type processing tasks. It is our hope that this work will drive L2 inquiry to novel areas of research and theorizing. In order to unravel the mystery of how adults learn a new language, researchers must start cracking how L2ers mentally process it in real time (Fodor, 1998; Sharwood Smith & Truscott, 2014; Truscott & Sharwood Smith, 2004). Without knowing how L2 sentences are processed, across L2 proficiency levels, it is impossible to truly understand how Interlanguage develops.

APPENDIX A

INSTRUCTIONS FOR THE EXPERIMENTS IN WEN (2007) AND CHAPTER 2

A. Instructions for Wen (2007)

This is a sentence reading task. Press the NEXT button to read instructions.

You will read sentences in a word-by-word fashion. Each sentence states a fact or point, which may be true or false. Your task is to decide whether each statement is true or false. The sentences were drawn from typical writing. Some of them may contain grammar errors, but the errors are minor and won't affect your decisions. Try to concentrate on the meaning of each sentence and ignore any grammar errors.

Try to read as quickly as possible until you reach the end of a sentence, and then decide whether the fact or the point stated is true or false. Press *TRUE* button if it's true. Press *FALSE* button if it's false. Press *NEXT* for the beginning word of each sentence, but press *CONTINUE* button to read the other words of the sentence. Now press *NEXT* to begin practice.

B. Instructions for the Study of Chapter 2

This is a sentence reading task. Press the NEXT button to read instructions.

You will read sentences in a word-by-word fashion. After you finish reading each sentence, decide whether it is acceptable in terms of its well-formedness. For example, you read: *Tom begin to study English two years ago*. This sentence is **not** acceptable. The acceptable version is: *Tom began to study English two years ago*. Note that some sentences are acceptable, although they may state a wrong fact or point; for example: *We all know that there are 12 weeks in a year*.

Try to read as quickly as possible until you reach the end of a sentence, and then decide whether the sentence is acceptable. Press *YES* button if it's acceptable. Press *NO* button if it's not acceptable. Press *NEXT* for the beginning word of each sentence, but press *CONTINUE* button to read the other words of the sentence. Now press *NEXT* to begin practice.

APPENDIX B
EXPERIMENTAL ITEMS, FILLERS, AND PRACTICE SENTENCES USED IN
THE EXPERIMENTS IN WEN (2007) AND CHAPTER 2

A. Experimental Items

For clarity, the following list gives all of the conditions of each experimental item:

- a = Agree-Vsg
- b = Disagree-Vsg
- c = Agree-Vpl
- d = Disagree-Vpl

True sentences

- 1) a. Applied scientific research usually caters to specific and practical needs.
b. * Very profitable businesses usually caters to specific and practical needs.
c. Very profitable businesses usually cater to specific and practical needs.
d. * Applied scientific research usually cater to specific and practical needs.
- 2) a. Beautiful mountain scenery often attracts people from big industrial cities.
b. * Beautiful tourist sites often attracts people from big industrial cities.
c. Beautiful tourist sites often attract people from big industrial cities.
d. * Beautiful mountain scenery often attract people from big industrial cities.
- 3) a. Common sense knowledge sometimes prevents ordinary people from thinking creatively.
b. * Common sense ideas sometimes prevents ordinary people from thinking creatively.
c. Common sense ideas sometimes prevent ordinary people from thinking creatively.
d. * Common sense knowledge sometimes prevent ordinary people from thinking creatively.
- 4) a. Excessively indulgent treatment often spoils young children who lack self-control.
b. * Excessively indulgent parents often spoils young children who lack self-control.
c. Excessively indulgent parents often spoil young children who lack self-control.
d. * Excessively indulgent treatment often spoil young children who lack self-control.
- 5) a. Good quality time usually helps couples to maintain close relations.
b. * Good quality talks usually helps couples to maintain close relations.
c. Good quality talks usually help couples to maintain close relations.
d. * Good quality time usually help couples to maintain close relations.
- 6) a. Highly sensational information generally spreads quickly in the news media.
b. * Highly sensational reports generally spreads quickly in the news media.
c. Highly sensational reports generally spread quickly in the news media.
d. * Highly sensational information generally spread quickly in the news media.
- 7) a. Truly original writing often comes from long and concentrated effort.
b. * Truly original thoughts often comes from long and concentrated effort.
c. Truly original thoughts often come from long and concentrated effort.
d. * Truly original writing often come from long and concentrated effort.

- 8) a. Very careful driving often reduces the probability of traffic accidents.
 b. * Clear traffic signs often reduces the probability of traffic accidents.
 c. Clear traffic signs often reduce the probability of traffic accidents.
 d. * Very careful driving often reduce the probability of traffic accidents.

False sentences

- 9) a. Cold winter air often raises the temperature of the earth.
 b. * Cold winter snowfalls often raises the temperature of the earth.
 c. Cold winter snowfalls often raise the temperature of the earth.
 d. * Cold winter air often raise the temperature of the earth.
- 10) a. Efficient time management never benefits people wanting to achieve success.
 b. * Efficient communication skills never benefits people wanting to achieve success.
 c. Efficient communication skills never benefit people wanting to achieve success.
 d. * Efficient time management never benefit people wanting to achieve success.
- 11) a. Extremely hard work always causes old people to feel relaxed.
 b. * Awfully hard trips always causes old people to feel relaxed.
 c. Awfully hard trips always cause old people to feel relaxed.
 d. * Extremely hard work always cause old people to feel relaxed.
- 12) a. Regular class attendance never improves the chances of successful learning.
 b. * Effective teaching methods never improves the chances of successful learning.
 c. Effective teaching methods never improve the chances of successful learning.
 d. * Regular class attendance never improve the chances of successful learning.
- 13) a. Successful language learning never happens to adult or child learners.
 b. * Really interesting things never happens to adult or child learners.
 c. Really interesting things never happen to adult or child learners.
 d. * Successful language learning never happen to adult or child learners.
- 14) a. Terribly bad behavior often makes a good impression upon people.
 b. * Terribly bad manners often makes a good impression upon people.
 c. Terribly bad manners often make a good impression upon people.
 d. * Terribly bad behavior often make a good impression upon people.
- 15) a. Truly practical advice only aids people who have university degrees.
 b. * Truly practical suggestions only aids people who have university degrees.
 c. Truly practical suggestions only aid people who have university degrees.
 d. * Truly practical advice only aid people who have university degrees.
- 16) a. Very hot water never hurts children if they touch one.
 b. * Very hot irons never hurts children if they touch one.
 c. Very hot irons never hurt children if they touch one.
 d. * Very hot water never hurt children if they touch one.

(Note that the last word *one* in [16a] and [16c] is ungrammatical. Fortunately, the errors occurred in the sentence-final position and did not affect the RT results. Hence, they were not excluded in the data analysis.)

B. Fillers

True, grammatical sentences

- 1) An active social life may be good for people's health.
- 2) The family is often an important influence on young adults.
- 3) Quite many schools now ask students to evaluate their teachers.
- 4) People can strengthen their immune systems by doing physical exercises.
- 5) Some famous athletes can earn millions of dollars every year.
- 6) It's hard for everyone to be satisfied with their lives.
- 7) Some young children spend a lot of time practicing sports.
- 8) Obesity rates among children have been increasing in rich countries.
- 9) It's safer for bicyclists to wear a helmet while bicycling.
- 10) The number of animal species is reducing almost every day.
- 11) Playing a game is often fun especially when you win.
- 12) Driving for long hours is likely to increase accident rates.
- 13) People often remember special gifts or presents they have received.
- 14) People have different ways of dealing with stress in life.
- 15) Some people get up early in the morning every day.
- 16) Extremely dirty places can cause people to have poor health.
- 17) Some people have the habit of eating rice for dinner.
- 18) History lessons can help people to make important decisions today.
- 19) The rapid pace of life today can cause psychological problems.
- 20) Society and people around us can define our life's attitudes.
- 21) Money has been in use for a very long time.
- 22) Most professions and academic fields require both imagination and knowledge.
- 23) The primary way to communicate ideas is through using language.
- 24) Banks are places where people can deposit or withdraw money.

True, ungrammatical sentences

- 25) Students should not learn only facts with his college education.
- 26) Conventional ways of thinking often limit us in make decisions.
- 27) One goal of technological advancement is to increasing people's efficiency.
- 28) Competition generally has both negative and positive effects about people.
- 29) People can obtain useful information by listen to radio programs.
- 30) Effective education is normally meet individual student's needs and interests.
- 31) College and university education should not free for all students.
- 32) Too many emphasis on role models harms our independent thinking.

False, grammatical sentences

- 1) Earthquakes are never thought of as a destructive natural phenomenon.
- 2) Fruit flies are tiny organisms much beloved by people everywhere.
- 3) It is impossible for hurricanes to cause any real damage.
- 4) Tremendously large plants often eat insects as their daily food.
- 5) Almost all the world's animals consume rotten meat every day.
- 6) Many plants in the world cannot survive in warm places.
- 7) Really important discoveries or creations always come unexpectedly by accident.
- 8) People living near rivers are often troubled by tidal waves.
- 9) Sleep is usually not quite necessary for most of us.
- 10) It causes you many health problems if you drink water.
- 11) Listening to light music can always make people feel unhappy.
- 12) The only way to solve vision problems is wearing sunglasses.
- 13) College teachers and students generally hate weekends and long vacations.

- 14) Older people who eat much meat will enjoy better health.
- 15) There is no real and useful information on the internet.
- 16) Shyness usually makes a person much smarter and more confident.
- 17) Eating less than normal can usually help people gain weight.
- 18) It usually makes people much sleepier if they drink coffee.
- 19) Success in life never comes to people who work hard.
- 20) Watching television shows featuring violent scenes always make people polite.
- 21) Modern people mostly spend their entire lives in one place.
- 22) Eating green vegetables every day always causes severe heart diseases.
- 23) University teachers today never require their students to attend classes.
- 24) Dancing is the best method for improving students' reading skills

False, ungrammatical sentences

- 25) Most of planets in the solar system have living beings.
- 26) All planets on the universe have fresh water and air.
- 27) Technologies used in schools always distracting students from real learning.
- 28) Human bodies created new blood cells only after each meal.
- 29) The earth's atmosphere contained nothing but some oxygen and water.
- 30) People lived in cities are never troubled by air pollutions.
- 31) Everyone agrees that swimming often does harm in people's health.
- 32) Space telescopes can only help us watching things on earth.

C. Practice Sentences

- 1) Reading books helps us to growing in knowledge.
- 2) Air pollution is usually good for the environment.
- 3) Human beings are the most intelligent beings in this world.
- 4) Living things cannot surviving without air or water.
- 5) Vegetables cannot provide any nutrition we need.
- 6) Most parts of the world have four seasons during year.

APPENDIX C
MEAN ACCEPTABILITY-JUDGMENT ACCURACY RATES FOR THE
EXPERIMENTAL SENTENCES AND FILLERS IN THE EXPERIMENT OF
CHAPTER 2

Group	Condition				All Conditions
	Agree-Vsg	Disagree-Vsg	Agree-Vpl	Disagree-Vpl	
	Experimental Sentences ($k = 16$)				
Natives ($n = 16$)	.79 (.28)	.80 (.21)	.89 (.19)	.88 (.18)	.84 (.22)
L2ers ($n = 16$)	.89 (.18)	.78 (.23)	.85 (.19)	.87 (.16)	.85 (.19)
	Fillers ($k = 64$)				
Natives ($n = 16$)	N/A	N/A	N/A	N/A	.84(.13)
L2ers ($n = 16$)	N/A	N/A	N/A	N/A	.80(.11)

Note. Standard deviations are in parentheses.

APPENDIX D
TRIMMED RAW RT MEANS (IN MILLISECONDS) FOR THE
EXPERIMENTAL SENTENCES USED IN THE EXPERIMENT OF CHAPTER 2

Condition	Region									
	1	2	3	4	5	6	7	8	9	10
Natives ($n = 16$)										
Agree-Vsg	586 (151)	531 (145)	501 (141)	491 (154)	525 (177)	494 (165)	475 (130)	467 (136)	491 (142)	635 (219)
Disagree-Vsg	603 (154)	510 (168)	544 (208)	505 (151)	607 (270)	550 (152)	398 (122)	413 (171)	419 (145)	493 (185)
Agree-Vpl	631 (173)	522 (167)	585 (191)	520 (138)	521 (146)	473 (146)	503 (148)	476 (129)	469 (110)	658 (223)
Disagree-Vpl	589 (179)	479 (135)	502 (196)	519 (182)	580 (213)	528 (144)	403 (139)	360 (134)	354 (132)	474 (190)
L2ers ($n = 16$)										
Agree-Vsg	665 (153)	563 (165)	608 (130)	588 (140)	663 (141)	534 (84)	578 (129)	510 (94)	553 (83)	658 (178)
Disagree-Vsg	650 (151)	601 (145)	659 (136)	637 (157)	730 (230)	558 (182)	430 (161)	396 (117)	414 (143)	494 (127)
Agree-Vpl	613 (152)	562 (134)	653 (160)	646 (167)	608 (165)	614 (116)	607 (146)	543 (128)	580 (128)	688 (156)
Disagree-Vpl	684 (171)	544 (127)	654 (170)	593 (151)	717 (264)	588 (157)	543 (205)	422 (117)	434 (111)	522 (133)

Note. Standard deviations are in parentheses.

APPENDIX F
THE CHINESE PROFICIENCY TEST

填空练习 Fill-in-the-blank exercise

指示说明：下面两篇短文略去了一些字。请你根据上下文的意思，在空格处填写略去的字，使意思完整。每个空格填一个字。不会写的字请用拼音代替。请在10分钟之内完成，每篇不超过5分钟。

Directions: The two texts below have left out some Chinese characters. Fill in each blank with just one character so that the texts make sense. If you do not know how to write a character, please use its *pinyin* form instead. You have a total of 10 minutes to complete the exercise, that is, 5 minutes for each passage.

第一篇 (5分钟内完成)

从历史上看，姓和名往往不是同时出现的。比____，汉语中先有姓，很____以后才出现名。英语中的情____恰好相____。起____，英语里只有名，没有姓。后来，因____人口数量不断增____，有限的几个人名被越来越____的人使用，重名现____就成了问____，给人们带____许多不____。于____，人们开____在名字后____附加上一些分辨身____的词语以便区____重名____。这____的做____逐渐系统____。人们使用家长的名字、家____住址、职业名____以及相貌特____等来做分辨词。这些分辨身份的词语就演____成了今天英语中的姓。

第二篇 (5分钟内完成)

如今，我们每天都特别忙，好像总有做不完的事。有时甚____连吃____和睡____的时____都没有。可是，忙碌似____并没有给我们多____充实和成就感。许多时____，经____一天紧____的劳碌，我们反而感到更____空虚和失____。这到____是什么原____造____的呢？现在，人们不用像从____那样为衣____担忧，生活水____确实提____了,但不____为什么，依____感觉生活缺____保障。这或____就是人们日复一____、无休无____劳累的原因吧。看来，这种困惑也并____仅仅是我们这一代人才有的。但愿未来这一切会有所好转。

APPENDIX G
SCORING METHOD AND ANSWER KEY FOR THE CHINESE PROFICIENCY
TEST

打分方法和答案 Scoring Method and Answer Key

两篇短文总共有50个空处。每个空只能填写一个字。每个空1分。满分为50分。答案准确的（即写对该汉字的）给1分。答案接近正确的，给0.75分，这包括答案的拼音、错拼字（即写错笔画但可辨认出是答案要求的字）以及同音别字。如果考生填了与标准答案不同但能说得上通的字，即“说得通的答案”，给0.5分，这也包括该答案的拼音、错拼字以及同音别字。另外，如果考生使用了拼音变体，要按照正确的拼音计算分数。其他情况均为0分。

There are 50 blanks altogether in the two texts. Each blank should be filled with just 1 character. The maximum total score is 50 points. Here is how to score each response: (a) **1** point for each **accurate response** (i.e., exactly the same as the correct answer); (b) **0.75** point for each **close-to-correct response** (i.e., failed to write the correct character but supplied the *pinyin*, misspelt character, or homophone); (c) **0.5** point for each **plausible response** (i.e., answer that is possible but deviates a little from the target answer, counting those written in *pinyin*, misspelt character, or homophone); (d) **0** point for all other cases. A variant form of *pinyin* is always treated as the correct use of the *pinyin* form.

The underlined bold-typed characters are target answers. Plausible answers are put in parentheses. (划线黑体字是标准答案。括号里的是说得通的答案。)

第一篇

从历史上看，姓和名往往不是同时出现的。比如，汉语中先有姓，很久/晚以后才出现名。英语中的情况(形)恰好相反(对)。起初(先、始、时)，英语里只有名，没有姓。后来，因为人口数量不断增加(长(多、大))，有限的几个人名被越来越多的人使用，重名现象(在、状)就成了问题，给人们带来许多不便。于是，人们开始在名字后面(边(再))附加上一些分辨身份的词语以便区别(分)重名者(的、人)。这样的做法逐渐系统化(了)。人们使用家长的名字、家庭(长、居、乡、族、人、的)住址、职业名称(字)以及相貌特征(点)等来做分辨词。这些分辨身份的词语就演变(化、绎)成了今天英语中的姓。

第二篇

如今，我们每天都特别忙，好像总有做不完的事，有时甚至连吃饭(喝)和睡觉(眠)的时间(候)都没有。可是，忙碌似乎并没有给我们多少充实和成就感。许多时候(间)，经过(历、常)一天紧张的劳碌，我们反而感到更加(多、深、有、为)空虚和失落(望、意、败)。这到底是什么原因造成的呢？现在，人们不用像从前那样为衣食(住)担忧，生活水平(准)确实提高(升)了，但不知(懂)为什么，依然(旧)感觉生活缺乏(少(点、欠))保障。这或许(者)就是人们日复一日、无休无止劳累的原因吧。看来，这种困惑也并非(不)仅仅是我们这一代人才有的。但愿未来这一切会有所好转。

APPENDIX H
THE SPECIAL TYPE OF FILLER USED IN THE ONLINE TASKS

A. In CFSPR

English translations are provided in the parentheses.

1. 总是不睡觉对身体会有害的。(If people don't sleep, it will do harm to their health.)
2. 不是所有的人都喜欢冷天气。(Not everyone likes cold weather.)
3. 天天不休息, 人们会累的。(People will get tired if they don't rest every day.)
4. 天太热时, 人们一般会出汗。(People will sweat if the weather is too hot.)
5. 学生不学习一般不会学得很好。(Students can't learn well if they don't study.)
6. 一个星期总共有七天。(There are altogether seven days in a week.)
7. 通常, 人每天都要吃饭。(Normally, people need to eat every day.)
8. 很少有人不知道一年有多少天。(Very few people don't know how many days there are in a year.)

B. In AJSPR

Glossing is provided because the sentences are ungrammatical. The relevant ungrammatical parts are shaded.

1. * 英语 可能 没是 很 容易 学好的。
English possibly not-be very easy learn-well-DE
2. * 他 把 送给 人 那个 房子。
he BA give person that house
3. * 他 喜欢 中国, 讲 经常 汉语。
he like China speak often Chinese
4. * 他 一遍 刚才 看 了 书。
he one-CL just-now read -LE book
5. * 很多 学生 都 汉语 会 讲。
many student all Chinese can speak
6. * 她 已经 不 有 东西 吃了。
she already not have thing eat-LE
7. * 他们 喜欢 很 这 的 生活。
they like very this DE life
8. * 他们 很 没 高兴 地 走了。
they very not happy DE go-LE

APPENDIX I

INSTRUCTIONS USED IN THE ONLINE TASKS

A. Instructions in CFSPR Tasks

English translations are provided after each sentence.

非常欢迎你参加此汉语阅读实验! (Welcome to this Chinese reading experiment!)

请按 NEXT (开始) 键继续看后面的说明。(Press the NEXT button to continue to read the following instructions.)

下面你要看一些汉语句子。(Below, you will read some Chinese sentences.)

是一个词一个词或者一个字一个字地看。(You will read in a word-by-word (character-by-character) fashion.)

看的速度由你自己控制, 由按键快慢决定。(You will control your reading speed by how fast you press the designated keys on the keyboard.)

请尽量用你的正常速度阅读每句话。(Please try to read each sentence with your normal speed.)

看完后请根据句子的意思回答问题。(After you finish reading a sentence, please answer the comprehension question according to the meaning of the sentence.)

比如, 你看了这句话: 我很喜欢看她写的书。(For example, you read this sentence: I really like reading the books she wrote.)

你要回答的问题是: 我不喜欢看她写的书, 对吗? (The comprehension question you need to answer is: I don't like the books she wrote, right?)

答案是: 不对。(The correct answer is: Not right.)

回答问题时, 请按 Yes (对) 键或者 No (不对) 键。(To answer the question, please press the Yes (Right) key or the No (Not Right) key.)

开始每句话, 总是按 NEXT 键。(Always press the NEXT key to start each sentence.)

连续读一句话则要使用 CONTINUE (继续) 键。(Press the CONTINUE key to continue reading each sentence.)

现在, 请按 NEXT (开始) 键开始练习。(Now, please press the NEXT (Start) key to begin your practice.)

.....

练习结束。请按 NEXT (下一句开始) 键开始实验。(End of practice. Please press the NEXT (Start the Next Sentence) key to begin the experiment.)

B. Instructions in AJSPR Tasks

English translations are provided in the parentheses.

非常欢迎你参加此汉语阅读实验！(Welcome to this Chinese reading experiment!)

请按 NEXT (开始) 键继续看后面的说明。(Press the NEXT button to continue to read the following instructions.)

下面你要看一些汉语句子。(Below, you will read some Chinese sentences.)

是一个词一个词或者一个字一个字地看。(You will read in a word-by-word (character-by-character) fashion.)

看的速度由你自己控制，由按键快慢决定。(You will control your reading speed by how fast you press the designated keys on the keyboard.)

请尽量用你的正常速度阅读每句话。(Please try to read each sentence with your normal speed.)

看完后请判断句子通顺不通顺。(After you finish reading a sentence, please judge whether the sentence is acceptable.)

也就是看它合不合汉语的习惯、规范。(That is, decide whether it conforms to the Chinese conventional ways of expressing it.)

比如后面这句就不通顺：**我很她喜欢**。(For example, the following sentence is not acceptable: I really her like.)

我们一般这样说：**我很喜欢她**。(We usually say it this way: I really like her.)

做判断时，请按 **Yes (通顺)** 键或者 **No (不通顺)** 键。(To make the judgment, please press the Yes (Acceptable) key or the No (Not Acceptable) key.)

开始每句话，总是按 NEXT 键。(Always press the NEXT key to start each sentence.)

连续读一句话则要使用 CONTINUE (继续) 键。(Press the CONTINUE key to continue reading each sentence.)

现在，请按 NEXT (开始) 键开始练习。(Now, please press the NEXT (Start) key to begin your practice.)

.....

练习结束。请按 NEXT (下一句开始) 键开始实验。(End of practice. Please press the NEXT (Start the Next Sentence) key to begin the experiment.)

APPENDIX J

PRACTICE ITEMS FOR THE SELF-PACED READING TASKS

A. Practice Items in CFSPR Tasks (Comprehension Questions in Parentheses)

English translations are provided. The relevant ungrammatical parts are shaded.

1. 小张刚要去跑步，他弟弟就来了。(谁来了? A. 小张 B. 他弟弟)
Just as Xiao Zhang was about to go jogging, his younger brother came. (Who came? A. Xiao Zhang B. His younger brother)
2. 我们的爱好根本就不一样。(我们俩的爱好一样，对吗?)
Our hobbies are totally different. (Our hobbies are the same, right?)
3. 他们坐了同一辆车去学校。(他们坐了同一辆车，对吗?)
They took the same bus to school. (They took the same bus, right?)
4. 小青是我中学时候的朋友。(小青是我以前的朋友，对吗?)
Xiao Qing is my high school friend. (Xiao Qing is my former friend, right?)
5. 小张特别喜欢学习英语，这不假。(小张喜欢学习英语，对吗?)
Xiao Zhang especially likes learning English, and this is true. (Xiao Zhang likes learning English, right?)
6. 老师和家长关心他的学习。(他们都不关心他的学习，对吗?)
His teachers and parents are concerned about his studies. (They are not concerned about his studies, right?)
7. 上海的天气和北京的不一样。(上海和北京的天气一样，对吗?)
The weather in Shanghai is different from that in Beijing. (Shanghai and Beijing have similar weather, right?)
8. *小李既喜欢跳舞既喜欢唱歌。(小李不喜欢跳舞，对吗?)
*Xiao Li not only likes dancing not only likes singing. (Xiao Li doesn't like dancing, right?)

B. Practice Items in AJSPR Tasks

English translations are provided only for the grammatical sentences. The ungrammatical sentences are glossed (since they cannot be translated directly). The relevant ungrammatical parts are shaded. Note that some of the sentences, especially the grammatical sentences, are similar or identical to those appearing in the CFSPR tasks. Most of the ungrammatical sentences are modified from the corresponding grammatical sentences in the CFSPR tasks.

1. 小张刚要去跑步，他妈就来了。
Just as Xiao Zhang was about to go jogging, his mother came.

2. * 我们的 爱好 根本 就 有 一样。
our hobbies fundamentally just have the-same
3. 他们坐了同一辆车去学校。
They took the same bus to school.
4. 小青是我中学时候的朋友。
Xiao Qing is my high school friend.
5. 小张特别喜欢学习英语，这不假。
Xiao Zhang especially likes learning English, and this is true.
6. * 老师 和 家长 被 关心 他的 学习。
Teacher and parent BEI care-about his study
7. * 上海的 天气 和 北京的 不 样。
Shanghai's weather and Beijing-DE not style
8. * 小李 既 喜欢 跳舞 既 喜欢 唱歌。
Xiao Li not-only like dancing not-only like singing

APPENDIX K
EXPERIMENTAL ITEMS FOR EXPERIMENTS 4.1 AND 4.2

These are the experimental items used in this study, with English glossing below each.

1. 他	刚/没/不	看	过/了	这	本	书。
he	just/mei/bu	look	-GUO/-LE	this	CL	book
2. 他们	刚/没/不	去	过/了	那	个	地方。
they	just/mei/bu	go	-GUO/-LE	that	CL	place
3. 她们	刚/没/不	见	过/了	那	个	老师。
they	just/mei/bu	see	-GUO/-LE	that	CL	teacher
4. 我们	刚/没/不	问	过/了	那	个	问题。
we	just/mei/bu	go	-GUO/-LE	that	CL	place
5. 我	刚/没/不	听说	过/了	这	件	事。
I	just/mei/bu	hear	-GUO/-LE	this	CL	matter
6. 她	刚/没/不	学	过/了	那	门	课。
she	just/mei/bu	study	-GUO/-LE	that	CL	course
7. 她们	刚/没/不	买	过/了	这	件	衣服。
they	just/mei/bu	buy	-GUO/-LE	this	CL	clothes
8. 我们	刚/没/不	坐	过/了	那里	的	地铁。
we	just/mei/bu	sit	-GUO/-LE	there	DE	subway
9. 他们	刚/没/不	吃	过/了	那样	的	米饭。
they	just/mei/bu	eat	-GUO/-LE	that-kind	CL	rice
10. 他	刚/没/不	做	过/了	那	个	菜。
he	just/mei/bu	do	-GUO/-LE	that	CL	dish
11. 我们	刚/没/不	叫	过/了	那	个	车。
we	just/mei/bu	call	-GUO/-LE	that	CL	car
12. 她	刚/没/不	唱	过/了	那	首	歌。
she	just/mei/bu	sing	-GUO/-LE	that	CL	song
13. 我们	刚/没/不	练习	过/了	那	句	话。
we	just/mei/bu	practice	-GUO/-LE	that	CL	sentence
14. 她	刚/没/不	洗	过/了	那	条	裤子。
she	just/mei/bu	wash	-GUO/-LE	that	CL	pants
15. 他	刚/没/不	介绍	过/了	这	位	先生。
he	just/mei/bu	introduce	-GUO/-LE	this	CL	gentleman

16. 他们 they	刚/没/不 just/mei/bu	玩 play	过/了 -GUO/-LE	这 this	个 CL	游戏。 game
17. 她们 they	刚/没/不 just/mei/bu	写 write	过/了 -GUO/-LE	那 that	个 CL	题目。 topic
18. 我 I	刚/没/不 just/mei/bu	读 read	过/了 -GUO/-LE	那 that	篇 CL	文章。 article
19. 我 I	刚/没/不 just/mei/bu	解释 explain	过/了 -GUO/-LE	那 that	个 CL	单词。 word
20. 我们 we	刚/没/不 just/mei/bu	庆祝 celebrate	过/了 -GUO/-LE	他 he	的 DE	生日。 birthday
21. 他们 they	刚/没/不 just/mei/bu	表达 express	过/了 -GUO/-LE	这 this	个 CL	想法。 idea
22. 他 he	刚/没/不 just/mei/bu	打 beat	过/了 -GUO/-LE	那 this	个 CL	同学。 classmate
23. 她 she	刚/没/不 just/mei/bu	骂 scold	过/了 -GUO/-LE	这 this	个 CL	人。 person
24. 她们 they	刚/没/不 just/mei/bu	用 use	过/了 -GUO/-LE	这 this	台 CL	电脑。 computer

APPENDIX L
GRAMMATICAL *BU* FILLERS FOR EXPERIMENTS 4.1 AND 4.2

Below are the 8 filler items with grammatical uses of the negator *bu*. English translations are provided in parentheses.

1. 他不喜欢这里的天气。(He doesn't like the weather here.)
2. 她不想每天去跑步。(She doesn't want to go jogging every day.)
3. 他不去那里学习外语。(He won't go there to study foreign languages.)
4. 他们不是这里的老师。(They are not teachers here.)
5. 我不认识那个学生。(I'm not acquainted with that student.)
6. 我们不知道学校在哪里。(I don't know where the school is located.)
7. 他不觉得汉语很难。(I don't think Chinese is hard to learn.)
8. 她不关心孩子的学习。(She doesn't care about her children's study.)

APPENDIX M
MEAN COMPREHENSION ACCURACY RATES FOR THE EXPERIMENTAL
SENTENCES IN EXPERIMENT 4.1

Group	Condition						
	<i>gang-guo</i>	<i>mei-guo</i>	* <i>bu-guo</i>	<i>gang-le</i>	* <i>mei-le</i>	* <i>bu-le</i>	All
Natives (<i>n</i> = 30)	.97 (.09)	.99 (.05)	.98 (.08)	.98 (.08)	.96 (.09)	.90 (.19)	.96 (.11)
Advanced L2ers (<i>n</i> = 30)	.99 (.05)	.98 (.08)	.93 (.11)	1.00 (.00)	.95 (.12)	.92 (.17)	.96 (.11)
Intermediate L2ers (<i>n</i> = 24)	.91 (.16)	.91 (.14)	.92 (.14)	.95 (.10)	.89 (.15)	.86 (.21)	.91 (.15)

Note. Standard deviations are in parentheses.

APPENDIX N
TRIMMED RAW RT MEANS (IN MILLISECONDS) FOR THE
EXPERIMENTAL SENTENCES IN EXPERIMENT 4.1

Condition	Region						
	1	2	3	4	5	6	7
Natives (<i>n</i> = 30)							
<i>gang-guo</i>	489 (126)	336 (118)	344 (131)	327 (125)	326 (121)	314 (94)	392 (150)
<i>mei-guo</i>	519 (150)	343 (130)	341 (125)	346 (98)	314 (94)	332 (100)	386 (127)
* <i>bu-guo</i>	503 (145)	324 (98)	345 (132)	358 (112)	417 (155)	374 (117)	464 (180)
<i>gang-le</i>	511 (141)	333 (99)	344 (105)	329 (96)	326 (96)	321 (104)	385 (161)
* <i>mei-le</i>	491 (146)	338 (120)	343 (133)	346 (120)	434 (147)	367 (107)	458 (168)
* <i>bu-le</i>	515 (125)	351 (125)	349 (120)	361 (121)	378 (123)	349 (87)	467 (168)
Advanced L2ers (<i>n</i> = 30)							
<i>gang-guo</i>	669 (176)	528 (90)	662 (212)	560 (132)	522 (119)	511 (131)	762 (201)
<i>mei-guo</i>	755 (216)	562 (152)	612 (167)	533 (111)	498 (108)	490 (114)	723 (200)
* <i>bu-guo</i>	713 (209)	499 (108)	628 (188)	569 (126)	600 (228)	554 (132)	738 (253)
<i>gang-le</i>	704 (211)	518 (112)	637 (196)	497 (107)	503 (126)	463 (112)	728 (221)
* <i>mei-le</i>	683 (173)	544 (122)	637 (134)	526 (107)	589 (190)	514 (139)	731 (180)
* <i>bu-le</i>	724 (222)	495 (112)	594 (166)	519 (122)	574 (153)	521 (106)	746 (180)
Intermediate L2ers (<i>n</i> = 24)							
<i>gang-guo</i>	777 (300)	638 (180)	769 (219)	661 (219)	597 (145)	578 (186)	916 (259)
<i>mei-guo</i>	814 (285)	639 (202)	718 (232)	646 (166)	642 (220)	611 (170)	959 (232)
* <i>bu-guo</i>	734 (194)	548 (149)	695 (169)	674 (265)	618 (181)	564 (162)	907 (245)
<i>gang-le</i>	754 (237)	625 (147)	767 (261)	588 (153)	594 (141)	536 (126)	868 (217)
* <i>mei-le</i>	750 (215)	593 (145)	735 (176)	549 (145)	595 (152)	568 (181)	846 (257)
* <i>bu-le</i>	770 (285)	513 (118)	731 (222)	595 (185)	598 (157)	604 (239)	861 (228)

Note. Standard deviations are in parentheses.

APPENDIX O
MEAN ACCEPTABILITY-JUDGMENT ACCURACY RATES FOR THE
EXPERIMENTAL SENTENCES IN EXPERIMENT 4.2

Group	Condition						
	<i>gang-guo</i>	<i>mei-guo</i>	* <i>bu-guo</i>	<i>gang-le</i>	* <i>mei-le</i>	* <i>bu-le</i>	All
Natives (<i>n</i> = 30)	.93 (.16)	.95 (.1)	.96 (.12)	.88 (.18)	.90 (.21)	.94 (.14)	.93 (.16)
Advanced L2ers (<i>n</i> = 28)	.73 (.32)	.89 (.14)	.80 (.28)	.86 (.17)	.54 (.38)	.74 (.31)	.76 (.30)
Intermediate L2ers (<i>n</i> = 25)	.73 (.23)	.88 (.15)	.59 (.35)	.81 (.18)	.45 (.32)	.45 (.31)	.65 (.31)

Note. Standard deviations are in parentheses.

APPENDIX P
TRIMMED RAW RT MEANS (IN MILLISECONDS) FOR THE
EXPERIMENTAL SENTENCES IN EXPERIMENT 4.2

Condition	Region						
	1	2	3	4	5	6	7
Natives ($n = 30$)							
<i>gang-guo</i>	577 (138)	434 (107)	459 (134)	430 (118)	397 (104)	422 (101)	559 (171)
<i>mei-guo</i>	552 (143)	423 (101)	445 (128)	432 (84)	409 (81)	402 (66)	568 (194)
* <i>bu-guo</i>	607 (169)	437 (150)	457 (150)	452 (112)	474 (117)	400 (90)	518 (161)
<i>gang-le</i>	559 (174)	435 (119)	468 (148)	431 (131)	421 (128)	416 (124)	561 (232)
* <i>mei-le</i>	576 (160)	441 (129)	451 (154)	489 (166)	486 (92)	399 (98)	515 (149)
* <i>bu-le</i>	577 (149)	425 (109)	483 (156)	483 (139)	477 (120)	415 (99)	528 (153)
Advanced L2ers ($n = 28$)							
<i>gang-guo</i>	721 (204)	620 (159)	760 (196)	637 (144)	643 (193)	594 (216)	890 (307)
<i>mei-guo</i>	719 (224)	605 (153)	757 (204)	639 (168)	589 (178)	541 (143)	878 (272)
* <i>bu-guo</i>	721 (250)	546 (155)	736 (225)	707 (229)	649 (208)	478 (204)	677 (325)
<i>gang-le</i>	685 (191)	636 (163)	742 (216)	567 (191)	638 (234)	564 (136)	864 (328)
* <i>mei-le</i>	700 (196)	577 (143)	766 (224)	648 (231)	675 (216)	548 (158)	848 (402)
* <i>bu-le</i>	709 (194)	539 (143)	799 (222)	650 (190)	728 (206)	520 (144)	806 (310)
Intermediate L2ers ($n = 25$)							
<i>gang-guo</i>	686 (194)	622 (186)	806 (193)	790 (349)	711 (216)	592 (145)	954 (391)
<i>mei-guo</i>	678 (200)	626 (151)	814 (247)	714 (268)	695 (190)	574 (152)	1006 (345)
* <i>bu-guo</i>	721 (234)	574 (148)	774 (217)	731 (259)	709 (221)	567 (160)	929 (318)
<i>gang-le</i>	710 (276)	680 (212)	839 (222)	630 (165)	680 (197)	635 (210)	1000 (282)
* <i>mei-le</i>	703 (229)	657 (146)	809 (216)	715 (233)	705 (215)	636 (233)	1024 (414)
* <i>bu-le</i>	678 (168)	525 (88)	810 (241)	673 (199)	707 (194)	627 (219)	1011 (366)

Note. Standard deviations are in parentheses.

APPENDIX Q
ANALYSES OF VARIANCE FOR THE EFFECTS OF CONDITION BEFORE
AND AFTER THE EXCLUSION OF THE INCORRECT TRIALS IN
EXPERIMENT 4.2

Sentence type	Region	Before removal		After removal	
		F_1	F_2	F_1	F_2
Natives ($n = 30$)					
- <i>guo</i> sentences	R4	(2, 58) 1.01	(2, 46) 1.32	(2, 58) 0.41	(2, 46) 0.62
	R5	(2, 58) 12.34***	(2, 46) 13.25***	(2, 58) 10.78***	(2, 46) 14.31***
- <i>le</i> sentences	R4	(2, 58) 5.32**	(2, 46) 5.64**	(2, 58) 3.78*	(2, 46) 6.20**
	R5	(2, 58) 5.94**	(2, 46) 5.80**	(2, 58) 5.89**	(2, 46) 6.32**
Advanced L2ers ($n = 28$)					
- <i>guo</i> sentences	R4	(2, 54) 4.29*	(2, 46) 3.25*	(2, 50) 2.29	(2, 44) 1.96
	R5	(2, 54) 1.15	(2, 46) 0.77	(2, 48) 0.82	(2, 44) 0.07
- <i>le</i> sentences	R4	(2, 54) 4.67*	(2, 46) 3.99*	(2, 42) 4.18*	(2, 42) 4.44*
	R5	(2, 54) 2.82 ⁺	(2, 46) 2.43	(2, 42) 2.58 ⁺	(2, 42) 2.35
Intermediate L2ers ($n = 25$)					
- <i>guo</i> sentences	R4	(2, 48) 1.06	(2, 46) 1.96	(2, 38) 1.66	(2, 40) 5.61** ^a
	R5	(2, 48) 0.08	(2, 46) 0.21	(2, 40) 0.01	(2, 42) 0.09
- <i>le</i> sentences	R4	(2, 48) 2.38	(2, 46) 1.89	(2, 32) 0.51	(2, 36) 2.37
	R5	(2, 48) 0.35	(2, 46) 0.61	(2, 30) 0.09	(2, 36) 1.28

Note. Shaded areas indicate changes in the level of statistical significance.

⁺ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

^aThe significant difference here is, however, in the opposite direction, with the *gang-guo* condition having the longest mean RT and the **bu-guo* condition having the shortest mean RT ($M_{\text{gang-guo}} = 801.1$, $SD = 261.4$; $M_{\text{mei-guo}} = 713.9$, $SD = 164.8$; $M_{\text{bu-guo}} = 613.8$, $SD = 141.0$).

APPENDIX R

ANALYSES OF VARIANCE FOR THE EFFECTS OF CONDITION BEFORE AND AFTER THE REMOVAL OF THE ACHIEVEMENT-PREDICATE ITEM

1. Experiment 4.1

Sentence type	Region	Before removal		After removal	
		F_1	F_2	F_1	F_2
Natives ($n = 30$)					
- <i>guo</i> sentences	R5	(2, 58) 16.71 ^{***}	(2, 46) 19.00 ^{***}	(2, 58) 12.07 ^{***}	(2, 44) 15.57 ^{***}
	R6	(2, 58) 6.37 ^{**}	(2, 46) 8.04 ^{**}	(2, 58) 5.46 ^{**}	(2, 44) 7.75 ^{**}
- <i>le</i> sentences	R5	(2, 58) 17.95 ^{***}	(2, 46) 16.43 ^{***}	(2, 58) 19.06 ^{***}	(2, 44) 13.53 ^{***}
	R6	(2, 58) 6.25 ^{**}	(2, 46) 6.10 ^{**}	(2, 58) 6.91 ^{**}	(2, 44) 5.13 [*]
Advanced L2ers ($n = 30$)					
- <i>guo</i> sentences	R5	(2, 58) 4.77 [*]	(2, 46) 7.09 ^{**}	(2, 58) 4.64 [*]	(2, 44) 7.20 ^{**}
	R6	(2, 58) 3.34 [*]	(2, 46) 1.92	(2, 58) 4.30 [*]	(2, 44) 2.36
- <i>le</i> sentences	R5	(2, 58) 4.83 [*]	(2, 46) 5.05 [*]	(2, 58) 4.44 [*]	(2, 44) 3.97 [*]
	R6	(2, 58) 3.31 [*]	(2, 46) 2.48 ⁺	(2, 58) 2.65 ⁺	(2, 44) 2.26
Intermediate L2ers ($n = 24$)					
- <i>guo</i> sentences	R5	(2, 46) 0.84	(2, 46) 0.22	(2, 46) 0.81	(2, 44) 0.10
	R6	(2, 46) 0.65	(2, 46) 0.08	(2, 46) 0.80	(2, 44) 0.02
- <i>le</i> sentences	R5	(2, 46) 0.01	(2, 46) 0.29	(2, 46) 0.00	(2, 44) 0.16
	R6	(2, 46) 1.82	(2, 46) 1.01	(2, 46) 2.35	(2, 44) 1.16

Note. Shaded areas indicate changes in the level of statistical significance.

⁺ $p < .10$; ^{*} $p < .05$; ^{**} $p < .01$; ^{***} $p < .001$.

2. Experiment 4.2

Sentence type	Region	Before removal		After removal	
		F_1	F_2	F_1	F_2
Natives ($n = 30$)					
- <i>guo</i> sentences	R4	(2, 58) 1.01	(2, 46) 1.32	(2, 58) 0.77	(2, 44) 1.13
	R5	(2, 58) 12.34 ^{***}	(2, 46) 13.25 ^{***}	(2, 58) 12.65 ^{***}	(2, 44) 13.11 ^{***}
- <i>le</i> sentences	R4	(2, 58) 5.32 ^{**}	(2, 46) 5.64 ^{**}	(2, 58) 4.89 [*]	(2, 44) 3.00 ⁺
	R5	(2, 58) 5.94 ^{**}	(2, 46) 5.80 ^{**}	(2, 58) 5.97 ^{**}	(2, 44) 4.12 [*]
Advanced L2ers ($n = 28$)					
- <i>guo</i> sentences	R4	(2, 54) 4.29 [*]	(2, 46) 3.25 [*]	(2, 54) 4.18 [*]	(2, 44) 3.70 [*]
	R5	(2, 54) 1.15	(2, 46) 0.77	(2, 54) 0.82	(2, 44) 0.39
- <i>le</i> sentences	R4	(2, 54) 4.67 [*]	(2, 46) 3.99 [*]	(2, 54) 5.00 [*]	(2, 44) 3.12 ⁺
	R5	(2, 54) 2.82 ⁺	(2, 46) 2.43	(2, 54) 3.39 [*]	(2, 44) 2.24
Intermediate L2ers ($n = 25$)					
- <i>guo</i> sentences	R4	(2, 48) 1.06	(2, 46) 1.96	(2, 46) 0.78	(2, 44) 2.67 ⁺
	R5	(2, 48) 0.08	(2, 46) 0.21	(2, 46) 0.13	(2, 44) 0.21
- <i>le</i> sentences	R4	(2, 48) 2.38	(2, 46) 1.89	(2, 46) 2.06	(2, 44) 1.60
	R5	(2, 48) 0.35	(2, 46) 0.61	(2, 46) 0.12	(2, 44) 0.37

Note. Shaded areas indicate changes in the level of statistical significance.

⁺ $p < .10$; ^{*} $p < .05$; ^{**} $p < .01$; ^{***} $p < .001$.

APPENDIX S

EXPERIMENTAL SENTENCES FOR EXPERIMENTS 5.1 AND 5.2 AND THE CORRESPONDING COMPREHENSION QUESTIONS IN EXPERIMENT 5.1

The experimental sentences and the comprehension questions (in parentheses) are given below, with English translations provided for each.

Part 1. Experimental Sentences with Common Verification Questions

- 1a. 她爱护宿舍的同学应该是真心的。(她爱护同学, 对吗?)
She was sincere when she took care of the classmates in her dormitory. (She took care of the classmates, right?)
- 1b. 她答应宿舍的同学应该是真心的。(她答应同学, 对吗?)
She was sincere when she promised the classmates in her dormitory. (She promised her classmates, right?)
- 2a. 他帮助公司的经理是很可笑的。(他帮助经理, 对吗?)
It's ridiculous that he helped the company's manager. (He helped the manager, right?)
- 2b. 他亲吻公司的经理是很可笑的。(他亲吻经理, 对吗?)
It's ridiculous that he kissed the company's manager. (He kissed the manager, right?)
- 3a. 他们称赞文章的作者肯定有目的。(他们称赞作者, 对吗?)
They had a purpose when they praised the article's author. (They praised the author, right?)
- 3b. 他们打击文章的作者肯定有目的。(他们打击作者, 对吗?)
They had a purpose when they attacked the article's author. (They attacked the author, right?)
- 4a. 她们害怕手机的用户有些反常。(她们害怕用户, 对吗?)
It's a little unusual that they feared the cellphone's user. (They feared the user, right?)
- 4b. 她们感谢手机的用户有些反常。(她们感谢用户, 对吗?)
It's a little unusual that they thanked the cellphone's user. (They thanked the user, right?)
- 5a. 我们离开楼房的居民是不对的。(我们离开居民, 对吗?)
It's not right that we left the residents of the building. (We left the residents, right?)
- 5b. 我们反对楼房的居民是不对的。(我们反对居民, 对吗?)
It's not right that we opposed the residents of the building. (We opposed the residents, right?)
- 6a. 他去找房子的主人可能是认真的。(他去找主人, 对吗?)
He may have been serious when he went to find the owner of the house. (He went to find the owner, right?)
- 6b. 他相信房子的主人可能是认真的。(他相信主人, 对吗?)
He may have been serious when he trusted the owner of the house. (He trusted the owner, right?)
- 7a. 我喜爱电视剧的演员是有原因的。(我喜爱演员, 对吗?)
There is a reason that I loved the actor of the TV drama. (I loved the actor, right?)
- 7b. 我同意电视剧的演员是有原因的。(我同意演员, 对吗?)
There is a reason that I agreed with the actor of the TV drama. (I agreed with the actor, right?)
- 8a. 她们想看汽车的司机已经很长时间了。(她们想看司机, 对吗?)
It's been a long time since they wanted to see the car's driver. (They wanted to see the driver, right?)
- 8b. 她们原谅汽车的司机已经很长时间了。(她们原谅司机, 对吗?)
It's been a long time since they forgave the car's driver. (They forgave the driver, right?)

Part 2. Experimental Sentences with Shallow-Targeted vs. Non-Shallow-Targeted Questions

- 9a. 我们告别小区的邻居已经有几年了。(我们告别小区, 对吗?)
It's been a few years since we bid farewell to the neighbor in the district. (We bid farewell to the district, right?)
- 9b. 我们误会小区的邻居已经有几年了。(我们没误会邻居, 对吗?)
It's been a few years since we misunderstood the neighbor in the district. (We didn't misunderstand the neighbor, right?)
- 10a. 他关注农村的学生非常好。(他关注农村, 对吗?)
It's very good that he gave attention to the students in the rural areas. (He gave attention to the rural areas, right?)
- 10b. 他欢迎农村的学生非常好。(他没欢迎学生, 对吗?)
It's very good that he welcomed the students in the rural areas. (He didn't welcome the students, right?)
- 11a. 她们恨医院的医生真地不应该。(她们恨医院, 对吗?)
It really shouldn't have happened that they hated the doctors in the hospital. (They hated the hospital, right?)
- 11b. 她们打医院的医生真地不应该。(她们没打医生, 对吗?)
It really shouldn't have happened that they beat the doctors in the hospital. (They didn't beat the doctors, right?)
- 12a. 她怀念家乡的朋友是真的。(她怀念家乡, 对吗?)
It's true that she missed her friend from home. (She missed home, right?)
- 12b. 她伤害家乡的朋友是真的。(她没伤害朋友, 对吗?)
It's true that she hurt her friend from home. (She didn't hurt her friend, right?)
- 13a. 他们了解工厂的工人是应该的。(他们了解工厂, 对吗?)
It was necessary that they understood the factory's workers. (They understood the factory, right?)
- 13b. 他们拥抱工厂的工人是应该的。(他们没拥抱工人, 对吗?)
It was necessary that they hugged the factory's workers. (They didn't hug the workers, right?)
- 14a. 我们讨厌小说的作者是对的。(我们讨厌小说, 对吗?)
It's reasonable that we disliked the novel's author. (We disliked the novel, right?)
- 14b. 我们劝说小说的作者是对的。(我们没劝说作者, 对吗?)
It's reasonable that we persuaded the novel's author. (We didn't persuade the author, right?)
- 15a. 他们想念学校的同事完全可以理解。(他们想念学校, 对吗?)
It's understandable that they missed the colleague at the school. (They missed the school, right?)
- 15b. 他们去送学校的同事完全可以理解。(他们没去送同事, 对吗?)
It's understandable that they went to send off the colleague at the school. (They didn't go to send off the colleague, right?)
- 16a. 她喜欢高中的同伴已经两年了。(她喜欢高中, 对吗?)
It's been two years since she began to like her high school friend. (She began to like her high school, right?)
- 16b. 她认识高中的同伴已经两年了。(她没认识同伴, 对吗?)
It's been two years since she got to know her high school friend. (She didn't get to know her friend, right?)

APPENDIX T

SENTENCES IN NORMING STUDY PART 1 FOR EXPERIMENTS 5.1 AND 5.2

English translations are given for each sentence.

- 1a. 她不用爱护那些同学。 She doesn't need to take care of those classmates.
- 1b. 她不用答应那些同学。 She doesn't need to promise those classmates.

- 2a. 他不应该帮助那些经理。 He shouldn't help those managers.
- 2b. 他不应该亲吻那些经理。 He shouldn't kiss those managers.

- 3a. 他们常常称赞那些作者。 They often praise those authors.
- 3b. 他们常常打击那些作者。 They often attack those authors.

- 4a. 我们不应该告别那些邻居。 We shouldn't bid farewell to those neighbors.
- 4b. 我们不应该误会那些邻居。 We shouldn't misunderstand those neighbors.

- 5a. 他不应该去关注那些学生。 He shouldn't give attention to those students.
- 5b. 他不应该去欢迎那些学生。 He shouldn't welcome those students.

- 6a. 她们常常害怕那些用户。 They often fear those users.
- 6b. 她们常常感谢那些用户。 They often thank those users.

- 7a. 她们不应该恨那些医生。 They shouldn't hate those doctors.
- 7b. 她们不应该打那些医生。 They shouldn't beat those doctors.

- 8a. 她常常怀念那些朋友。 She often misses those friends.
- 8b. 她常常伤害那些朋友。 She often hurts those friends.

- 9a. 他们不必要去了解那些工人。 They don't need to understand those workers.
- 9b. 他们不应该去拥抱那些工人。 They don't need to hug those workers.

- 10a. 我们没有离开那些居民。 We didn't leave those residents.
- 10b. 我们没有反对那些居民。 We didn't oppose those residents.

- 11a. 他不应该去找那些主人。 He shouldn't go to find those owners.
- 11b. 他不应该相信那些主人。 He shouldn't trust those owners.

- 12a. 我们不应该讨厌那些作者。 We shouldn't dislike those authors.
- 12b. 我们不应该劝说那些作者。 We shouldn't persuade those authors.

- 13a. 我不应该喜爱那些演员。 I shouldn't love those actors.
- 13b. 我不应该同意那些演员。 I shouldn't agree with those actors.

- 14a. 她们不想看那些司机。 They don't want to see those drivers.
- 14b. 她们不原谅那些司机。 They don't forgive those drivers.

- 15a. 他们不应该想念那些同事。 They shouldn't miss those colleagues.
- 15b. 他们不应该去送那些同事。 They shouldn't go to send off those colleagues.

- 16a. 她不应该喜欢那些同伴。 She shouldn't like those friends.
- 16b. 她不应该认识那些同伴。 She shouldn't get to know those friends.

APPENDIX U

SENTENCES IN NORMING STUDY PART 2 FOR EXPERIMENTS 5.1 AND 5.2

1. 她____(A. 答应 B. 爱护)宿舍应该是真心的。
She was sincere when she ____ (A. promised B. took care of) her dormitory.
2. 他____(A. 亲吻 B. 帮助)公司是很可笑的。
It's ridiculous that he ____ (A. kissed B. helped) the company.
3. 他们____(A. 打击 B. 称赞)文章肯定有目的。
They had a purpose when they ____ (A. attacked B. praised) the article.
4. 我们____(A. 告别 B. 误会)小区已经有几年了。
It's been a few years since we ____ (A. bid farewell to B. misunderstood) the district.
5. 他____(A. 关注 B. 欢迎)农村非常好。
It's very good that he ____ (A. gave attention to B. welcomed) the rural areas.
6. 她们____(A. 感谢 B. 害怕)手机有些反常。
It's a little unusual that they ____ (A. thanked B. feared) the cellphone.
7. 她们____(A. 恨 B. 打)医院真地不应该。
It really shouldn't have happened that they ____ (A. hated B. beat) the hospital.
8. 她____(A. 怀念 B. 伤害)家乡是真的。
It's true that she ____ (A. missed B. hurt) her hometown.
9. 他们____(A. 了解 B. 拥抱)工厂是应该的。
It was necessary that they ____ (A. understood B. hugged) the factory.
10. 我们____(A. 反对 B. 离开)楼房是不对的。
It's not right that we ____ (A. opposed B. left) the building.
11. 他____(A. 相信 B. 去找)房子可能是认真的。
He may have been serious when he ____ (A. trusted B. went to find) the house.
12. 我们____(A. 讨厌 B. 劝说)小说是对的。
It's reasonable that we ____ (A. disliked B. persuaded) the novel.
13. 我____(A. 同意 B. 喜爱)电视剧是有原因的。
There is a reason that I ____ (A. agreed with B. loved) the TV drama.
14. 她们____(A. 原谅 B. 想看)汽车已经很长时间了。
It's been a long time since they ____ (A. forgave B. wanted to see) the car.
15. 他们____(A. 想念 B. 去送)学校完全可以理解。
It's understandable that they ____ (A. missed B. went to send off) the school.
16. 她____(A. 喜欢 B. 认识)高中已经两年了。
It's been two years since she ____ (A. began to like B. got to know) her high school.

APPENDIX V
MEAN COMPREHENSION ACCURACY RATES FOR THE EXPERIMENTAL
SENTENCES IN EXPERIMENT 5.1

1. For All the Experimental Sentences

Group	Condition	
	Locally plausible ($k = 16$)	Locally implausible ($k = 16$)
Natives ($n = 30$)	.81 (.14)	.81 (.15)
Advanced L2ers ($n = 18$)	.70 (.17)	.87 (.15)
Intermediate L2ers ($n = 19$)	.59 (.18)	.70 (.13)

Note. Standard deviations are in parentheses.

2. For the Shallow-Targeted Questions vs. the Non-Shallow-Targeted Questions

Group	Question Type	
	Shallow-targeted ($k = 8$)	Non-shallow-targeted ($k = 8$)
Natives ($n = 30$)	.86 (.20)	.94 (.11)
Advanced L2ers ($n = 18$)	.65 (.29)	.96 (.10)
Intermediate L2ers ($n = 19$)	.49 (.27)	.80 (.20)

Note. Standard deviations are in parentheses.

3. For the Common Verification Questions

Group	Condition	
	Locally plausible ($k = 8$)	Locally implausible ($k = 8$)
Natives ($n = 30$)	.76 (.20)	.68 (.23)
Advanced L2ers ($n = 18$)	.75 (.21)	.78 (.30)
Intermediate L2ers ($n = 19$)	.68 (.30)	.59 (.25)

Note. Standard deviations are in parentheses.

APPENDIX W
TRIMMED RAW RT MEANS (IN MILLISECONDS) FOR THE
EXPERIMENTAL SENTENCES IN EXPERIMENT 5.1

1. For All Trials

Condition	Region						
	1	2	3	4	5	6	7
Natives ($n = 30$)							
Locally plausible	623 (201)	439 (143)	432 (138)	389 (93)	413 (121)	457 (129)	606 (242)
Locally implausible	619 (225)	438 (122)	454 (149)	433 (134)	435 (135)	440 (135)	589 (284)
Advanced L2ers ($n = 18$)							
Locally plausible	567 (276)	580 (279)	589 (284)	399 (113)	505 (264)	485 (194)	873 (461)
Locally implausible	554 (229)	624 (340)	598 (300)	465 (147)	517 (227)	500 (188)	817 (392)
Intermediate L2ers ($n = 19$)							
Locally plausible	715 (219)	1021 (467)	959 (434)	535 (131)	748 (352)	658 (214)	1077 (342)
Locally implausible	753 (310)	971 (412)	914 (365)	573 (209)	758 (371)	681 (316)	1139 (407)

Note. Standard deviations are in parentheses.

2. For Correct Trials Only

Condition	Region						
	1	2	3	4	5	6	7
Natives ($n = 30$)							
Locally plausible	630(204)	439(148)	435(138)	384(96)	415(128)	457(133)	606(245)
Locally implausible	620(221)	437(122)	462(161)	434(126)	442(134)	448(133)	606(289)
Advanced L2ers ($n = 18$)							
Locally plausible	560(278)	545(261)	603(314)	405(115)	509(265)	475(208)	880(449)
Locally implausible	553(238)	624(337)	601(316)	465(145)	503(200)	501(188)	817(373)
Intermediate L2ers ($n = 19$)							
Locally plausible	710(240)	1046(467)	980(440)	530(127)	729(339)	627(216)	1108(350)
Locally implausible	741(307)	938(423)	941(419)	567(213)	765(438)	690(346)	1049(420)

Note. Standard deviations are in parentheses.

APPENDIX X
PAIRED-SAMPLE *T* TESTS BASED ON ALL TRIALS VS. ON CORRECT TRIALS ONLY IN EXPERIMENT 5.1

Region	For all trials		For correct trials only	
	<i>t</i> ₁	<i>t</i> ₂	<i>t</i> ₁	<i>t</i> ₂
Natives (<i>n</i> = 30)				
R3	(29) 1.63	(15) 1.13	(29) 1.62	(15) 0.21
R4	(29) 2.65*	(15) 2.77*	(29) 2.97**	(15) 2.77*
R5	(29) 1.17	(15) 1.18	(29) 1.32	(15) 0.58
R6	(29) 1.09	(15) 1.01	(29) 0.48	(15) 1.59
Advanced L2ers (<i>n</i> = 18)				
R3	(17) 0.35	(15) 0.35	(17) 0.08	(15) 1.14
R4	(17) 4.05**	(15) 2.85*	(17) 3.17**	(15) 1.98 ⁺
R5	(17) 0.34	(15) 0.16	(17) 0.18	(15) 1.12
R6	(17) 0.56	(15) 0.11	(17) 0.80	(15) 0.24
Intermediate L2ers (<i>n</i> = 19)				
R3	(18) 0.88	(15) 0.67	(18) 0.73	(15) 0.93
R4	(18) 1.15	(15) 1.02	(18) 1.13	(15) 0.53
R5	(18) 0.34	(15) 0.49	(18) 0.75	(15) 0.08
R6	(18) 0.44	(15) 0.57	(18) 1.10	(15) 0.32

Note. Shaded areas indicate differences in the level of statistical significance.

⁺*p* < .10; **p* < .05; ***p* < .01.

APPENDIX Y
MEAN ACCEPTANCE RATES FOR THE EXPERIMENTAL SENTENCES IN
EXPERIMENT 5.2

Group	Condition	
	Locally plausible ($k = 16$)	Locally implausible ($k = 16$)
Natives ($n = 30$)	.60 (.17)	.60 (.23)
Advanced L2ers ($n = 18$)	.63 (.24)	.65 (.22)
Intermediate L2ers ($n = 18$)	.47 (.22)	.50 (.21)

Note. Standard deviations are in parentheses.

APPENDIX Z
TRIMMED RAW RT MEANS (IN MILLISECONDS) FOR THE
EXPERIMENTAL SENTENCES IN EXPERIMENT 5.2

Condition	Region						
	1	2	3	4	5	6	7
Natives ($n = 30$)							
Locally plausible	665 (228)	543 (231)	569 (259)	475 (162)	595 (267)	576 (214)	777 (361)
Locally implausible	670 (234)	555 (232)	599 (280)	515 (184)	567 (266)	573 (254)	864 (426)
Advanced L2ers ($n = 18$)							
Locally plausible	584 (222)	779 (369)	750 (326)	486 (139)	669 (279)	674 (285)	1094 (475)
Locally implausible	607 (275)	782 (386)	832 (352)	506 (144)	663 (296)	693 (342)	1145 (548)
Intermediate L2ers ($n = 18$)							
Locally plausible	624 (212)	1022 (408)	1003 (336)	615 (203)	876 (337)	710 (180)	1250 (383)
Locally implausible	717 (270)	1068 (419)	1059 (362)	624 (174)	898 (297)	772 (285)	1274 (332)

Note. Standard deviations are in parentheses.

APPENDIX AA
EXPERIMENTAL ITEMS FOR THE ONLINE READING TASKS IN
EXPERIMENTS 6.1 AND 6.2

The experimental sentences are listed below. English translations are provided only for the null-object conditions.

1. 已经 不早了; 他 正要 看(,) 电影 就 放完 了。
It was already a little late; just as he was about to watch [it], the movie was over.
电影 很有名; 他 正要 看(,) 电影 就 放完 了。
The movie was very famous; just as he was about to watch [it], the movie was over.
2. 天 要下雨了; 她 刚要 拿(,) 杯子 就 破 了。
It was going to rain; just as she was about to take [it], the cup broke.
杯子 还是新的; 她 刚要 拿(,) 杯子 就 破 了。
The cup was still new; just as she was about to take [it], the cup broke.
3. 这里 很安静; 他们 还没 听(,) 音乐 就 中断 了。
It was very quiet here; although they hadn't listened to [it], the music was interrupted.
音乐 很好听; 他们 还没 听(,) 音乐 就 中断 了。
The music was very nice; although they hadn't listened to [it], the music was interrupted.
4. 已经 很晚了; 他 正要 喝(,) 牛奶 就 流出来 了。
It was already very late; just as he was about to drink [it], the milk leaked out.
牛奶 不太热; 他 正要 喝(,) 牛奶 就 流出来 了。
The milk was not too hot; just as he was about to drink [it], the milk leaked out.
5. 外面 起风了; 她 刚要 擦(,) 箱子 就 划破 了。
The wind picked up outside; just as she was about to wipe [it], the box was scratched.
箱子 很脏; 她 刚要 擦(,) 箱子 就 划破 了。
The box was very dirty; just as she was about to wipe [it], the box was scratched.
6. 快到 中午了; 她们 正要 玩(,) 游戏 就 开始 了。
It was almost noon; just as they were about to play [it], the game started.
游戏 很不错; 她们 正要 玩(,) 游戏 就 开始 了。
The game was very good; just as they were about to play [it], the game started.
7. 这里 很安静; 他 还没 开(,) 车 就 坏 了。
It was very quiet here; although he hadn't driven [it], the car broke down.
车 准备好了; 他 还没 开(,) 车 就 坏 了。
The car was ready; although he hadn't driven [it], the car broke down.
8. 快要 下课了; 她 正要 用(,) 电脑 就 关 了。
It was almost time for the class to end; just as she was about to use [it], the computer shut down.
电脑 不太稳定; 她 正要 用(,) 电脑 就 关 了。
The computer was not stable; just as she was about to use [it], the computer shut down.

9. 天 就要黑了; 他们 正要 打扫(,) 房间 就 整齐 了。
It was getting dark; just as they were about to clean [it], the room got tidy.
房间 很脏了; 他们 正要 打扫(,) 房间 就 整齐 了。
The room got very dirty; just as they were about to clean [it], the room got tidy.
10. 快到 晚上了; 他 还没 参观(,) 工厂 就 关门 了。
It was almost evening; although he hadn't visited [it], the factory was closed.
工厂 很特别; 他 还没 参观(,) 工厂 就 关门 了。
The factory was very unusual; although he hadn't visited [it], the factory was closed.
11. 天 暖了; 她们 刚要 买(,) 手机 就 便宜 了。
It was getting warm; just as they were about to buy [it], the cellphone got cheaper.
手机 一直很贵; 她们 刚要 买(,) 手机 就 便宜 了。
The cellphone was always expensive; just as they were about to buy [it], the cellphone got cheaper.
12. 时间 不早了; 她们 正要 参加(,) 比赛 就 结束 了。
It was getting late; just as they were about to take part in [it], the competition was over.
比赛 很重要; 她们 正要 参加(,) 比赛 就 结束 了。
The competition was very important; just as they were about to take part in [it], the competition was over.
13. 快到 下午了; 他们 还没 吃(,) 饭 就 凉 了。
It was almost afternoon; just as they were about to eat [it], the meal got cold.
饭 做好了; 他们 还没 吃(,) 饭 就 凉 了。
The meal was ready; just as they were about to eat [it], the meal got cold.
14. 时候 不早了; 他 刚要 打(,) 电话 就 响 了。
It was getting late; just as he was about to call [it], the phone rang.
电话 特别忙; 他 刚要 打(,) 电话 就 响 了。
The phone was very busy; just as he was about to call [it], the phone rang.
15. 外面 下雪了; 她 还没 穿(,) 衣服 就 变旧 了。
It was snowing outside; just as she was about to put [them] on, the clothes got old.
衣服 是刚买的; 她 还没 穿(,) 衣服 就 变旧 了。
The clothes were just bought; just as she was about to put [them] on, the clothes got old.
16. 应该 起床了; 他 刚要 关(,) 电灯 就 没电 了。
It was time to get up; just as he was about to switch [it] off, the light ran out of power.
电灯 不好用; 他 刚要 关(,) 电灯 就 没电 了。
The light doesn't work well; just as he was about to switch [it] off, the light ran out of power.

APPENDIX AB
MEAN COMPREHENSION ACCURACY RATES FOR THE EXPERIMENTAL
SENTENCES IN THE OFFLINE TASK OF EXPERIMENT 6.1

Group	Condition			
	I-N	R-N	I-NN	R-NN
Natives ($n = 24$)	.98 (.10)	.99 (.05)	.93 (.11)	.97 (.08)
English Advanced L2ers ($n = 15$)	.84 (.23)	.90 (.16)	.83 (.26)	.93 (.11)
English Intermediate L2ers ($n = 15$)	.92 (.12)	.90 (.13)	.77 (.20)	.88 (.13)
Japanese Advanced L2ers ($n = 18$)	.86 (.21)	.88 (.15)	.81 (.22)	.90 (.19)
Japanese Intermediate L2ers ($n = 10$)	.78 (.25)	.75 (.24)	.70 (.23)	.73 (.28)

Note. Standard deviations are in parentheses. I-N = Irrelevant-Null; R-N = Relevant-Null; I-NN = Irrelevant-Nonnull; R-NN = Relevant-Nonnull.

APPENDIX AC
MEAN COMPREHENSION ACCURACY RATES FOR THE EXPERIMENTAL
SENTENCES IN THE ONLINE TASK OF EXPERIMENT 6.1

Group	Condition			
	I-N	R-N	I-NN	R-NN
Natives ($n = 24$)	.93 (.12)	.94 (.11)	.94 (.13)	.93 (.12)
English Advanced L2ers ($n = 15$)	.82 (.20)	.83 (.22)	.75 (.25)	.73 (.32)
English Intermediate L2ers ($n = 15$)	.58 (.26)	.63 (.21)	.58 (.20)	.55 (.27)
Japanese Advanced L2ers ($n = 18$)	.67 (.26)	.74 (.22)	.71 (.23)	.64 (.21)
Japanese Intermediate L2ers ($n = 10$)	.43 (.21)	.55 (.26)	.55 (.26)	.40 (.32)

Note. Standard deviations are in parentheses. I-N = Irrelevant-Null; R-N = Relevant-Null; I-NN = Irrelevant-Nonnull; R-NN = Relevant-Nonnull.

APPENDIX AD
TRIMMED RAW RT MEANS (IN MILLISECONDS) FOR THE
EXPERIMENTAL SENTENCES IN EXPERIMENT 6.1

1. For All Trials

Condition	R1	R2	R3	R4	R5	R6	R7	R8	R9
Natives ($n = 24$)									
Irrelevant-null	576 (179)	471 (214)	384 (104)	412 (139)	446 (196)	456 (147)	394 (105)	388 (118)	476 (168)
Relevant-null	545 (164)	474 (228)	442 (134)	427 (149)	398 (142)	431 (161)	364 (87)	408 (121)	425 (106)
Irrelevant-nonnull	637 (211)	463 (254)	405 (158)	405 (170)	376 (152)	406 (164)	388 (134)	399 (127)	488 (152)
Relevant-nonnull	598 (198)	494 (236)	460 (153)	416 (125)	396 (154)	413 (144)	383 (121)	382 (120)	461 (198)
English Advanced L2ers ($n = 15$)									
Irrelevant-null	879 (204)	1194 (329)	582 (134)	784 (210)	769 (222)	897 (202)	637 (132)	950 (333)	760 (273)
Relevant-null	1018 (197)	1255 (259)	630 (108)	851 (237)	786 (150)	731 (163)	500 (85)	903 (302)	683 (182)
Irrelevant-nonnull	906 (261)	1150 (270)	629 (166)	866 (203)	733 (132)	798 (275)	560 (104)	876 (237)	744 (216)
Relevant-nonnull	1001 (238)	1330 (220)	736 (213)	743 (175)	705 (210)	766 (220)	566 (172)	944 (237)	757 (246)
English Intermediate L2ers ($n = 15$)									
Irrelevant-null	914 (226)	1315 (255)	661 (186)	964 (256)	768 (170)	994 (317)	601 (144)	975 (255)	745 (178)
Relevant-null	1054 (285)	1447 (280)	656 (114)	911 (216)	739 (181)	840 (215)	538 (87)	883 (280)	652 (144)
Irrelevant-nonnull	985 (223)	1374 (310)	701 (215)	1000 (304)	774 (228)	883 (207)	581 (132)	1108 (415)	750 (257)
Relevant-nonnull	1084 (296)	1340 (169)	699 (204)	861 (236)	787 (154)	762 (126)	555 (90)	1025 (333)	714 (252)
Japanese Advanced L2ers ($n = 18$)									
Irrelevant-null	561 (260)	662 (230)	474 (112)	524 (197)	524 (234)	521 (197)	485 (162)	539 (225)	533 (183)
Relevant-null	566 (251)	733 (351)	550 (177)	531 (193)	448 (170)	467 (155)	432 (103)	471 (135)	437 (112)
Irrelevant-nonnull	558 (206)	647 (289)	526 (188)	520 (185)	475 (150)	509 (230)	441 (165)	473 (185)	529 (192)
Relevant-nonnull	532 (185)	769 (348)	551 (181)	456 (140)	453 (162)	470 (150)	392 (79)	457 (189)	514 (189)
Japanese Intermediate L2ers ($n = 10$)									
Irrelevant-null	576 (203)	733 (231)	622 (170)	573 (230)	432 (121)	530 (186)	433 (102)	474 (185)	498 (241)
Relevant-null	628 (257)	944 (329)	499 (142)	564 (202)	465 (113)	489 (146)	424 (111)	520 (220)	492 (150)
Irrelevant-nonnull	600 (256)	671 (240)	496 (118)	464 (101)	469 (138)	505 (154)	430 (129)	396 (114)	489 (267)
Relevant-nonnull	631 (169)	768 (219)	561 (104)	493 (204)	524 (240)	423 (133)	420 (144)	479 (126)	417 (71)

Note. Standard deviations are in parentheses.

2. For Correct Trials Only

Condition	R1	R2	R3	R4	R5	R6	R7	R8	R9
Natives ($n = 24$)									
Irrelevant-null	572 (174)	466 (208)	378 (106)	405 (144)	448 (200)	457 (148)	392 (106)	388 (118)	463 (178)
Relevant-null	541 (167)	477 (238)	444 (131)	431 (148)	396 (149)	426 (167)	363 (96)	403 (130)	420 (114)
Irrelevant-nonnull	629 (206)	458 (263)	399 (163)	390 (171)	364 (164)	395 (170)	380 (141)	400 (129)	469 (166)
Relevant-nonnull	593 (194)	487 (229)	454 (141)	404 (109)	393 (145)	403 (138)	381 (111)	378 (110)	456 (169)
English Advanced L2ers ($n = 15$)									
Irrelevant-null	892 (227)	1224 (333)	584 (192)	800 (224)	774 (253)	890 (235)	650 (175)	913 (344)	755 (331)
Relevant-null	1006 (211)	1231 (289)	628 (122)	884 (282)	783 (170)	707 (176)	489 (78)	819 (230)	663 (148)
Irrelevant-nonnull	909 (280)	1048 (253)	621 (185)	896 (260)	723 (171)	812 (337)	542 (123)	886 (399)	742 (235)
Relevant-nonnull	987 (277)	1265 (198)	762 (247)	785 (292)	784 (310)	790 (264)	583 (239)	1027 (232)	789 (254)
English Intermediate L2ers ($n = 15$)									
Irrelevant-null	849 (208)	1365 (409)	701 (234)	975 (318)	846 (404)	970 (378)	618 (213)	932 (200)	811 (233)
Relevant-null	1019 (357)	1520 (319)	651 (142)	919 (214)	764 (190)	891 (310)	522 (91)	858 (328)	601 (172)
Irrelevant-nonnull	989 (266)	1395 (324)	689 (207)	1041 (396)	806 (247)	958 (354)	578 (124)	1256 (538)	785 (358)
Relevant-nonnull	1061 (334)	1320 (267)	644 (243)	783 (237)	641 (219)	681 (199)	562 (104)	961 (422)	721 (263)
Japanese Advanced L2ers ($n = 18$)									
Irrelevant-null	543 (267)	628 (272)	482 (153)	545 (265)	489 (223)	512 (224)	495 (195)	489 (225)	513 (196)
Relevant-null	518 (217)	740 (430)	540 (230)	499 (208)	423 (153)	472 (176)	424 (125)	447 (158)	415 (99)
Irrelevant-nonnull	567 (220)	670 (304)	534 (196)	578 (279)	464 (157)	502 (271)	428 (217)	464 (195)	512 (211)
Relevant-nonnull	500 (170)	646 (331)	549 (209)	434 (127)	486 (248)	452 (166)	376 (101)	417 (129)	503 (256)
Japanese Intermediate L2ers ($n = 10$)									
Irrelevant-null	648 (411)	688 (217)	482 (76)	692 (457)	395 (128)	484 (164)	437 (149)	455 (210)	429 (125)
Relevant-null	672 (329)	1027 (405)	519 (140)	564 (213)	441 (99)	502 (283)	404 (123)	435 (204)	499 (210)
Irrelevant-nonnull	600 (281)	689 (364)	531 (237)	437 (65)	485 (182)	539 (234)	430 (159)	362 (107)	446 (160)
Relevant-nonnull	575 (252)	712 (335)	616 (196)	546 (247)	437 (103)	469 (216)	436 (193)	466 (143)	431 (108)

Note. Standard deviations are in parentheses.

APPENDIX AE
PAIRED-SAMPLE *T* TESTS BASED ON ALL TRIALS VS. ON CORRECT TRIALS ONLY IN EXPERIMENT 6.1

Effect	Region	For all trials		For correct trials only	
		t_1	t_2	t_1	t_2
Natives ($n = 24$)					
Discourse-context	R5	(23) 2.20*	(15) 3.01**	(23) 2.36*	(15) 2.64*
	R6	(23) 1.09	(15) 1.26	(23) 1.40	(15) 1.60
	R7	(23) 1.92 ⁺	(15) 3.30**	(23) 1.67	(15) 3.59**
Priming	R6	(23) 0.37	(15) 0.39	(23) 0.41	(15) 0.01
English Advanced L2ers ($n = 15$)					
Discourse-context	R5	(14) 0.44	(15) 0.28	(14) 0.16	(15) 0.39
	R6	(14) 2.65*	(15) 1.37	(14) 3.04**	(15) 1.14
	R7	(14) 3.55**	(15) 3.25**	(14) 3.65**	(15) 3.43**
Priming	R6	(14) 0.43	(15) 0.26	(13) 0.42	(13) 0.42
English Intermediate L2ers ($n = 15$)					
Discourse-context	R5	(14) 0.54	(15) 0.62	(14) 0.72	(13) 0.59
	R6	(14) 2.05 ⁺	(15) 1.88 ⁺	(13) 1.19	(14) 0.93
	R7	(14) 1.91 ⁺	(15) 2.96*	(14) 2.06 ⁺	(14) 1.81 ⁺
Priming	R6	(14) 2.98*	(15) 1.20	(14) 2.69*	(13) 1.67
Japanese Advanced L2ers ($n = 18$)					
Discourse-context	R5	(17) 2.03 ⁺	(15) 1.79 ⁺	(16) 1.25	(13) 0.72
	R6	(17) 2.06 ⁺	(15) 1.75 ⁺	(16) 1.12	(13) 0.33
	R7	(17) 2.21*	(15) 0.97	(16) 2.02 ⁺	(13) 0.76
Priming	R6	(17) 0.95	(15) 0.55	(17) 0.83	(14) 0.86
Japanese Intermediate L2ers ($n = 10$)					
Discourse-context	R5	(9) 1.63	(15) 0.73	(7) 2.17 ⁺ a	(7) 0.05
	R6	(9) 0.70	(15) 0.57	(7) 0.25	(7) 1.24
	R7	(9) 0.27	(15) 0.45	(7) 0.57	(7) 2.43*
Priming	R6	(9) 2.13 ⁺	(15) 1.40	(6) 1.16	(9) 1.08

Note. Shaded areas indicate changes in the level of statistical significance.

⁺ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

^aThe difference here is in the opposite direction, with the irrelevant-context condition having shorter mean RT than the relevant condition (see Appendix AD).

APPENDIX AF
MEAN COMPREHENSION ACCURACY RATES FOR THE EXPERIMENTAL
SENTENCES IN THE OFFLINE TASK OF EXPERIMENT 6.2

Group	Condition			
	I-N	R-N	I-NN	R-NN
Natives ($n = 24$)	.92 (.18)	.91 (.26)	.92 (.16)	.93 (.16)
English Advanced L2ers ($n = 15$)	.92 (.15)	.83 (.18)	.84 (.22)	.88 (.14)
English Intermediate L2ers ($n = 15$)	.91 (.12)	.87 (.20)	.75 (.20)	.79 (.26)
Japanese Advanced L2ers ($n = 18$)	.88 (.17)	.87 (.21)	.88 (.17)	.88 (.19)
Japanese Intermediate L2ers ($n = 10$)	.91 (.19)	.81 (.18)	.84 (.19)	.78 (.28)

Note. Standard deviations are in parentheses. I-N = Irrelevant-Null; R-N = Relevant-Null; I-NN = Irrelevant-Nonnull; R-NN = Relevant-Nonnull.

APPENDIX AG
MEAN ACCEPTANCE RATES FOR THE EXPERIMENTAL SENTENCES IN
THE ONLINE TASK OF EXPERIMENT 6.2

Group	Condition			
	I-N	R-N	I-NN	R-NN
Natives ($n = 24$)	.24 (.24)	.58 (.26)	.28 (.24)	.46 (.26)
English Advanced L2ers ($n = 16$)	.25 (.22)	.44 (.32)	.3 (.25)	.52 (.34)
English Intermediate L2ers ($n = 14$)	.43 (.32)	.63 (.29)	.39 (.23)	.55 (.22)
Japanese Advanced L2ers ($n = 22$)	.47 (.33)	.64 (.33)	.38 (.33)	.57 (.33)
Japanese Intermediate L2ers ($n = 8$)	.41 (.13)	.41 (.27)	.34 (.23)	.34 (.23)

Note. Standard deviations are in parentheses. I-N = Irrelevant-Null; R-N = Relevant-Null; I-NN = Irrelevant-Nonnull; R-NN = Relevant-Nonnull.

APPENDIX AH
TRIMMED RAW RT MEANS (IN MILLISECONDS) FOR THE
EXPERIMENTAL SENTENCES IN EXPERIMENT 6.2

Condition	Region								
	1	2	3	4	5	6	7	8	9
Natives (<i>n</i> = 24)									
Irrelevant-null	611 (200)	630 (215)	516 (125)	526 (184)	608 (281)	614 (293)	519 (188)	515 (142)	726 (370)
Relevant-null	672 (217)	671 (345)	532 (145)	550 (216)	529 (210)	518 (166)	449 (123)	498 (194)	636 (267)
Irrelevant-nonnull	662 (210)	642 (231)	523 (169)	539 (157)	524 (179)	536 (207)	468 (129)	490 (192)	663 (345)
Relevant-nonnull	603 (181)	649 (255)	532 (123)	497 (136)	489 (172)	517 (177)	468 (116)	498 (185)	675 (254)
English Advanced L2ers (<i>n</i> = 16)									
Irrelevant-null	951 (264)	1424 (449)	688 (207)	935 (313)	960 (297)	1110 (268)	806 (352)	994 (361)	1046 (339)
Relevant-null	1116 (386)	1405 (355)	680 (231)	950 (256)	840 (253)	896 (219)	625 (149)	1055 (312)	960 (375)
Irrelevant-nonnull	944 (269)	1297 (281)	758 (269)	958 (238)	829 (283)	905 (221)	698 (179)	1040 (332)	1257 (452)
Relevant-nonnull	1076 (343)	1382 (279)	719 (181)	951 (274)	787 (225)	887 (206)	606 (118)	996 (351)	1005 (387)
English Intermediate L2ers (<i>n</i> = 14)									
Irrelevant-null	984 (249)	1357 (304)	742 (198)	985 (238)	917 (281)	1069 (304)	633 (142)	894 (255)	919 (228)
Relevant-null	1164 (216)	1575 (298)	716 (170)	928 (200)	887 (218)	907 (169)	616 (135)	911 (225)	942 (330)
Irrelevant-nonnull	1112 (351)	1431 (401)	696 (159)	965 (317)	731 (199)	921 (219)	673 (194)	989 (284)	1030 (361)
Relevant-nonnull	1216 (333)	1485 (288)	711 (157)	996 (269)	836 (228)	877 (152)	772 (283)	1031 (324)	985 (394)
Japanese Advanced L2ers (<i>n</i> = 22)									
Irrelevant-null	581 (186)	821 (301)	534 (171)	585 (202)	616 (255)	638 (266)	542 (198)	617 (231)	673 (253)
Relevant-null	587 (194)	841 (343)	532 (128)	584 (226)	542 (211)	554 (184)	491 (121)	592 (251)	702 (319)
Irrelevant-nonnull	603 (205)	740 (283)	528 (144)	563 (163)	539 (188)	594 (220)	550 (200)	597 (256)	731 (308)
Relevant-nonnull	582 (212)	910 (369)	584 (171)	591 (232)	517 (148)	548 (207)	509 (188)	524 (182)	727 (354)
Japanese Intermediate L2ers (<i>n</i> = 8)									
Irrelevant-null	618 (237)	806 (261)	653 (123)	609 (216)	677 (183)	796 (260)	626 (149)	686 (298)	703 (317)
Relevant-null	572 (186)	1190 (365)	560 (127)	629 (149)	624 (163)	744 (236)	490 (117)	667 (258)	673 (249)
Irrelevant-nonnull	577 (207)	975 (355)	621 (124)	632 (185)	568 (120)	748 (146)	607 (187)	673 (219)	776 (227)
Relevant-nonnull	620 (211)	1184 (303)	655 (132)	629 (205)	565 (143)	638 (174)	470 (80)	653 (177)	805 (322)

Note. Standard deviations are in parentheses

REFERENCES

- Abe, J. (2009). Identification of null arguments in Japanese. In H. Hoshi (Ed.), *The dynamics of the language faculty: Perspectives from linguistics and cognitive neuroscience* (pp. 135–162). Tokyo, Japan: Kuroshio Publishers.
- Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research, 30*, 275–306.
- Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences, 2*, 146–152.
- Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31*, 685–712.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.
- Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191–238.
- Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J.-A. (1998). Late closure in context. *Journal of Memory and Language, 38*, 459–484.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–37). Cambridge, England: Cambridge University Press.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: John Wiley & Sons.
- Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language, 44*, 297–324.
- Blackwell, A., & Bates, E. (1995). Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience, 7*, 228–257.

- Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*, 175–198.
- Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia, 51*, 2492–2511.
- Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse: Context effects and their limits. *Journal of Memory and Language, 31*, 293–314.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal, 64*, 311–317.
- Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 687–696.
- Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*, 161–174.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368–407.
- Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*, 107–126.
- Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.
- Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*, 564–570.
- Clifton, C., Jr., Traxler, M. J., Taha Mohamed, M., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*, 317–334.
- Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics, 34*, 615–646.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen, & A.

- M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*. Cambridge, UK: Cambridge University Press.
- Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning, 63*, 766–799.
- Dekydtspotter, L. (2001). The Universal Parser and interlanguage: Domain-specific mental organization in the comprehension of *combien* interrogatives in English-French interlanguage. *Second Language Research, 17*, 91–143.
- Dekydtspotter, L., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition, 30*, 453–480.
- Dekydtspotter, L., & Miller, A. K. (2013). Inhibitive and facilitative priming induced by traces in the processing of wh-dependencies in a second language. *Second Language Research, 29*, 345–372.
- Dekydtspotter, L., & Renaud, C. (2014). On second language processing and grammatical development: The parser in second language acquisition. *Linguistic Approaches to Bilingualism, 4*, 131–165.
- Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). The Comparative Fallacy in L2 processing research. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 33–40). Somerville, MA: Cascadilla Proceedings Project.
- Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology, 48*, 1203–1207.
- Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition, 25*, 529–557.

- Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of *wh*-gaps by Chinese-English second language speakers. *Second Language Research*, 26, 443–472.
- Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 10, 101–116.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23, 290–325.
- Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540–551.
- Ernst, T. (1995). Negation in Mandarin Chinese. *Natural Language & Linguistic Theory*, 13, 665–707.
- Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2012). Processing relative clauses in supportive contexts. *Cognitive Science*, 36, 471–497.
- Felser, C., & Clahsen, H. (2009). Grammatical processing of spoken language in child and adult language learners. *Journal of Psycholinguistic Research*, 38, 305–319.
- Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, 34, 67–98.
- Felser, C., & Roberts, L. (2007). Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23, 9–36.
- Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453–489.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164–203.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.

- Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good enough language processing: A satisficing approach. In N. Taatgen & H. v. Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 413–418). Austin, TX: Cognitive Science Society.
- Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A. J., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 324–335.
- Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71–83.
- Fodor, J. D. (1998). Learning To Parse? *Journal of Psycholinguistic Research*, *27*, 285–319.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124.
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*, 379–399.
- Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*, 226–248.
- Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, *17*, 213–247.
- Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Philadelphia, PA: John Benjamins.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, *24*, 437–468.

- Frazier, L., & Clifton, C., Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93–126.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291–325.
- Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, 21, 175–198.
- Frenck-Mestre, C., German, E. S., & Foucart, A. (2014). Qualitative differences in native and nonnative semantic processing as revealed by ERPs. In R. R. Heredia & J. Altarriba (Eds.), *Foundations of bilingual memory* (pp. 237–255). New York: Springer.
- Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 50A, 119–148.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84.
- Frisson, S., & Traxler, M. J. (2013). Semantic interpretation of sentences. In R. P. G. van Gompel (Ed.), *Sentence processing* (pp. 160–188). New York: Psychology Press.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gibson, E., & Pearlmuter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, 2, 262–268.
- Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish: Evidence from Event Related Potentials. *Neuropsychologia*, 1651–1659.
- Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22, 1870–1887.
- Goad, H., & White, L. (2004). Ultimate attainment of L2 inflection: Effects of L1 prosodic structure *EUROSLA Yearbook*, 4, 119–145.

- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*, 311–347.
- Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science*, *22*, 389–424.
- Gregg, K. R. (1996). The logical and developmental problems of second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 49–81). San Diego, CA: Academic Press.
- Gregg, K. R. (2003). SLA theory: Construction and assessment. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 831–865). Malden, MA: Blackwell.
- Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition*, *95*, 275–296.
- Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*, 251–266.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*, 123–141.
- Harrington, M. (2001). Sentence processing. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 91–124). Cambridge, England: Cambridge University Press.
- Hawkins, R., & Chan, C. Y.-h. (1997). The partial availability of Universal Grammar in second language acquisition: The “failed functional features hypothesis”. *Second Language Research*, *13*, 187–226.
- Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The lexicon-syntax interface in second language acquisition* (pp. 21–44). Philadelphia, PA: John Benjamins.
- Hoeks, J. C. J., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language*, *46*, 99–119.

- Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition: Grammar and processing*. Groningen, The Netherlands: Grodil Press.
- Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12, 463–483.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120, 901–931.
- Hsu, C.-C., & Hsieh, S. C.-Y. (2013). The effects of syntactic knowledge and extra-grammatical factors on sentence acceptability judgment: An empirical study on Chinese topic constructions. *Taiwan Journal of Linguistics*, 11(1), 91–124.
- Huang, C.-T. J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 15, 531–574.
- Huang, C.-T. J. (1988). *Wǒ pǎo de kuài* and Chinese phrase structure. *Language*, 64, 274–311.
- Huang, C.-T. J. (1991). Remarks on the status of the null object. In R. Freidin (Ed.), *Principles and parameters in comparative grammar* (pp. 56–76). Cambridge, MA: MIT Press.
- Huang, C.-T. J. (1998). *Logical relations in Chinese and the theory of grammar*. New York, NY: Garland Publishers.
- Huang, C.-T. J., Li, Y.-H. A., & Li, Y. (2009). *The syntax of Chinese*. Cambridge, England: Cambridge University Press.
- Jackson, C. N., & Bobb, S. C. (2009). The processing and comprehension of *wh*-questions among second language speakers of German. *Applied Psycholinguistics*, 30, 603–636.
- Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, 12, 65–82.
- Jackson, C. N., & Roberts, L. (2010). Animacy affects the processing of subject-object ambiguities in the second language: Evidence from self-paced reading with German second language learners of Dutch. *Applied Psycholinguistics*, 31, 671–691.

- Jacob, G., & Felser, C. (in press). Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology*.
- Jafarpur, A. (1999). Can the C-test be improved with classical item analysis? *System*, 27, 79–89.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603–634.
- Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57, 1–33.
- Juffs, A. (1998a). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning*, 48, 107–147.
- Juffs, A. (1998b). Some effects of first language argument structure and morphosyntax on second language sentence processing. *Second Language Research*, 14, 406–424.
- Juffs, A. (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics*, 21, 207–223.
- Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, 102, 199–225.
- Juffs, A. (2005). The influence of first language on the processing of *wh*-movement in English as a second language. *Second Language Research*, 21, 121–151.
- Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in *wh*-extraction. *Studies in Second Language Acquisition*, 17, 483–516.
- Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning*, 46, 283–320.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238.

- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*, 113–147.
- Katsika, K. (2009). Exploring the minimal structure in prepositional phrase attachment ambiguities: Evidence from Greek. *Lingua*, *119*, 1482–1500.
- Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*, 503–535.
- Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, *37*, 1–32.
- Kinney, P. R., & Gray, C. D. (2004). *SPSS 12 made simple*. New York: NY: Psychology Press.
- Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: An appraisal. *Language Testing*, *14*, 47–84.
- Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*, 68–74.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*, 223–237.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.
- Lee-Ellis, S. (2009). The development and validation of a Korean C-Test using Rasch Analysis. *Language Testing*, *26*, 245–274.
- Lee, P. P.-l., & Pan, H. (2001). The Chinese negation marker *bu* and its association with focus. *Linguistics*, *39*, 703–731.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

- Li, M. (1999). *Negation in Chinese*. Doctoral dissertation, University of Manchester, Manchester, U.K.
- Li, P. (1998). Crosslinguistic variation and sentence processing: The case of Chinese. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective* (pp. 33–53). San Diego, CA: Academic Press.
- Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, *32*, 169–192.
- Lin, J.-w. (2003). Aspectual selection and negation in Mandarin Chinese. *Linguistics*, *41*, 425–459.
- Lu, C.-C., Bates, E., Li, P., Tzeng, O., Hung, D., Tsai, C.-H., et al. (2000). Judgments of grammaticality in aphasia: The special case of Chinese. *Aphasiology*, *14*, 1021–1054.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.
- Mak, W. M., Vonk, W., & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory & Cognition*, *36*, 170–181.
- Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, *19*, 144–161.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*, 53–78.
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226–228.
- Marslen-Wilson, W. D., Brown, C. M., & Tyler, L. K. (1988). Lexical representations in spoken language comprehension. *Language and Cognitive Processes*, *3*, 1–16.
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*, 381–401.
- McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., et al. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(Suppl. 2), 123–150.

- Miao, X. (1999). Sentence understanding in Chinese. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 279–295). Mahwah, NJ: Lawrence Erlbaum Associates.
- Miller, A. K. (2011). *Processing filler-gap dependencies in second language French: Computational complexity and lexical access*. Unpublished Doctoral dissertation, Indiana University, Bloomington, IN.
- Miller, A. K. (2014). Accessing and maintaining referents in L2 processing of *wh*-dependencies. *Linguistic Approaches to Bilingualism*, 4, 167–191.
- Miller, A. K. (2015). Facilitating the task for second language processing research: A comparison of two testing paradigms. *Applied Psycholinguistics*, 36, 613–637.
- Moreno, E. M., Rodríguez-Fornells, A., & Laine, M. (2008). Event-related potentials (ERPs) in the study of bilingual language processing. *Journal of Neurolinguistics*, 21, 477–508.
- Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research*, 21, 152–174.
- Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, 17, 1229–1244.
- Myers, E. B., & Blumstein, S. E. (2005). Selectional restriction and semantic priming effects in normals and Broca's aphasics. *Journal of Neurolinguistics*, 18, 277–296.
- Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, 24, 1205–1223.
- O'Grady, W. (2012). Language acquisition without an acquisition device. *Language Teaching*, 45(01), 116–130.
- O'Grady, W. (2013a). Facts and issues: Response to the commentaries. *Linguistic Approaches to Bilingualism*, 3, 373–383.
- O'Grady, W. (2013b). The illusion of language acquisition. *Linguistic Approaches to Bilingualism*, 3, 253–285.

- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*, 1212–1228.
- Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*, 563–588.
- Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*, 221–236.
- Pan, H.-Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, *19*, 298–313.
- Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research*, *21*, 98–120.
- Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*, 501–528.
- Phillips, C., & Ehrenhofer, L. (in press). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*.
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 940–961.
- Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic parsing. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2 ed., pp. 455–503). New York, NY: Elsevier.
- Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16*, 167–182.
- Prévost, P., & White, L. (2000). Accounting for morphological variation in second language acquisition: Truncation or missing inflection? In M.-A. Friedemann & L. Rizzi (Eds.), *The acquisition of syntax: Studies in comparative developmental linguistics* (pp. 202–235). Harlow, England: Longman.

- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago, IL: University of Chicago Press.
- Rah, A., & Adone, D. (2010). Processing of the reduced relative clause versus main verb ambiguity in L2 learners at different proficiency levels. *Studies in Second Language Acquisition*, 32, 79–109.
- Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85, 618–660.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Roberts, L. (2010). Parsing the L2 input, an overview: Investigating L2 learners processing of syntactic ambiguities and dependencies in real-time comprehension. *Language, Interaction and Acquisition / Langage, Interaction et Acquisition*, 1, 189–205.
- Roberts, L. (2012a). Individual differences in second language sentence processing. *Language Learning*, 62(Suppl. 2), 172–188.
- Roberts, L. (2012b). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, 28, 113–127.
- Roberts, L. (2013). Sentence processing in bilinguals. In R. P. G. van Gompel (Ed.), *Sentence processing* (pp. 221–246). New York: Psychology Press.
- Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, 32, 299–331.
- Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, 30, 333–357.
- Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research*, 36, 175–188.
- Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, 35, 213–235.

- Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*, 2030–2048.
- Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*, 2022–2039.
- Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning, 61*, 80–116.
- Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes, 42*, 99–108.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences, 6*, 382–386.
- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*, 40–72.
- Sharwood Smith, M., & Truscott, J. (2014). *The multilingual mind: A modular processing perspective*. New York: Cambridge University Press.
- Slabakova, R. (2009). What is easy and what is hard to acquire in a second language? In M. Bowles, T. Ionin, S. Montrul, & A. Tremblay (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)* (pp. 280–294). Somerville, MA: Cascadilla Proceedings Project.
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language, 69*, 104–120.
- Smith, C. S. (1994). Aspectual viewpoint and situation type in Mandarin Chinese. *Journal of East Asian Linguistics, 3*, 107–146.
- Song, Y. (2015). L2 processing of plural inflection in English. *Language Learning, 65*, 233–267.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism, 1*, 1–33.

- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*, 447–481.
- Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*, 135–162.
- Su, I.-R. (2001). Context effects on sentence processing: A study based on the Competition Model. *Applied Psycholinguistics*, *22*, 167–189.
- Su, I.-R. (2004). The effects of discourse processing with regard to syntactic and semantic cues: A competition model study. *Applied Psycholinguistics*, *25*, 587–601.
- Sun, C. (2006). *Chinese: A linguistic introduction*. Cambridge, England: Cambridge University Press.
- Swets, B., Desmet, T., Clifton, C., Jr., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201–216.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.
- Swinney, D. A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition*, *7*, 159–165.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Handbook in perception and cognition, Vol. 11: Speech language and communication* (2 ed., pp. 217–262). New York: Academic Press.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
- Traxler, M. J., & Tooley, K. M. (2007). Lexical mediation and context effects in sentence processing. *Brain Research*, *1146*, 59–74.

- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition*, 33, 339–372.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.
- Truscott, J., & Sharwood Smith, M. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition*, 7, 1–20.
- Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research*, 23, 215–242.
- van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147–182.
- van Gompel, R. P. G., & Pickering, M. J. (2007). Sentence parsing. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 289–307). Oxford: Oxford University Press.
- van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, 26, 43–74.
- VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52, 755–803.
- VanPatten, B., Keating, G. D., & Leiser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism*, 2, 109–140.
- Vendler, Z. (1967). *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.
- Wang, S., Chen, H.-C., Yang, J., & Mo, L. (2008). Immediacy of integration in discourse comprehension: Evidence from Chinese readers’ eye movements. *Language and Cognitive Processes*, 23, 241–257.

- Wang, S., Mo, D., Xiang, M., Xu, R., & Chen, H.-C. (2013). The time course of semantic and syntactic processing in reading Chinese: Evidence from ERPs. *Language and Cognitive Processes, 28*, 577–596.
- Wang, W. S. Y. (1965). Two aspect markers in Mandarin. *Language, 41*, 457–470.
- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review, 14*, 770–775.
- Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition, 111*, 132–137.
- Weinberg, A. (1999). A minimalist theory of human sentence processing. In S. D. Epstein & N. Hornstein (Eds.), *Working minimalism* (pp. 287–315). Cambridge, MA: MIT Press.
- Wen, Z. (2007). Online processing of subject-verb agreement in English by first and second language speakers. In H. Caunt-Nulton, S. Kulatilake, & I.-h. Woo (Eds.), *Proceedings of the 31st annual Boston University Conference on Language Development* (Vol. 2, pp. 668–679). Somerville, MA: Cascadilla Press.
- Wen, Z., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. D. (2010). Proficiency effects and distance effects in nonnative processing of English number agreement. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *Proceedings of the 34th annual Boston University Conference on Language Development* (pp. 445–456). Somerville, MA: Cascadilla Press.
- Wen, Z., & Schwartz, B. D. (2012). Task demands and ‘depth’ of L2 processing of subject-verb number agreement. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th annual Boston University Conference on Language Development* (pp. 673–685). Somerville, MA: Cascadilla Press.
- Wen, Z., & Schwartz, B. D. (2014). Fully-specified L2 processing of negation-aspect interactions in Chinese. In C.-Y. Chu, C. E. Coughlin, B. L. Prego, U. Minai, & A. Tremblay (Eds.), *Selected proceedings of the 5th Conference on Generative Approaches to Language Acquisition North America (GALANA 2012)* (pp. 128–139). Somerville, MA: Cascadilla Proceedings Project.

- Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9, 71–88.
- Williams, J. N., Mobius, P., & Kim, C. (2001). Native and non-native processing of English *wh*- questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22, 509–540.
- Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, 33, 419–456.
- Xiao, R., & McEnery, T. (2008). Negation in Chinese: A corpus-based study. *Journal of Chinese Linguistics*, 36, 274–320.
- Xiao, R., Rayson, P., & McEnery, T. (2009). *A frequency dictionary of mandarin Chinese: Core vocabulary for learners*. New York: Routledge.
- Yang, C. L., Gordon, P. C., Hendrick, R., & Wu, J. T. (1999). Comprehension of referring expressions in Chinese. *Language and Cognitive Processes*, 14, 715–743.
- Yang, F., Mo, L., & Louwerse, M. M. (2013). Effects of local and global context on processing sentences with subject and object relative clauses. *Journal of Psycholinguistic Research*, 42, 227–237.
- Yang, J., Wang, S., Chen, H.-C., & Rayner, K. (2009). The time course of semantic and syntactic processing in Chinese sentence comprehension: Evidence from eye movements. *Memory & Cognition*, 37, 1164–1176.
- Ye, Z., Luo, Y.-j., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain Research*, 1071, 186–196.
- Yu, J., & Zhang, Y. (2008). When Chinese semantics meets failed syntax. *Neuroreport*, 19, 745–749.
- Zhang, Y., Li, P., Piao, Q., Liu, Y., Huang, Y., & Shu, H. (2013). Syntax does not necessarily precede semantics in sentence processing: ERP evidence from Chinese. *Brain and Language*, 126, 8–19.
- Zhang, Y., Shu, H., Zhang, H., & Zhou, X. (2002). The processing of ambiguous phrases in referential discourse context in Chinese. *Xinli Xuebao*, 34, 126–134.

- Zhang, Y., Yu, J., & Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 765–781.
- Zhao, Y. (2011). A tree in the wood: A review of research on L2 Chinese acquisition. *Second Language Research*, *27*, 559–572.
- Zhou, X., Jiang, X., Ye, Z., Zhang, Y., Lou, K., & Zhan, W. (2010). Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An ERP study. *Neuropsychologia*, *48*, 1551–1562.
- Zhou, X., & Shu, H. (2011). Neurocognitive processing of the Chinese language. *Brain and Language*, *119*, 59.
- Zhou, X., Ye, Z., Cheung, H., & Chen, H.-C. (2009). Processing the Chinese language: An introduction. *Language and Cognitive Processes*, *24*, 929–946.
- Zhuang, H., & Liu, Z. (2011). Negative marker *bu* in Chinese: Its nature and features. *International Journal of Asian Language Processing*, *21*, 107–160.