BEHAVIOURAL AND ELECTROPHYSIOLOGICAL INVESTIGATIONS OF CHINESE TRANSLATION ACTIVATION DURING ENGLISH WORD RECOGNITION IN CHINESE-ENGLISH BILINGUALS

YUN WEN

温韫

BA, MA

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

JULY, 2017

Abstract

Bilinguals have the unique ability to translate words between their languages. Although translation is a seemingly deliberate and conscious process, recent research has shown that first language (L1) translation equivalents can be automatically and quickly activated during second language (L2) word reading. Automatic translation activation strongly supports the idea of non-selective lexical access. This thesis investigates L1 (Chinese) translation activation during L2 (English) word reading in Chinese-English bilinguals, mainly through using the hidden translation repetition paradigm. In a series of behavioural and electrophysiological experiments using carefully selected stimuli, English words were subliminally or visibly presented to Chinese-English bilinguals in an attempt to seek the source of automatic translation activation (phonology: segment and/or tone, and/or orthography) and to explore to what extent translation activation is automatic. In contrast to previous studies, the behavioural investigations revealed that automatic translation only occurs for target words, which were visible to bilinguals, but not for the invisible masked primes. In addition, in the electrophysiological study, the event-related brain potentials and event-related brain oscillations provided evidence for the dominant role of Chinese segmental (consonants and vowels) activation during English word reading. Possible interpretations for these new findings are provided. Theoretical and methodological implications of the present thesis are also discussed.

Declaration

I hereby declare that this is my own work conducted for my PhD project at the University of Nottingham. This thesis has resulted so far in two papers published in peer-reviewed journals, namely the English-Chinese translation norms reported in Chapter 2, and the meta-analysis presented in Chapter 6.

Manuscripts submitted or accepted for publication, or in preparation

Wen, Y., Filik, R. & van Heuven, W. J. B. (in preparation). Cross-language phonological activation through translation in Chinese-English bilinguals: Evidence from brain potentials and brain oscillations.

Wen, Y., & van Heuven, W. J. B. (in preparation). Limited automatic translation in masked priming: Behavioural evidence from Chinese-English bilinguals.

Wen, Y., & van Heuven, W. J. B. (2016). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analytic review. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-016-1151-1

Wen, Y., & van Heuven, W. J. B. (2016). Chinese translation norms for 1429 English words. *Behavior Research Methods*. doi: 10.3758/s13428-016-0761-x

Acknowledgements

I never expect my PhD journey to be easy, but I did not know it was going to be that difficult. Throughout this journey, I have always felt deeply grateful to many people, who have contributed to my survival, directly or indirectly.

First and foremost, massive thanks go to my principal supervisor Dr. Walter J. B. van Heuven for giving me the opportunity to work with him and guiding me towards an independent researcher. All along my PhD journey, Walter has been great source of knowledge, inspiration and encouragement. His door is always open for suggestion. To me, Walter is the definition of supervisor. My gratitude also goes to my second supervisor Dr. Ruth Filik, who has been another great source of support, feedbacks and encouragement. Because of my fantastic supervisors, I have everything that a PhD student would hope for.

I want to express my thanks to all the people who have helped in various way during my academic journey. Thanks to the University of Nottingham and School of Psychology for funding my PhD project. Thanks to my internal examiner Dr. Nicola Pitchford for all the insightful comments and warm encouragement. Thanks to members in Bilingual Research Group and Cognition and Language Group, in particular Dr. Kathy Conklin, for all the comments and feedbacks. Thanks to Dr. Jonathan Peirce for his kind help in solving technical issues and his thought-provoking words. Thanks to Prof. Ton Dijkstra for his kind help and encouragement when I applied for my PhD, and for having time for discussions during my visits in Nijmegen. Thanks to Prof. Yanping Dong, who supervised my first psycholinguistic experiment and my first ERP study for my master project. It is under Prof. Dong's supervision that my curiosity for language processing grew, which motivated me to embark on a PhD journey. Thanks to Prof. Ping Li who kindly gave me support, encouragement and suggestion for an academic career. Thanks to Prof. Yuanhua Xie who supervised my first empirical study in my undergraduate project. I specially thank all the lovely Chinese-English bilinguals who participated my experiments, especially those who circulated the information and asked their friends to come.

I am thankful to many people that I have met in my PhD years. Thank you my dearest friend Elisa for travelling this journey with me. Things would have been much more difficult without you. Thank you my wonderful officemates Paula and Andrey for all

the chats and help. Thank you Nayana for your companion for the most difficult year in my PhD journey. Thanks to Jingjia Yuan and Huibin Jia for answering my questions on EEG with great patience. Thanks to Xiangbin Teng for sharing the useful information of EEG analyses. My thanks also go to Beerelim, Lawrence, Antonios, Nandini, Persa, Alex, Liz, Sacha, Anne, Nan Lin, Fengling, Qiyang Nie and in particular to Yang and Beinan (and thanks 52brain.com).

I am extremely grateful to my friends who are far away. Thank you, Guanyin Zhao. You're the ultimate source of wisdom! Thanks to Dr. Xian Zhang for offering help and encouragement all along. Thanks to Dr. Nan Zhao and Dr. Zhenguang Cai for all the suggestion and encouragement at every critical stage of my career. Thanks to Dr. Jing Yang for everything. I am very lucky to have a special group of friends from Tucaoxinlingjiayuan who always listen: Ya, Annette, Flora, Su, Wenfeng, Sibo, Junqing, Out, Luoluo, Xiongmao and Taiyang. I am also extremely blessed to have so many friends who bring joys and support to me for all these years: Caixia Zhang, Zhihang Mai, Xiaoqiong Chen, Huihang Xue, Lu Cao, Yunzhen Mo, Xiaomeng Zeng, Zhibin Yu, Ruicong Chen, Tinghui Yu, Nan Xu, Quan Quan, Zhenguang Pang, Shujian Zeng, Danna Zheng, Ting Lin and Weibo Ji. Thank you all for being the glimmers of light whenever I feel like walking alone in a long and dark tunnel.

Finally, my special thanks go to my family, in particular my aunt Fei Wang, my cousin Ye Bao, and most of all to my parents Zhi Wen and Weiwei Wang for understanding, love and support.

Contents

Abstract.....i

Declaration.....ii

Manuscripts submitted or accepted for publication, or in preparation iii

Acknowledgements iv

Contents vi

List of Figures..... xiii

List of Tables.....xvii

Chapter 1: Cross-language Activation in Bilinguals.....1

1.1 General Introduction.....1

1.2 Translation Recognition 5

1.2.1 Introduction to Interference Effects in Translation Recognition 5

1.2.2 Translation Recognition: Behavioural Findings 10

1.2.3 Translation Recognition: ERP Findings.....12

1.3 Masked Translation Priming 15

1.3.1 Introduction to Non-cognate Masked Translation Priming Paradigm 15

1.3.2 Non-cognate Masked Translation Priming: Behavioural Findings.....16

1.3.3 Non-cognate Masked Translation Priming: Behavioural Findings with Chinese-English Bilinguals.....23

1.3.4 Non-cognate Masked Translation Priming: ERP Findings 30

1.4 Translation Activation in Hidden Repetition Priming.....33

1.5 Translation Activation in Non-linguistic Tasks 36

1.5.2 Translation Activation in Non-linguistic Tasks: Eye-tracking Findings.....37

1.5.3 Translation Activation in Non-linguistic Tasks: ERP Findings 39

1.6 Summary of Empirical Studies Investigating Translation Processing.....40

1.7 Models on Bilingual Word Recognition and Lexical Representation 40

1.7.1 The Revised Hierarchical Model 41
1.7.2 The BIA+ and BIA Models 47
1.7.3 The Multilink Model 51
1.7.4 The Sense Model 53
1.7.5 The Distributed Representation Model 55
1.7.6 The DevLex-II Model 56
1.7.7 Summary 58
1.8 Language Processing in Chinese-English Bilinguals 58
1.8.1 A Brief Introduction to the Chinese Language 59
1.8.2 The Role of Characters in the Recognition of Two-character Chinese Words 67
1.8.3 The Role of Phonology in Chinese Reading 70
1.8.3.1 Phonological Activation in Chinese Reading 70
1.8.3.2 Segmental and Tonal Activation in Chinese Character Recognition 76
1.8.3.3 Summary 78
1.8.4 Cross-language Phonological Activation 78
1.9 Overview of the Present Thesis 81
Chapter 2: English to Chinese Translation Norms 84
2.1 Introduction 84
2.1.1 Word Frequency and Concreteness 86
2.1.2 Second Language Proficiency 87
2.1.3 Dominant and Subdominant Translations 88
2.1.4 Relationship between Source and Target Language 88
2.2 Experiment 1: Translation Norming Study 90
2.2.1 Method 90
2.2.1.1 Participants 90
2.2.1.2 Stimuli 91
2.2.1.3 Procedure 93
2.2.1.4 Measures of Word Frequency 93

2.2.1.5 Language Proficiency Test (LexTALE) 94
2.3 Results and Discussion 94
2.3.1 Translation Norms 94
2.3.2 English (L2) Word Frequency, Concreteness and Translation Ambiguity 97
2.3.3 Language Proficiency and Translation Ambiguity 100
2.3.4 Dominant and Subdominant Translations 102
2.3.5 Correlation between Source and Target Language Word Frequencies 103
2.4 Summary and Conclusions 106
Chapter 3: Phonological Activation of Chinese Translations during English Word Reading: Behavioural Investigations 108
3.1 Introduction 108
3.2 Experiment 2: Semantic Relatedness Judgment Task 111
3.2.1 Method 111
3.2.1.1 Participants 111
3.2.1.2 Materials and Design 111
3.2.1.3 Procedure 114
3.2.2 Results 115
3.2.3 Discussion 115
3.3 Experiment 3: Lexical Decision Task 116
3.3.1 Method 116
3.3.1.1 Participants 116
3.3.1.2 Materials and Design 117
3.3.1.3 Procedure 118
3.3.2 Results 120
3.3.3 Discussion 120
3.4 General Discussion 121
Chapter 4 Phonological Activation of Chinese Translation during English Word Reading: Electrophysiological Investigations 125
4.1 Introduction 125

4.2 Experiment 4: English Semantic Relatedness Judgment Task 131
4.2.1 Method 131
4.2.1.1 Participants 131
4.2.1.2 Materials and Design 132
4.2.1.3 Procedure 134
4.2.1.4 EEG Recordings 135
4.2.2 Data Analysis 135
4.2.2.1 Preprocessing 135
4.2.2.2 ERP Analysis 136
4.2.2.3 Statistical Analysis 136
4.2.3 Results 137
4.2.3.1 Behavioural Data 137
4.2.3.2 ERP Data (N400 time window) 137
4.2.4 Discussion 139
4.3 Experiment 5: Chinese Semantic Relatedness Judgment Task 143
4.3.1 Methods 143
4.3.1.1 Participants 143
4.3.1.2 Materials and Design 143
4.3.1.3 Procedure 145
4.3.1.4 EEG Recordings 145
4.3.2 Data Analysis 145
4.3.2.1 Preprocessing 145
4.3.2.2 ERP Analysis 145
4.3.2.4 Statistical Analysis 145
4.3.3 Results 147
4.3.3.1 Behavioural Data 147
4.3.3.2 ERP Data 148
4.3.4 Discussion 150
4.4 Time-frequency analysis of Experiment 4 and Experiment 5 154
4.4.1 Time-frequency Analysis 154
4.4.2 Statistical Analysis 154
4.4.3 Results 155

4.4.3.1 Chinese (Experiment 5) 155
4.4.3.2 English (Experiment 4) 158
4.4.4 Discussion 160
4.5 General Discussion 161
Chapter 5 Orthographical and Phonological Activation of Chinese Translation during English Word Reading: Behavioural Investigations 164
5.1 Introduction 164
5.2 Experiment 6: Lexical Decision Task 166
5.2.1 Method 166
5.2.1.1 Participants 166
5.2.1.2 Materials and Design 167
5.2.2 Results and discussion 169
5.3 Experiment 7: Lexical Decision Task 171
5.3.1 Method 171
5.3.1.1 Participants 171
5.3.1.2 Materials and Design 171
5.3.1.3 Procedure 172
5.3.2 Results and discussion 172
5.4 Experiment 8: Exact Replication of Zhang et al. (2011) 176
5.4.1 Method 176
5.4.1.1 Participants 176
5.4.1.2 Materials and Design 177
5.4.1.3 Procedure 179
5.4.2 Results 180
5.4.2.1 Task 1: Masked Hidden Repetition Piming with English Primes and Targets 181
5.4.2.2 Task 2: Masked Cross-language Priming with Chinese Primes and English Targets 181
5.4.2.3 Task 3: Chinese-English Translation Recognition 181
5.4.2.4 Evaluate Replication Attempts 182
5.5 General Discussion 183
Chapter 6: A Meta-analysis of Masked Translation Priming 191

6.1 Introduction 191

6.2 Method 192

6.2.1 Literature Search and Study Selection 192

6.2.2 Meta-analysis 193

6.2.2.1 Effect Sizes 193

6.2.2.2 Moderator Coding 194

6.2.2.3 Data Analyses 194

6.3 Results 195

6.3.1 L1-L2 Translation Priming 195

6.3.2 L2-L1 Translation Priming 197

6.4 General Discussion 198

Chapter 7 General Discussion and Conclusions 203

7.1 Introduction 203

7.2 Summary of Findings 203

7.3 Methodological and Theoretical Implications 206

7.3.1 The Importance of Translation Norms 206

7.3.2 Measures of Second Language Proficiency 207

7.3.3 The Issue of Effect Size and Power 209

7.3.4 The Relationship between Electrophysiological and Behavioral Evidence 214

7.3.5 The Time-frequency Analysis as an Insightful Approach to Investigate Visual Word Recognition 215

7.3.6 The Origin of Hidden Repetition Priming Effects 216

7.3.7 The Origin of Masked Translation Priming 219

7.4 Limitations and Future Directions 221

7.5 Conclusions 223

References 224

Appendix 260

Appendix 1. Database based on Experiment 1 in Chapter 2 260

Appendix 2. Stimuli used in Experiment 2 262

Appendix 3. Stimuli used in Experiment 3 265

Appendix 4. Stimuli used in Experiment 4 267

Appendix 5. Stimuli used in Experiment 5 271

Appendix 6. Stimuli used in Experiment 6 275

Appendix 7. Stimuli used in Experiment 7 277

Appendix 8. Stimuli used in Experiment 8 Task 1 279

Appendix 9. Stimuli used in Experiment 8 Task 2 and Task 3 281

Appendix 10. Results of mixed-effects modelling analysis of RTs for Experiment 6-8 283

Appendix 11. Results of additional analysis for Chapter 6 286

List of Figures

Figure 1. Illustration of ERP waveform 13
Figure 2. Illustration of visual events in masked translation priming paradigm 15
Figure 3. The Revised Hierarchical Model (figure adapted from Kroll & Stewart, 1994) 42
Figure 4. The Revised Hierarchical Model of Translation Ambiguity adapted to account for synonym translation-ambiguous pairs (figure adapted from Eddington & Tokowicz, 2013) 46
Figure 5. The Revised Hierarchical Model of Translation Ambiguity adapted to account for meaning translation-ambiguous pairs (figure adapted from Eddington & Tokowicz, 2013) 46
Figure 6. The BIA mode (figure adapted from van Heuven, Dijkstra, & Grainger, 1998) 48
Figure 7. The BIA+ Model (figure adapted from Dijkstra & van Heuven, 2002; van Heuven & Dijkstra, 2010) 49
Figure 8. The theoratical framework of the Multilink Model for word translation (figure adapted from Dijkstra & Rekké, 2010) 52
Figure 9. The Sense Model (figure adapted from Finkbeiner, Forster, Nicol, & Nakamura, 2004) 54
Figure 10. The Distributed Representation Model (figure adapted from Schoonbaert, Duyck, Brysbaert, & Hartsuiker, 2009) 55
Figure 11. The structure of the DevLex-II Model (figure adpated from Zhao & Li, 2010, 2013) 57
Figure 12. Overview of number of characters in Chinese words based on SUBTLEX-CH (Cai & Brysbaert, 2010) 60
Figure 13. Distribution of tones in Chinese charaters based on SUBTLEX-CH (Cai & Brysbaert, 2010) 64
Figure 14. Distribution of tones in 1,181 syllables based on SUBTLEX-CH (Cai & Brysbaert, 2010) 65
Figure 15. Histogram of English word frequencies (N = 1429) 92
Figure 16. Histogram of the translation accuracy 95
Figure 17. Histogram of the number of translations 95
Figure 18. The interaction between word frequency and concreteness ratings 99

Figure 19. Relationship between word frequency between English words and their dominant Chinese translations 103
Figure 20. Relationship between English and Chinese word frequency for translation-unambiguous pairs (N = 307) 104
Figure 21. Relationship of English and Chinese word frequency for cross-script cognates (N = 41) 105
Figure 22. Illustration of visual events in Experiment 2 114
Figure 23. Illustration of visual events in Experiment 3 119
Figure 24. Illustration of visual events in Experiment 4 134
Figure 25. Results of ERP analysis of critical trials in Experiment 4. ERPs time-locked to the onset of target words averaged across 9 electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz, all shaded in the schematic head) on the scalp with topography of ERP effect (-*Segment* minus +*Segment*) 138
Figure 26. Results of ERP analysis for semantic related (fillers) vs. unrelated trials (-*Segment* -*Tone*)) in Experiment 4. ERPs time-locked to the onset of target words averaged across 9 electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz, all shaded in the schematic head) on the scalp with topography of ERP effect (-*Segment* -*Tone* minus related fillers) 141
Figure 27. ERPs time-locked to the onset of target words average across all the unrelated conditions at the Cz electrode with the schematic head depicting the location of Cz. 146
Figure 28. Results of ERP analysis of critical trials in Experiment 5. ERPs time-locked to the onset of target words averaged across 3 electrodes (Pz, P1, P2, all shaded in the schematic head) on the scalp with topography of ERP effect (+*Segment* minus -*Segment*) 149
Figure 29. Results of ERP analysis semantic related (fillers) vs. unrelated trials (-*Segment* -*Tone*) in Experiment 5. ERPs time-locked to the onset of target words averaged across 9 electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz, all shaded in the schematic head) on the scalp with topography of ERP effect (-*Segment* -*Tone* minus related fillers) 151
Figure 30. Results of ERP analysis of -*Segment* -*Tone* vs. +*Segment* +*Tone* in Experiment 5. ERPs time-locked to the onset of target words averaged across 3 electrodes (FC1, FC2, FCz), all shaded in the schematic head) on the scalp with topography of ERP effect (-*Segment* -*Tone* minus +*Segment* +*Tone*) 152

Figure 31. Results of time-frequency analysis for the effect of Segment in Experiment 5. (A) Time-frequency representations at the Cz electrode for -Segment (left) and +Segment (right). (B) Time-frequency representations of the contrast -Segment minus +Segment at the Cz electrode with a mask showing the gamma frequency range (left); the corresponding topography of gamma frequency range from 531 to 676 ms with the white star marking the location of the Cz electrode (right). The colour scale indicates the relative power changes (0 = no difference).....156
Figure 32. Results of time-frequency analysis for the effect of Tone in Experiment 5. (A) Time-frequency representations at the Cz electrode for -Tone (left) and +Tone (right) conditions. (B) Time-frequency representations of the contrast -Tone minus +Tone at the Cz electrode with a mask showing the gamma frequency range (left); the corresponding topography of gamma frequency range from 465 to 800 ms with the white star marking the location of the Cz electrode (right). The colour scale indicates the relative power changes (0 = no difference).....157
Figure 33. Results of time-frequency analysis for the effect of Segment in Experiment 4. (A) Time-frequency representations of Cz electrode for *–Segment* (left) and *+Segment* (right) conditions. (B) Time-frequency representations of the contrast *Segmen*t minus *+Segment* in FCz electrode with a mask showing the gamma frequency range (30 - 60 Hz) from 578 to 632 ms (left); the corresponding topography of gamma frequency range from 578 to 632 ms with the white star marking the location of the Cz electrode (right). The colour scale indicates the relative power changes (0 = no difference).....159
Figure 34. The origins of the character repetition priming effect.....186
Figure 35. Mean response times for Task 2 in Experiment 8 with error bars indicating the standard errors.....188
Figure 36. Overall effect sizes for L1-L2 and L2-L1 non-cognate masked translation priming (with their 95% confidence intervals).....195
Figure 37. Observed effect sizes for L1-L2 non-cognate masked translation priming ordered by magnitude of the effect size and the overall effect sizes (with their 95% confidence intervals).....196
Figure 38. Observed effect sizes for L2-L1 masked non-cognate translation priming ordered by magnitude of the effect size and the overall effect sizes (with their 95% confidence intervals).....198

Figure 39. The relationship between power and number of participants estimated with the overall effect sizes for L1-L2 and L2-L1 masked non-cognate translation priming 200
Figure 40. Results of simulated experiments with a true priming effect of 4, 8 12 ms 212
Figure 41. Results of simulated experiments with a true priming effect of 16, 21 and 25 ms. 213
Figure 42. The origins of the phonological repetition priming effect in Chinese translations during English word reading. 219
Figure 43. The origins of masked translation priming effects with L1 primes and L2 targets 221

List of Tables

Table 1. Summary of behavioural experiments with a translation recognition task 8
Table 2. Description of studies using non-cognate masked L1-L2 translation priming paradigm with a lexical decision task 18
Table 3. Description of studies using non-cognate masked L2-L1 translation priming paradigm with a lexical decision task 21
Table 4. Description of studies published in peer-reviewed English journals with Chinese-English bilinguals using masked translation priming paradigm 26
Table 5. Six types of Chinese characters according to Shuowen Jiezi (Xu, 1985) 62
Table 6. Summary of language background data 91
Table 7. Results of fixed-effects regression analysis 99
Table 8. Results of mixed-effects logistic regression analysis 102
Table 9. Experimental design and stimulus characteristics of Experiments 2 (Means with SDs in brackets) 113
Table 10. Mean error rates and reaction times (with SD in brackets) of Experiment 2 115
Table 11. Experimental design and stimulus characteristics of Experiments 3 (Means with SDs in brackets) 118
Table 12. Mean error rates and reaction times (with SD in brackets) of Experiment 3 120
Table 13. Experimental design and stimulus examples in Experiment 4 133
Table 14. Mean error rates and reaction times (with SD in brackets) of Experiment 4 137
Table 15. Experimental design and stimulus examples of Experiment 5 144
Table 16. Mean error rates and reaction times (with SD in brackets) of Experiment 5 147
Table 17. Summary of participants' language background data from Experiments 6-8 167
Table 18. Experimental design and stimulus examples of Experiment 6 168
Table 19. Mean error rates and reaction times (with SD in brackets) of Experiment 6 170
Table 20. Experimental design and stimulus examples of Experiment 7 172

Table 21. Mean error rates and reaction times (with SD in brackets) in Experiment 7 174
Table 22. Experimental design and stimulus examples of Experiment 8 task 1 177
Table 23. Experimental design and stimulus examples of Experiment 8 task 2 & 3 178
Table 24. Mean error rates and reaction times (with SD in brackets) for tasks 2 and 3 of Experiment 8 180
Table 25. Mean error rates and reaction times (with SD in brackets) in Experiment 8 task 1 180
Table 26. Effect sizes and Meta BF of two replications against Zhang et al. (2011) 182
Table 27. Meta-regression analysis with one moderator for L1-L2 translation priming 197
Table 28. Meta-regression analysis with one moderator for L2-L1 translation priming 197

Chapter 1: Cross-language Activation in Bilinguals

1.1 General Introduction

It has been estimated that the majority of people in the world regularly use two languages (Grosjean, 2010; Grosjean & Li, 2013; Traxler, 2011), and those people are referred to as bilinguals. Bilinguals may differ in many aspects, for example, language proficiency, age of language acquisition (e.g., early bilinguals vs. late bilinguals) and language dominance (e.g., balanced bilinguals with similar proficiency in both languages vs. unbalanced bilinguals with higher proficiency in one language). Irrespective of the inevitable diversity, bilingualism has become common across the world (de Groot, 2011; Lemhöfer et al., 2008).

Due to the prevalence of bilingualism, researchers have come to realise that studying bilinguals can shed more light on language and cognition. The last few decades have witnessed a huge increase in the number of bilingual studies in the literature (Bialystok, 2016; van Heuven & Dijkstra, 2010) as well as an increase in the general interest in bilingualism. This trend is illustrated by the change of the word *bilingual* in terms of word frequency based on the Google Ngram Database (a corpus of digitized books published from 1800 to 2000, Michel et al., 2011). The frequency of the word *bilingual* was 0.17 per million in 1900, 0.72 per million in 1950, and 5 per million in 2000. Evidence of a growing interests in research on bilingualism comes from entering the search string ‘‘bilingualism’’ in the Web of Science, which revealed only 23 published papers in 1995, 76 in 2005, and 412 in 2015 (see Kroll & Bialystok, 2013 for a detailed survey from 1993 to 2013). The existing literature dedicated to bilingualism covers a wide range of topics, e.g., lexical representation and access of two languages, syntactic processing in second languages, conceptual transitions in bilinguals’ mind, cognitive changes as a consequence of bilingualism, as well as how second languages impact people’s decision-making (for reviews on these topics, see Athanasopoulos, 2015; Bialystok, Craik, & Luk, 2012; Hayakawa, Costa, Foucart, & Keysar, 2016; Jiang, 2015; Kotz, 2009; Thierry, 2016).

Amid this growing body of bilingual research, a fundamental topic is bilingual word recognition. The central question dominating this field is whether a presented word only activate word representations from the relevant language (*selective lexical access*), or word representations from both languages (*non-selective lexical access*). Although selective lexical access seems to be more consistent with people's intuition, this view has only received limited support from a handful of studies (e.g., Rodriguez-Fornells, Rotte, Heinze, Nosselt, & Munte, 2002). In contrast, the idea of non-selective lexical access (e.g., Dijkstra, 2005; Dijkstra & van Heuven, 2002; Grainger & Dijkstra, 1992; van Heuven et al., 1998) has been overwhelmingly supported in the literature (for reviews, see de Groot, 2011; Dijksta, 2005; Jiang, 2015; van Heuven & Dijkstra, 2010).

Convincing evidence for non-selective lexical access comes, for example, from studies using interlingual homographs (i.e., words with identical orthography but different meanings across languages, e.g., *coin* is an interlingual English-French homograph, which means "corner" in French). Beauvillain and Grainger (1987) presented English-French bilinguals with letter strings preceded by French primes in an English lexical decision task. The results showed that English target words were facilitated by related homographic primes (e.g., *coin* - *money*) compared to unrelated primes (e.g., *cuir* - *money*, *cuir* means "leather" in French). This indicated that the English meanings of the homographs were also activated even though bilinguals were informed that the primes were French words. The co-activation of two languages in processing interlingual homographs was further confirmed by more behavioural studies (Dijkstra, De Bruijn, Schriefers, & Ten Brinke, 2000; Dijkstra, Grainger, & van Heuven, 1999; Dijkstra, Van Jaarsveld, & Brinke, 1998; Lemhöfer & Dijkstra, 2004) as well as studies using event-related potentials (ERPs) and functional magnetic resonance imaging (fMRI) and involving different language pairs (e.g., Hoshino & Thierry, 2012; Kerkhofs, Dijkstra, Chwilla, & de Bruijn, 2006; Macizo, Bajo, & Martín, 2010; Paulmann, Elston-Güttler, Gunter, & Kotz, 2006; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008).

Compelling evidence for non-selective lexical access also comes from studies manipulating the number of interlingual orthographic neighbours of target words. Orthographic neighbours of a target can be generated by substituting one letter in the target with another letter (Coltheart, Davelaar, Jonasson, & Besner, 1977), for example, *pool*, *fool*, *cool*, *wool* and *toil* are neighours of the English target *tool*. In these example,

neighbours come from the same language as the target. When neighbours come from another language, they are referred to as interlingual orthographic neighbours (e.g., French word *clair* and English word *chair*). In their seminal study, van Heuven, Dijkstra, and Grainger (1998) found that Dutch-English bilinguals responded slower to English words with many Dutch neighbours than those with few Dutch neighbours. This result indicates that the orthographic information of the irrelevant language (i.e., Dutch) is automatically activated. Evidence of cross-language activation has also been obtained in ERP studies using interlingual orthographic neighbours (e.g., Grossi, Savill, Thomas, & Thierry, 2012; Midgley, Holcomb, van Heuven, & Grainger, 2008).

Further support for non-selective lexical access comes from a wealth of studies using cognates (i.e., words that have the same meanings in two language and that have large overlap in orthography or/and phonology across languages, e.g., *silence* is a cognate between French and English). The rationale of using cognates is straightforward because any difference observed when processing cognates and matched control words can only be explained by the co-activation of two languages. Robust facilitation effects have been found in various tasks, for example, relative to non-cognate translation equivalents, cognates are recognized faster (e.g., Dijkstra et al., 1999; Dijkstra et al., 1998; Lemhöfer & Dijkstra, 2004), named faster (e.g., Allen & Conklin, 2013; Costa, Caramazza, & Sebastian-Galles, 2000; Hoshino & Kroll, 2008) and translated faster (e.g., de Groot, 1992; de Groot, Dannenburg, & van hell, 1994; Sáchez-Casas, García-Albea, & Davis, 1992). In line with behavioural studies, the cognate facilitation effect has also been reported in studies using ERP (e.g., Midgley, Holcomb, & Grainger, 2011; Peeters, Dijkstra, & Grainger, 2013; Strijkers, Costa, & Thierry, 2009; Yudes, Macizo, & Bajo, 2010) as well as eye-tracking (e.g., Duyck, Assche, Drieghe, & Hartsuiker, 2007; Libben & Titone, 2009; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009).

In light of the converging evidence, it is generally agreed that bilingual lexical access is fundamentally non-selective (for reviews, see Dijksta, 2005; van Heuven & Dijkstra, 2010). Based on these findings and conclusions of non-selective lexical access, several models of bilingual word recognition have been put forward (e.g., Dijkstra & van Heuven, 2002; van Heuven et al., 1998 to be discussed further in Section 1.8). These models, which are used to guide new studies, further advance the investigation of

bilingual word recognition. It is noteworthy that the idea of non-selective lexical access not only is important for understanding bilingual word recognition, but also is the foundation for investigating the cognitive effects of bilingualism. This is because that co-activation of two languages in comprehension as well as production (Colomé, 2001; Hermans, Bongaerts, De Bot, & Schreuder, 1998; Jared & Kroll, 2001; Kroll, Bobb, & Wodniecka, 2006) requires a constant involvement of monitoring and control during bilingual language processing. This linguistic experience is unique to bilinguals, which is believed to enhance bilinguals' general cognitive abilities as a result (e.g., Coderre & van Heuven, 2014; Wu & Thierry, 2013; Zhou & Krott, 2016).

The above-mentioned evidence for non-selective lexical access comes from a group of special words (e.g., interlingual homographs and cognates). Obviously, these special words do not form the majority of the whole lexicon. In addition, studies supporting non-selective lexical access primarily tested bilinguals with same-script languages (e.g., English and French, in contrast to the different-script languages like English and Chinese). Therefore, evidence is needed to confirm the idea of non-selective lexical access for other types of words as well for cross-script bilinguals.

To investigate cross-language activation in both same-script and different-script bilinguals, non-cognate translation equivalents (henceforth translation equivalents) have been used extensively (e.g., Hoshino, Midgley, Holcomb, & Grainger, 2010; Marian & Spivey, 2003a; Midgley, Holcomb, & Grainger, 2009; Mishra & Singh, 2014; Moon & Jiang, 2012; Schoonbaert et al., 2011; Singh & Mishra, 2015; Spivey & Marian, 1999; Thierry & Wu, 2007; Wu & Thierry, 2010, 2011). For example, recent research has revealed that Chinese translations are automatically activated when Chinese-English bilinguals read English words (e.g., Thierry & Wu, 2007; Wu & Thierry, 2010; Zhang, van Heuven, & Conklin, 2011 to be presented further in Section 1.4). Because there are a large number of translation equivalents across languages, translation equivalents do not belong to a group of special words and any cross-language activation in terms of translation activation would strongly support the idea of non-selective lexical access. To understand the underlying mechanism of translation activation, research of the field so far mainly focuses on four critical aspects of translation activation, namely, whether the translation equivalent of a given word is activated before or after the word is processed up to the semantic level, whether translation

activation is task-independent or task-dependent, to what extent translation activation is automatic, and what type of lexical information is activated (phonology or/and orthography).

To provide an important step forward with respect to specifying the mechanisms of translation activation, this thesis explores L1 (first language) translation activation during L2 (second language) word reading in different-script bilinguals (Chinese-English) by using behavioural and electrophysiological methods. In this chapter, I present an overview of cross-language activation in bilingual word recognition focussing, in particular, on studies using translation equivalents. The chapter is organized as follows: Section 1.2 focuses on behavioural and ERP studies using the translation recognition task. Section 1.3 reviews behavioural and ERP studies involving the masked translation priming paradigm. Section 1.4 introduces recent investigations of L1 translation activation in a purely L2 context. Section 1.5 discusses translation activation in non-linguistic tasks using novel experimental paradigms. Following a summary of all the empirical studies in Section 1.6, Section 1.7 further examines cross-language activation by presenting theoretical models on bilingual processing to shed light on the experimental findings. After this comprehensive overview, Section 1.8 then presents a selective review of language processing in Chinese-English bilinguals, that were tested throughout this thesis to investigate translation activation. Finally, Section 1.9 provides the rationales for the studies in this thesis and presents an overview of the thesis.

1.2 Translation Recognition

1.2.1 Introduction to Interference Effects in Translation Recognition

Translation tasks, which explicitly require two active languages, have a long history in bilingual research. For example, the word translation task, in which bilinguals verbally translate the visually presented words from one language into the other language, has been used to investigate the organisation of bilingual mental lexicon (e.g., de Groot, 1992; de Groot et al., 1994; de Groot & Poot, 1997; Kroll & Curley, 1988; Kroll & Stewart, 1994; La Heij et al., 1990; La Heij, Hooglander, Kerling, & van der Velden,

1996; Potter, So, Von Eckardt, & Feldman, 1984; Sholl, Sankaranarayanan, & Kroll, 1995; Tokowicz & Kroll, 2007; van Hell & de Groot, 1998b). Using a word translation task, de Groot and colleges found that translation latencies were shorter in concrete words than in abstract words (e.g., de Groot, 1992; de Groot et al., 1994; van Hell & de Groot, 1998b). These findings have led to the conclusion that concrete translation equivalents share more semantic features than abstract translation equivalents in bilingual mental lexicon. Although word translation studies have produced fruitful findings, de Groot and Comijs (1995) pointed out that these studies often contain lots of omissions and errors in participants' verbal responses, and are labour-intensive and time-consuming to analyse. To overcome these issues, de Groot (1992) invented the translation recognition task. This paradigm is easier to analyse because it induces almost no omissions and fewer errors from participants. The paradigm has gained popularity among researchers ever since de Groot's introduction.

In a translation recognition task, bilinguals are presented with a word pair in which one word of the word pair is presented first (prime) followed by the other word of the word pair (target). The prime and target are presented in different languages and participants have to judge whether the word pair are translation equivalent or not. Thus, the translation recognition task requires bilinguals to use both of their languages. Early translation recognition studies focused on translation equivalents pairs, which required "Yes" responses. For example, de Groot (1992) examined the impact of word frequency and imageability on translation recognition and revealed that the responses were faster in high-frequency words than low-frequency words and faster in high- imageability words than low- imageability words.

However, later studies often used the translation recognition task to investigate interference effects in non-translation-equivalent pairs. For example, for Spanish-English pairs in a translation recognition task, a "No" response is required for pairs like *cara - fact* or *care-head*, because the correct English translation of the Spanish word *cara* is *face*. However, the correct translation *face* is similar to *fact* at the form level, so an interference effect at the form level (e.g., longer reaction time compared to control pairs with no form overlap) is expected for *fact* if participants translate the Spanish word into English. This effect is known as *through-translation form interference*

effects.[1] Similarly, because *cara* is semantically related to *head*, *semantic interference effects* are expected if the meaning of *cara* is activated. Studies in the literature that used through-translation form interference distractors (e.g., *fact*) and semantic interference distractors (e.g., *head*) have focused on the presence or absence of interference to investigate whether translation equivalents and/or semantic meanings are activated during translation recognition (see Table 1 for an overview of all the existing studies up to August 2016). In the next section I will summarize the findings of these studies.

[1] The word *form* is not limited to orthographical form, but refers to lexical form which contained orthographic and phonological information.

Table 1. Summary of behavioural experiments with a translation recognition task

Study	**L1**	**L2**	**direction**	**Prime (ms)**	**ISI (ms)**	**group division criteria**	**language proficiency assessment**	**NoP**	**group**	**Interference**	
										through-translation	**semantic**
Talamas, Kroll, and Dufour (1999)	English	Spanish	both	400	100	acc of the task (threshold: 85%)	self-rated	22	less fluent	114#	20#
								17	more fluent	43#	124#
Ferré, Sánchez-Casas, and Guasch (2006)	Spanish	Catalan	Catalan (L2)-Spanish (L1)	500	0	Age of acquisition	self-rated	28	late non-proficient	64*	very close: 10; close: 23
	Spanish	Catalan						28	late proficient	41*	very close: 43; close: 15
	balanced Spanish-Catalan							49	early	37**	very close:47 **; close:7
Sunderman and Kroll (2006)	English	Spanish	L2-L1	400	100	length of L2 learning	self-rated; reading span; picture naming	63	less proficient	44#	88#
								44	more proficient	6#	81#
Guasch, Sánchez-Casas, Ferré, and García-Albea (2008)	Spanish	Catalan	L2-L1	750	0	length of L2 learning	self-rated	22	beginning	113*	very close:-79 ; close: 9
								16	intermediate	57**	very close:81**; close:30
								75	proficient	77**	very close:40**; close:12
Linck, Kroll, and Sunderman (2009)	English	Spanish	L2-L1	400	100	learning environment	self-rated/acc and RT of translation recognition	25	immersed	25#	About 120#
								20	class-room	65#	About 65#
Qasem and Foote (2010)	Arabic	English	L2-L1	400	100	class; self-rated proficiency; length of stay in US	self-rated	30	low proficiency	orth:48#; morph:216#	136#
								30	high proficiency	ortho:109#; morph:177#	137#

Table 1. (*Continued*)

Study	**L1**	**L2**	**direction**	**Prime (ms)**	**ISI (ms)**	**language proficiency assessment**	**NOP**	**Interference**	
								through-translation	**semantic**
Moldovan, Sánchez-Casas, Demestre, and Ferré (2012)	Catalan /Spanish	Catalan /Spanish	Catalan-Spanish	250	0	self-rated	85 (48 Catalan dominant & 37 Spanish dominant)	67*	very close: 54*; close:20
			Spanish-Catalan				71 (50 Catalan dominant & 21 Spanish dominant)	71*	very close: 51*; close:16
Sunderman and Priya (2012)	Hindi	English	L1-L2	400	100	self-rated	34	125[#]	/
			L2-L1	400	100			47[#]	/
Guo, Misra, Tam, and Kroll (2012)	Chinese	English	L1-L2	500	250	self-rated/LDT	20	About 76***	About 75***
				250	50		18	About 70***	About 74***
Moldovan, Demestre, Ferré, and Sánchez-Casas (2016)	balanced Spanish-Catalan		Catalan-Spanish	250	0	self-rated	25	67*	very close: 86*; close:36*

Note. **NoP**: number of participants; orth: orthographic similar; morph: morphologic similar; **ISI**: Inter-Stimulus Interval, i.e., interval between the offset of prime and the onset of target; [#] indicating studies did not conduct planned-contrast tests to compare whether each type of interference is statistically significant;
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1.2.2 Translation Recognition: Behavioural Findings

Talamas et al. (1999) used the translation recognition paradigm for the first time to investigate through-translation form and semantic interference effects. In their study, English-Spanish were tested in an English-Spanish and Spanish-English translation recognition task. Importantly, in order to investigate whether interference effects were modulated by the bilinguals' proficiency, bilinguals were divided into less fluent and more fluent groups (terms used by Talamas et al.) based on the accuracy in their experiment (threshold: > 85% correct). Talamas et al. (1999) found that the less fluent group showed a larger through-translation form interference effect than a semantic interference effect (114 ms vs. 20 ms), whereas in the more fluent group the semantic interference effect was larger than the through-translation form interference effect (124 ms vs. 43 ms). According to Talamas et al. (1999), the results suggest that lexical connections between translation equivalents are stronger in less fluent bilinguals (who were less proficient) than more fluent bilinguals (who were more proficient). In contrast, the access of semantic meanings is faster in more fluent bilinguals than less fluent bilinguals.

Following Talamas's et al. (1999) seminal study, numerous studies examined through-translation form and semantic interference effects in groups of bilinguals with different linguistic profiles. The existing findings in this line of research are hard to evaluate for two major reasons. First of all, different criteria have been used to divide bilinguals into separate groups. Some studies used length of L2 learning as a criteria (Guasch et al., 2008; Sunderman & Kroll, 2006), whereas others divided the bilinguals into groups based on their learning environment (Linck et al., 2009) or age of acquisition (Ferré et al., 2006). More confusingly, as shown in Table 1, different terms have been used to refer to less and more proficient bilinguals (e.g., less/more fluent, beginning/intermediate/proficient bilinguals). Importantly, the existing studies predominantly used self-assessed proficiency ratings, which are a less reliable estimate of the bilinguals' L2 proficiency (Brysbaert, 2013; Khare, Verma, Kar, Srinivasan, & Brysbaert, 2013; Lemhöfer & Broersma, 2012), and no standard objective language proficiency measures were used. Therefore, it is problematic to compare the bilinguals' proficiency across the different studies to assess the findings. Secondly, different statistical analyses have been used in the literature. Because different sets of control

words were created for through-translation form distractors and semantic detractors to calculate the interference effects, some researchers conducted separate t-tests to directly test whether through-translation form interference effects and/or semantic interference effects were statistically significant (Ferré et al., 2006; Guasch et al., 2008; Guo et al., 2012; Moldovan et al., 2016; Moldovan et al., 2012). Lack of such direct comparison in other studies makes it difficult to interpret the existing findings (Linck et al., 2009; Qasem & Foote, 2010; Sunderman & Kroll, 2006; Sunderman & Priya, 2012; Talamas et al., 1999).

Guo et al. (2012) claimed that through-translation form interferences effects were generally only observed in less proficient bilinguals. However, in the five studies which separately tested through-translation form interference effects and/or semantic interference effects (Ferré et al., 2006; Guasch et al., 2008; Guo et al., 2012; Moldovan et al., 2016; Moldovan et al., 2012), significant through-translation form interference effects have consistently been reported (see Table 1), and significant semantic interference effects were found more often in more proficient bilinguals. In the other studies (Linck et al., 2009; Qasem & Foote, 2010; Sunderman & Kroll, 2006; Sunderman & Priya, 2012; Talamas et al., 1999), the average through-translation form interference effect was 85 ms (range from 6 to 216 ms) and the average semantic interference effect was 96 ms (range from 20 to 137 ms). Therefore, it seems that both types of interference effects do occur. Thus, the translation equivalents and semantic information of the primes are activated when bilinguals perform the translation recognition task.

To explore the crucial role of semantic activation in the translation recognition task, other studies further manipulated the semantic distance in the semantic distractors by including more related semantic distractors and less related distractors (Ferré et al., 2006; Guasch et al., 2008; Moldovan et al., 2016; Moldovan et al., 2012). For example, for the Catalan word *ruc*, which means “donkey”, the Spanish word for “horse” *caballo* is considered as more related to the correct translation *donkey* than the Spanish word for “bear” *oso*. These two types of distractors are referred to in the literature as a very close semantic distractor and a close semantic distractor respectively. The general pattern of highly proficient bilinguals that emerges from the literature (see Table 1) is

that the very close semantic distractors cause reliable interference effects, whereas close semantic distractors did not always induce significant interference effects.

Some studies have focused more on through-translation form interference effects. For example, Sunderman and Priya (2012) found with Hindi-English bilinguals asymmetric through-translation form interference effects, with stronger interference in the L1-L2 direction than the L2-L1 direction. This is so far the only study that statistically compared the magnitude of interference effects in two translation directions in unbalanced bilinguals. Therefore, more systematic investigations of the impact of translation directions on through-translation form interference effects are needed for the field. The role of orthography and morphology has also been explored in through-translation form distractors in a study by Qasem and Foote (2010), which revealed that Arabic-English bilinguals showed a larger interference effects in morphological distractors than in orthographic distractors.

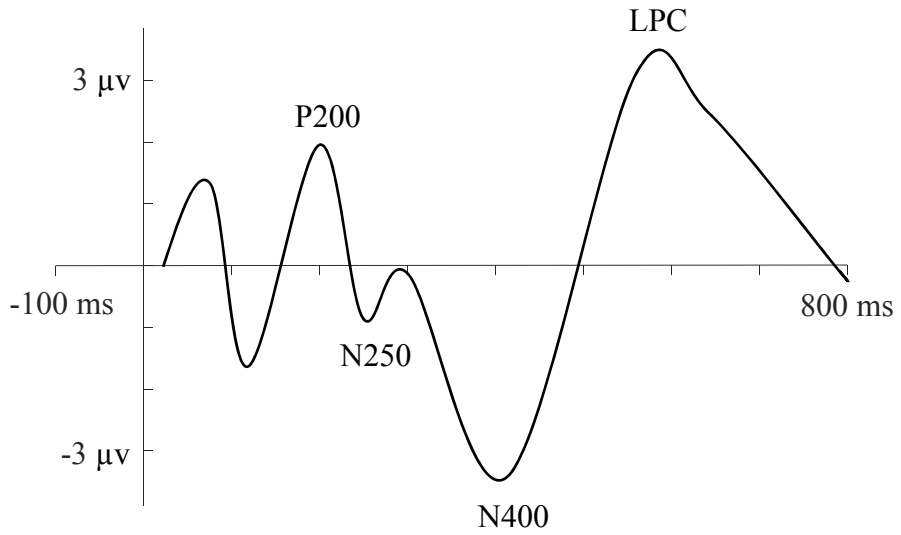
In terms of the origin of the through-translation form interference, it is important to point out that the majority of the existing studies assume that bilinguals translate the primes (the first word of the pairs) into the other language in the translation recognition task (e.g., for the Spanish-English pair *cara* - *fact*, the Spanish word *cara* is translated into English as *face*, the English word *fact* would therefore cause interference). Interestingly, Ferré et al. (2006) is the only study in which targets (the second words in the pairs) must be translated to obtain through-translation form interference effects. For example, in the Catalan-Spanish pair *ruc* - *riego*, the Catalan translation of *riego* is *rec,* similar to the prime *ruc*. Significant through-translation form interference was found for these pairs by Ferré et al. (2006), which suggests during the translation recognition task not only primes are translated but also targets.

1.2.3 Translation Recognition: ERP Findings

The evidence of the behavioural studies presented in section 1.2.2 suggests that translation equivalents and semantic information of the primes are both activated. A critical question is whether translation equivalents are activated *before* or *after* the semantic information is activated (e.g., when presented with Spanish word *cara*, Spanish-English bilinguals first access its English translation *face* and then access the

semantic information of the word *cara*, or vice versa). In other words, the time course of translation activation and semantic retrieval is unclear. To investigate the time course of translation and semantic activation, some researchers have conducted electrophysiological studies. Guo et al. (2012) conducted the first electrophysiological translation recognition study to investigate through-translation form interference effects and semantic interference effects. Two groups of unbalanced proficient Chinese-English bilinguals were tested in an English-Chinese (L2-L1) translation recognition task using a 700-ms SOA (i.e., Stimuli Onset Asynchrony, the interval between the onset of prime and the onset of target) in Experiment 1 and a 300-ms SOA in Experiment 2. Although through-translation form distractors and semantic distractors slowed down the bilinguals' responses, the electrophysiological data revealed different ERP patterns for the two types of interference effects. In Experiment 1 and 2, semantic interference reduced the amplitude of the N400 component (i.e., a negative-going wave peaking at around 400 ms after target presentation, see Figure 1), which is associated with semantic processing (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). Interestingly, only in Experiment 1 through-translation form interference elicited a larger amplitude in P200 component (i.e., a positive-going wave peaking at around 200 ms after target presentation, see Figure 1), which is associated with lexical form overlaps (Liu, Perfetti, & Hart, 2003; Misra & Holcomb, 2003). Furthermore, in both experiments, both types of interference elicited the LPC effect (i.e., Late Positive Component, peaking from around 500 to 700 ms, see Figure 1), which has been assumed to reflect the conflict resolution at late

Figure 1. Illustration of ERP waveform



decision stages (Kolk & Chwilla, 2007; Misra & Holcomb, 2003). Based on their electrophysiological data, Guo et al. (2012) suggested that the presence of N400, P200 and LPC effects indicated the activation of semantics and translation equivalents in Experiment 1, which used a long SOA. However, when a shorter SOA was used in Experiment 2, the presence of the N400 and LPC effect and the absence of the P200 effect suggests that semantic information of L2 primes is activated earlier than their L1 translation equivalents. Taken together, Guo et al. (2012) argued that L1 translation equivalents are activated after the semantics of L2 words are activated. However, this only happens when the prime is presented for sufficient amount of time, and with a short SOA there is insufficient time for L1 translation equivalents activation during the early processing stage of L2 words.

In another ERP study that used a short SOA (250 ms), balanced Catalan-Spanish bilinguals were asked to perform a Catalan-Spanish translation recognition task (Moldovan et al., 2016). In addition to compare the through-translation form interference and semantic interference effects, Moldovan et al. (2016) also compared semantic interference in more related semantic distractors and less related semantic distractors. The results revealed that both types of semantic distractors elicited the reduced N400, with larger effects in more related semantic distractors. In contrast, through-translation form distractors only modulated the LPC amplitude. According to Moldovan et al. (2016), these findings replicated Experiment 2 of Guo et al. (2012). Consequently, Moldovan et al. (2016) also conclude that translation equivalents are activated after the semantic information of words is activated.

As far as I am aware, the above two studies are only two ERP studies that investigated interference effects during the translation recognition task. Despite the limited number of electrophysiological studies relative to the number of behavioural studies, the electrophysiological data have suggested that the semantic information of words is activated before the translation equivalents in the other language are activated. An important implication drawn from the conclusion of the translation recognition studies is that the translation equivalent of a given word is activated through the shared semantic representation rather than a direct lexical link between pairs of translation equivalents. Obviously, the translation recognition task explicitly requires bilinguals to use both of their languages, so the mechanism of translation activation needs to be

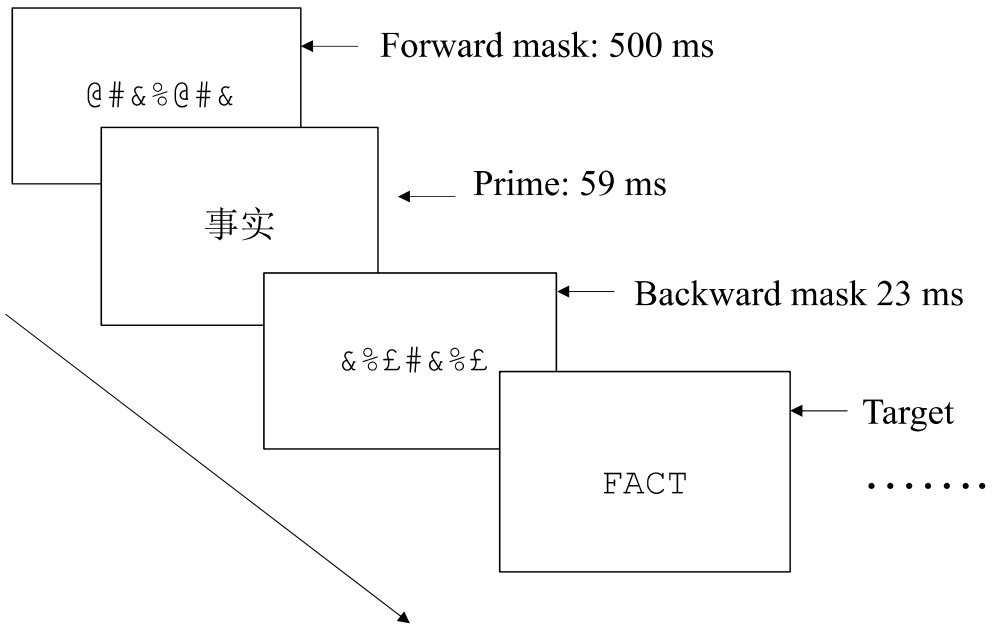
further investigated in a situation when bilinguals consciously attend to one of their languages, which is the focus on the next section.

1.3 Masked Translation Priming

1.3.1 Introduction to Non-cognate Masked Translation Priming Paradigm

This section focuses on the masked translation priming paradigm, which has been widely used in the last 25 years to investigate word processing in bilinguals. In contrast to the translation recognition task, the masked translation priming paradigm only requires bilinguals to focus on the language of the target word. In this paradigm (see Figure 2), a prime word is briefly presented (e.g., 50 ms) after a forward pattern mask (e.g., #####). The prime word is then replaced by a target word in another language, which is either the translation equivalent of the prime word or an unrelated word. An advantage of this paradigm is that the participants are generally unware of the masked primes (Forster & Davis, 1984; Kinoshita & Lupker, 2003) and any observed translation priming effect (e.g., shorter reaction times to target words primed by translation

Figure 2. Illustration of visual events in masked translation priming paradigm



equivalents) is therefore not influenced by strategic processing. Researchers predominantly have used non-cognate translation equivalents in this paradigm because non-cognate translation equivalents are generally orthographically and phonologically dissimilar, and any translation priming effect can only be attributed to the automatic lexico-semantic interactions between the two languages. Therefore, the following review focuses only on studies using non-cognate translation equivalents.

1.3.2 Non-cognate Masked Translation Priming: Behavioural Findings

Since the first translation priming effects were observed in de Groot and Nas's seminal study in 1991, other studies have used non-cognate masked translation priming in combination with various linguistic tasks. The existing behavioural evidence for translation priming mainly comes from studies using the lexical decision task, which requires participants to decide whether the presented letter string is a word or not. Those lexical decision studies have revealed a translation priming asymmetry in unbalanced bilinguals (see Table 2 and Table 3 for an overview of L1-L2 and L2-L1 translation priming respectively). The first evidence of translation priming asymmetry was reported by Gollan et al. (1997). Gollan et al. (1997) tested two groups of Hebrew-English and English-Hebrew bilinguals in either L1-L2 direction (L1 primes and L2 targets) or L2-L1 direction (L2 primes and L1 targets). For both Hebrew-English and English-Hebrew bilinguals, significant L1-L2 translation priming effects were observed, whereas no L2-L1 translation priming effects were found. Consistent with Gollan et al. (1997), early studies have consistently reported significant L1-L2 translation priming effects in contrast to no L2-L1 translation priming effects (e.g., Finkbeiner et al., 2004; Jiang, 1999; Jiang & Forster, 2001; Kim & Davis, 2003). It is important to note that this asymmetric pattern is only found in unbalanced bilinguals performing a lexical decision task (findings from other tasks to be presented in the next section). Studies testing balanced bilinguals using the same task have observed similar magnitudes of L1-L2 and L2-L1 translation priming (Duñabeitia, Perea, & Carreiras, 2010b; Perea, Dunabeitia, & Carreiras, 2008).

However, recent lexical decision studies with unbalanced bilinguals did observe significant translation priming effects in both directions across different language pairs (e.g., Greek-English: Dimitropoulou et al., 2011b; Dutch-English: Duyck & Warlop,

2009; Japanses-English: Nakayama et al., 2016). For example, Schoonbaert's et al. (2009), found significant L1-L2 and L2-L1 translation priming with short and long SOAs (100 ms & 250 ms) in unbalanced Dutch-English bilinguals. Although a translation priming asymmetry was also observed because priming effects for L2-L1 were smaller than those for L1-L2. Inconsistent findings in the field have been discussed in several recent narrative reviews (Altarriba & Basnight-Brown, 2007; Dimitropoulou et al., 2011b; Nakayama et al., 2016; Xia & Andrews, 2015). As summarized in Dimitropoulou et al. (2011b), only 8 out of 21 experiments reported significant L2-L1 translation priming (mean priming effect: 9 ms, range from -6 to 26 ms). Although robust L1-L2 translation priming effects were reported in all experiments (cf. Davis et al., 2010), the priming effects varied greatly from 16 to 100 ms (mean: 41 ms). Similar to Dimitropoulou et al. (2011b), as summarized Table 2 and Table 3, the average L1-L2 translation priming effect was 44 ms (range: 12 -127 ms), and the average L2-L1 translation priming effect was 11 ms (range: -4 - 48 ms).

Recent narrative reviews have tried to identify possible factors modulating the translation priming effects reported in the literature, such as factors related to statistical power (i.e., number of participants,), the experiment procedures (i.e., prime duration, ISI, SOA), the stimuli (i.e., number of experimental trials, the languages involved are same or different script) and processing speed (i.e., response speed from participants). Because the existing studies differ dramatically in terms of all these factors (see Table 2 and Table 3), no studies have investigated all the potential moderators at the same time. Furthermore, there are so far no meta-analytic reviews conducted in the literature that quantitatively assessed the effect sizes of L1-L2 and L2-L1 translation priming and the impacts of potential experimental moderators of translation priming effects. Therefore, it is unclear whether L2-L1 translation priming effects are reliable or not, and whether inconsistent findings for translation priming effects in both directions are caused by any potential factors.

Table 2. Description of studies using non-cognate masked L1-L2 translation priming paradigm with a lexical decision task

Study	**Experiment**	**L1**	**L2**	**Prime (ms)**	**ISI (ms)**	**SOA (ms)**	**NoP**	**NoI**	**RT_c (ms)**	**PE (ms)**
Williams (1994)	Exp2b	German/Italian /French	English	40	10	50	18	9	723	33*
Gollan et al. (1997)	Exp1	Hebrew	English	50	0	50	40	16	712	36***
	Exp2	English	Hebrew	50	0	50	30	16	979	52*
Jiang (1999)	Exp1	Chinese	English	50	0	50	52	16	755	45*
	Exp2	Chinese	English	50	0	50	44	16	814	68*
Jiang and Forster (2001)	Exp4	Chinese	English	50	0	50	22	16	717	41*
Kim and Davis (2003)	Exp1	Korean	English	50	0	50	22	12	674	40*
Basnight-Brown and Altarriba (2007)	Exp2	Spanish	English	100	0	100	48	16	636	33***
Voga and Grainger (2007)	Exp2	Greek	French	50	0	50	30	15	684	23*
	Exp3	Greek	French	50	0	50	30	15	691	22*
Duyck and Warlop (2009)	Exp1	Dutch	French	56	56	112	24	22	687	48***
Schoonbaert et al. (2009)	Exp1	Dutch	English	50	50	100	30	28	572	19***
	Exp1	Dutch	English	50	200	250	30	28	661	100***

Table 2. (*Continued*)

Study	**Experiment**	**L1**	**L2**	**Prime (ms)**	**ISI (ms)**	**SOA (ms)**	**NoP**	**NoI**	**RT_c (ms)**	**PE (ms)**
Dimitropoulou, Duñabeitia, and Carreiras (2011a)	Exp1a	Greek	Spanish	50	0	50	40	14	777	29**
	Exp3b	Greek	Spanish	50	50	100	44	14	818	31***
Dimitropoulou, Duñabeitia, and Carreiras (2011b)	Exp1a	Greek	English	50	0	50	36	58	723	31***
	Exp2a	Greek	English	50	0	50	36	58	739	28***
	Exp3a	Greek	English	50	0	50	36	58	695	28***
Schoonbaert, Holcomb, Grainger, and Hartsuiker (2011)	Exp1a	English	French	100	20	120	20	80	653	70***
Witzel and Forster (2012)	Exp1b	Chinese	English	50	0	50	32	16	678	39***
Luo et al. (2013)	Exp2a	Chinese	English	50	150	200	27	12	745	77**
	Exp3a	Chinese	English	50	150	200	24	12	803	75**
	Exp4	Chinese	English	50	150	200	27	12	598	30***
Wang (2013)	Exp1	English	Chinese	50	0	50	20	40	604	39**
Chen et al. (2014b)	Exp2	Chinese	English	50	150	200	40	14	678	28***
Aparicio and Lavaur (2015)	Exp1	French	English	50	17	67	24	20	574	14**

Table 2. (*Continued*)

Study	**Experiment**	**L1**	**L2**	**Prime (ms)**	**ISI (ms)**	**SOA (ms)**	**NoP**	**NoI**	**RT_c (ms)**	**PE (ms)**
Lupker, Perea, and Nakayama (2015)	Exp1	Spanish	English	50	0	50	20	50	620	12*
	Exp3	Japanese	English	60	0	60	36	30	668	45***
Wang and Forster (2015)	Exp2	Chinese	English	50	0	50	20	15	625	32*
Xia and Andrews (2015)	Exp1b	Chinese	English	50	150	200	34	16	854	127***
	Exp2b	Chinese	English	50	150	200	30	16	647	76***

Note. **ISI**: Inter-Stimulus Interval, i.e., interval between the offset of prime and the onset of target); **SOA**: Stimuli Onset Asynchrony, i.e., the interval between the onset of prime and the onset of target; **NoP**: number of participants; **NoI**: number of items per cell; **RT-c**: reaction times in the control conditions; **PE**: priming effects; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Description of studies using non-cognate masked L2-L1 translation priming paradigm with a lexical decision task

Study	**Experiment**	**L1**	**L2**	**Prime (ms)**	**ISI (ms)**	**SOA (ms)**	**NoP**	**NoI**	**RT_c (ms)**	**PE (ms)**
Gollan, Forster, and Frost (1997)	Exp3	Hebrew	English	50	0	50	40	16	574	9
Grainger and Frenck-Mestre (1998)	Exp1 (43 ms)	English	French	43	14	57	12	30	573	10
Jiang (1999)	Exp1	Chinese	English	50	0	50	52	16	581	13*
	Exp2	Chinese	English	50	50	50	44	16	635	3
	Exp3	Chinese	English	50	50	100	16	16	583	4
	Exp4	Chinese	English	50	200	250	18	16	541	7
Jiang and Forster (2001)	Exp1	Chinese	English	50	200	250	24	16	580	8
	Exp3, SOA 250 ms	Chinese	English	50	200	250	18	16	552	9
	Exp3, SOA 50 ms	Chinese	English	50	0	50	18	16	522	4
Finkbeiner et al. (2004)	Exp2	Japanese	English	50	150	200	18	26	525	-4

Table 3. (*Continued*)

Study	**Experiment**	**L1**	**L2**	**Prime (ms)**	**ISI (ms)**	**SOA (ms)**	**NoP**	**NoI**	**RT_c (ms)**	**PE (ms)**
Chen, Zhou, Gao, and Dunlap (2014b)	Exp1	Chinese	English	50	150	200	44	14	555	4
Sabourin, Brien, and Burkholder (2014)	Exp1 group2	English	French	50	0	50	25	15.5	<550,>525	14.2*
	Exp1 group3	English	French	50	0	50	16	15.5	<550,>525	0.8
	Exp1 group4	English	French	50	0	50	30	15.5	about 550	-0.4
Wang and Forster (2015)	Exp2	Chinese	English	80	0	80	20	15	530	5
Xia and Andrews (2015)	Exp1b	Chinese	English	50	150	200	34	16	620	12
	Exp2b	Chinese	English	50	150	200	30	16	591	14
Nakayama, Ida, and Lupker (2016)	Exp1	Japanese	English	50	0	50	36	30	518	10*
	Exp2	Japanese	English	60	0	60	34	30	533	22**
	Exp3	Japanese	English	60	0	60	34	30	509	3

Note. **ISI**: Inter-Stimulus Interval, i.e., interval between the offset of prime and the onset of target); **SOA**: Stimuli Onset Asynchrony, i.e., the interval between the onset of prime and the onset of target; **NoP**: number of participants; **NoI**: number of items per cell; **RT-c**: reaction times in the control conditions; **PE**: priming effects * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1.3.3 Non-cognate Masked Translation Priming: Behavioural Findings with Chinese-English Bilinguals

This section focuses specifically on non-cognate masked translation priming studies in Chinese-English bilinguals (see Table 4 for an overview). In the literature, a number of studies involving Chinese-English bilinguals have used other tasks in addition to the lexical decision task. For example, semantic categorization task (Chen et al., 2014a; Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015) or an episodic recognition task (Jiang & Forster, 2001; Witzel & Forster, 2012). Discussing the studies using different tasks in bilinguals with the same language pairs would provide a useful review on the impact of task on translation priming.

The results of lexical decision studies with Chinese-English bilinguals are similar to the findings discussed in the previous section. In unbalanced bilinguals, significant L1-L2 translation priming effects have consistently been reported (see Table 4: Chen et al., 2014b; Jiang, 1999; Luo et al., 2013; Wang & Forster, 2015; Xia & Andrews, 2015). In contrast, mixed findings have been reported in the L2-L1 direction. Although one experiment in Jiang (1999) reported a significant L2-L1 translation priming effect, Jiang (1999) failed to replicate this in later experiments. Using L2 words with many senses (i.e., semantic meanings) and L1 translations with fewer senses, an experiment of Luo et al. (2013) did report a significant L2-L1 translation priming. However, as can be seen in Table 4, all the other studies with unbalanced bilinguals failed to find significant L2-L1 translation priming effects (Chen et al., 2014b; Jiang, 1999; Jiang & Forster, 2001; Luo et al., 2013; Wang & Forster, 2015; Xia & Andrews, 2015). It is important to note that no L2-L1 translation priming effects were found despite the use of longer prime durations (e.g., 80 ms in Wang & Forster, 2015) or by including a backward mask up to 150 ms to increase the processing time of L2 primes (e.g., Chen et al., 2014b; Jiang, 1999; Xia & Andrews, 2015). In addition, a translation asymmetry was also found with balanced Chinese-English bilinguals (Wang, 2013).

To explain the translation priming asymmetry found in lexical decision experiments, two studies combined an episodic recognition task with the masked priming technique to test a episodic L2 hypothesis (Jiang & Forster, 2001; Witzel & Forster, 2012).

According to this hypothesis, L1 words are represented in lexical memory whereas L2 words are represented in episodic memory. This hypothesis predicts that if L2 words are activated in episodic memory, significant L2-L1 translation priming effects would occur. In order to move L2 words to episodic memory, bilinguals were first presented with a list of L2 words in a study phase. In the following recognition phase, bilinguals were asked to judge whether the presented L1 words, which were primed by masked L2 words, are old (presented in the study phase) or new (not presented in the study phase). Significant L2-L1 translation priming effects were found in old items, however, only when a 50 ms blank and a 150 ms backward mask were used (Jiang & Forster, 2001; Witzel & Forster, 2012). It is unclear why the effect disappears when the prime was immediately followed by the target (Experiment 3a in Jiang, 1999). More importantly, as Nakayama et al. (2016) pointed out, L2-L1 priming occurred "because episodic representations of L2 primes were being ACTIVATED, not because L2 word are episodically REPRESENTED" (p. 16).

There are three studies in the literature that have used the semantic categorization task with Chinese-English bilinguals (Chen et al., 2014a; Wang & Forster, 2010; Xia & Andrews, 2015). For example, in a colour categorization task, participants have to judge whether the presented words are colour words. They are instructed to press the "Yes" key in response to *black* (known as *exemplar*) and to press the "No" key in response to *apple* (known as a *non-exemplar*). These studies found that L2 words did prime L1 translations in exemplars. However, semantic categorization studies have potential methodological limitations. As noted by Nakayama et al. (2016), bilinguals with smaller L2 vocabulary size might generate a set of exemplars before the words are presented in the experiment. Therefore, it is likely that significant L2-L1 translation priming effects in exemplars are driven by episodic representations of L2 primes.

To conclude, the non-cognate masked translation priming paradigm has been widely used in studies with Chinese-English bilinguals. Despite that the interpretation of the existing findings are still debated (for a recent review, see Nakayama et al., 2016), it is clear that L1-L2 translation priming effects are robust in Chinese-English bilinguals. To illustrate, the average L1-L2 translation priming effect from 10 lexical decision experiments with Chinese-English bilinguals (see Table 4: Chen et al., 2014a; Chen et al., 2014b; Jiang, 1999; Luo et al., 2013; Wang & Forster, 2015; Xia & Andrews, 2015)

is 58 ms (range: 18 -127 ms). This magnitude is even slightly larger than the 41-ms (range: 12 - 100 ms) L1-L2 translation priming effect observed in 8 lexical decision experiments with same-script bilinguals (see Table 2: Alvarez, Holcomb, & J, 2003; Basnight-Brown & Altarriba, 2007; Duyck & Warlop, 2009; Lupker et al., 2015; Schoonbaert et al., 2009; Schoonbaert et al., 2011; Williams, 1994). Taken together, the results so far suggests that L1-L2 translation priming effects are independent of the scripts involved.

Table 4. Description of studies published in peer-reviewed English journals with Chinese-English bilinguals using masked translation priming paradigm

Study	Task	Place of Testing	NoP	L1	AoFEE	YoFEE	EPro	NoI	Prime (ms)	Blank (ms)	Post-mask (ms)	Chi-Eng (ms)	Eng-Chi (ms)
Xia and Andrews (2015)													
Exp 1A	SCT	Sydney, Australia	37	Chi	10	12	7 point self-rate: 4.9	16	50	/	150	E; 109** NE: 67*	E: 21* NE: 8
Exp 1B	LDT											127**	12
Exp 2A	SCT		32				7 point self-rate: 5.2					E: 98** NE: 79*	E: 21* NE: 7
Exp 2B	LDT											76**	14
Wang and Forster (2015)													
Exp 2	LDT	Arizona, USA	24	Chi	/	/	TOEFL: 602.29 7 point self-rate: 5.92 (reading)	15	Chi: 50 Eng: 80	/	/	32*	5
Chen et al. (2014b)													
Exp 1	LDT	Beijing, China	44	Chi	After 11	/	6 point self-rate: 2.40 (reading)	14	50	/	150	/	4
Exp 2	LDT		46	Chi		/	6 point self-rate: 2.48 (reading)					28***	/

Table 4. (*Continued*)

Study	Task	Place of Testing	NoP	L1	AoFEE	YoFEE	EPro	NoI	Prime (ms)	Blank (ms)	Post-mask (ms)	Chi-Eng (ms)	Eng-Chi (ms)
Chen et al. (2014a)													
Exp 1	LDT	Beijing, China	40	Chi	10.5	/	6 point self-rate: 2.98 (reading); Oxford Placement Test: 39.2	15	50	/	150	Con: 18[#] Abs: 18	Con: 2 Abs: 1
Exp 2	SCT		42	Chi	11.5	/	6 point self-rate: 2.95 (reading); Oxford Placement Test: 37.3					Con: 89 Abs: 76	Con: 9 Abs: 13
Wang (2013)													
Exp 1	LDT	Singapore	20	Eng	NA	NA	7 point self-rate: 6.3 (reading)	40	50	/	/	0	39*
Exp 2	LDT		20	Chi or Eng & Chi	NA	NA	7 point self-rate: 6.3 (reading)					13*	24*
Luo et al. (2013)													
Exp 1	LDT	Guangzhou, China	27	Chi	/	/	/	12	50	/	150	/	48**
Exp 2a	LDT		27									77**	/
Exp 2b	LDT		26									/	30
Exp 3a	LDT		24									75**	/
Exp 3b	LDT		27									/	-4
Exp 4	LDT		27									30***	/

Table 4. (*Continued*)

Study	Task	Place of Testing	NoP	L1	AoFEE	YoFEE	EPro	NoI	Prime (ms)	Blank (ms)	Post-mask (ms)	Chi-Eng (ms)	Eng-Chi (ms)
Wang and Forster (2010)													
Exp 1	SCT	Arizona, USA	18	Chi	/	>= 8	/	40	50	/	150	/	E: 20** NE: -2
Exp 2	SCT		24					32	50	/	150	/	E: 14* NE: -4
Exp 3	SCT		24					32	50	/	/	/	E: 5* NE:2
Exp 4	SJT		14					24	50	/	150 for Eng-Chi	34**	-3
Jiang (1999)													
Exp 1	LDT	Arizona, USA	42	Chi	/	>= 8	TOEFL score of >= 550	16	50	/	/	45*	13*
Exp 2			44									68*	3
Exp 3			16							50	/	/	4
Exp 4			18							50	150	/	7
Exp 5			18							50	150	/	-2
Witzel and Forster (2012)													
Exp 1A	ERT	Arizona, USA	32	Chi	/	>= 8	TOEFL score of >= 550	16	50	50	150	/	O: 20*; N: -3
	LDT												1
Exp 1B	ERT											O: 55*** N: 23*	/

Table 4. (*Continued*)

Study	Task	Place of Testing	NoP	L1	AoFEE	YoFEE	EPro	NoI	Prime (ms)	Blank (ms)	Post-mask (ms)	Chi-Eng (ms)	Eng-Chi (ms)
Jiang and Forster (2001)													
Exp 1	ERT	Arizona, USA	26	Chi	/	>= 8	TOEFL score of >= 550	O:16 N:16	50	50	150	/	O: 29* N: -8
	LDT							16					8
Exp 2	LDT		18					O:16 N:16	50	50	150		O: 9; N: 3
Exp 3a	ERT		18					O:16 N:16	50	/	/		O: 7 N: 0
	LDT							16					4
Exp 3b	ERT		18					O:16 N:16	50	50	150		O: 52* N: -17
	LDT							16					9
Exp 4	ERT		22					O:16 N:16	50	/	/	O:6 N: 13	/
	LDT							16				41*	
Exp 5	ERT	Pennsylvania, USA	18					O:16 N:16	50	/	200	O:11 N: 13	

Note. **NoP**: number of participants; **AoFEE**: age of formal English education; **YoFEE**: years of formal English education; **EPro**: English proficiency; **Eng**: English; **Chi**: Chinese; **NoI**: number of items per cell; **SCT**: semantic categorization task; **E**: Exemplars, **NE**: non-Exemplars; **LDT**: lexical decision task; **ERT**: Episodic recognition task; **O**: old items, N: new items ; **SJT:** Size judgement task (e.g., bigger than a brick); **Con**: Concrete words; **Abs**: Abstract words; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; # Chen, Liang, Cui, and Dunlap (2014) did not directly compare translation pairs vs. control pairs

1.3.4 Non-cognate Masked Translation Priming: ERP Findings

In addition to the considerable amount of behavioural investigations with the non-cognate masked translation priming paradigm, recent studies using electrophysiological measures have provided further insight into translation priming effects. The use of electrophysiological measures has the advantage that it can unravel the time course of processing, which cannot be captured with behavioural data (Duñabeitia, Dimitropoulou, Dowens, Molinaro, & Martin, 2016; van Hell & Kroll, 2013).

Electrophysiological studies in the literature mainly focused on two ERP components to investigate translation priming effects. The first one is the N250, which is a negative-going wave peaking at around 250 ms after target presentation. In visual word recognition, N250 is assumed to index form-level processes during which orthographic mappings between the sub-lexical representations (e.g., letters, bigrams) and whole-word representation occur (Holcomb & Grainger, 2006). Applying this interpretation to masked translation priming studies, a reduced N250 would reflect lexical form overlap due to co-activation of translation equivalents (Hoshino et al., 2010; Midgley et al., 2009). Although it is possible that translation priming effects revealed by the N250 reduction are caused by automatic translation activation via a direct lexical link, the predominant interpretation is that the prime pre-activates its translation equivalent via connections from the pre-activated semantics (Hoshino et al., 2010; Midgley et al., 2009). Another electrophysiological indicator of translation priming effects is the N400 reduction. The N400 component is assumed to index the mapping between orthographic form and lexical meaning during word recognition (Hoshino et al., 2010; Midgley et al., 2009). Because it has been well documented that a reduced N400 amplitude reflects the greater ease in semantic processing (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984), a N400 reduction observed in the translation priming pradigm is interpreated as a cross-languge priming effect at the semantic level (Hoshino et al., 2010; Midgley et al., 2009).

The first electrophysiological study of masked translation priming was a study conducted by Midgley et al. (2009). In this study, a group of unbalanced French-English bilinguals took part in a go/no-go animal detection task (i.e., participants only respond to the words that refer to animals, e.g., *tiger*) with a prime presented for 50 ms followed

by a 17-ms backward mask. Critically, all the experiment trials were no-go trials in which participants should not respond. Their results revealed that L1-L2 translation priming resulted in a N250 and N400 reduction, whereas L2-L1 translation only reduced the N400 component. According to Midgley et al. (2009), L1 primes activate semantic representations and then their L2 translation equivalents, thus resulting in a semantic priming (a N400 reduction) and a lexical form priming effect (a N250 reduction). In contrast, the processing of L2 primes is slower so there is only sufficient time for a semantic processing (a N400 reduction), but not for L1 translation activation via semantic feedback as indicated by an absence of a N250 effect. If this interpretation from Midgley et al. (2009) is correct, it is reasonable to predict that both N250 and N400 effects would be observed in the L2-L1 direction when sufficient processing time is provided for the prime. To test this hypothesis, the prime duration was increased from 50 ms to 100 ms with a 20-ms backward mask in Schoonbaert's et al. study (2011), in which a group of unbalanced English-French bilinguals were asked to perform a lexical decision task. In addition to significant behavioural priming effects observed in both directions, translation priming modulated both the N250 and N400 components in both directions as well. These findings strongly support Midgley's et al. (2009) interpretation of translation priming effects in the N250 and N400 components. Importantly, Schoonbaert et al. (2011) also observed asymmetric translation effects in the ERP data with a stronger N400 effect in L1-L2 direction but a more pronounced N250 effect in L2-L1 direction.

Following these two seminal studies, Hoshino et al. (2010) reported for the first time cross-script masked translation priming effects in an ERP study. Using go/no-go body-part detection task, unbalanced Japanese-English bilinguals were presented with 50-ms prime words followed by a 30-ms backward mask and then a target word. Interestingly, the results revealed that L1-L2 translation priming modulated both N250 and N400 amplitude, whereas no L2-L1 translation priming effects were observed. In line with the interpretation provided by Midgley et al. (2009), Hoshino et al. (2010) argued that the processing of L2 primes was incomplete due to insufficient time and failed to pre-activate the sematic representation, in contrast to the fast processing of L1 primes, which facilitated semantic processing and the activation of L2 translation equivalents.

In addition to electrophysiological evidence of masked translation priming effects from unbalanced bilinguals, similar but slightly different patterns were found in balanced bilinguals. In the study by Duñabeitia, Dimitropoulou, Uribe-Etxebarria, Laka, and Carreiras (2010a), balanced Basque-Spanish bilinguals conducted a go/no-go animal detection task, during which the prime words were presented for 50 ms immediately followed by target words. A N400 effect, indicating translation priming at the semantic level, were found in both Basque-Spanish and Spanish-Basque directions with a similar magnitude. Interestingly, no N250 reduction was found. These results seem to indicate that for balanced bilinguals the semantic processing of primes is fast in both directions and it does not trigger automatic translation up to the form level.

The masked priming asymmetry reported in the electrophysiological studies reviewed above is consistent with behavioural studies using unbalanced bilinguals (e.g., Duñabeitia et al., 2010b; Gollan et al., 1997; Jiang, 1999; Perea et al., 2008). A further consistent finding reported in electrophysiological and behavioural studies is that L2-L1 translation priming effects were more often found in studies involving same-script languages than in studies involving different-script languages. This conclusion is based on the existing studies so far and there is only one electrophysiological study using different-script languages which failed to observe any L2-L1 translation priming effects (Hoshino et al., 2010). An important issue is that the number of ERP studies is still very limited compared to the abundant behavioural evidence of non-cognate masked translation priming effects. Surprisingly, although the lexical decision task is widely used in behavioural masked translation priming studies, only one published electrophysiological study has used a lexical decision task (Schoonbaert et al., 2011), whereas the remaining three studies used a go/no-go semantic categorization task (Duñabeitia et al., 2010a; Hoshino et al., 2010; Midgley et al., 2009). In addition to the fundamental difference between the tasks involved, electrophysiological studies using a go/no-go semantic categorization task are not able to measure translation priming effects at the behavioural level because the critical translation equivalents pairs have targets that are non-exemplars (e.g., not animal or body part), which do not require any responses in the go/no-go task. It is important to note that using translation equivalent pairs as non-exemplars is not a common practise in behavioural studies, because behavioural semantic categorisation studies usually presented translation equivalents both as exemplars and non-exemplars and behavioural responses are required for both

(e.g., Wang & Forster, 2010; Xia & Andrews, 2015). Another inherent difference between electrophysiological and behavioural studies, which is less obvious, is that the number of experimental trials used varies. Electrophysiological studies extract event-related potentials (ERP) from electroencephalograms (EEG) by averaging across many trials to increase the signal-to-noise ratio. In language studies involving EEG, it is necessary to present many trials per condition (Steinhauer, 2014). For example, for the N400 component, it is recommended to present at least 30-40 trials per condition (Luck, 2014). In contrast, the number of experimental trials used in behavioural studies vary considerably and usually this number is less than 40 (see Table 2 and Table 3). Given that using more trials may increase priming effects by reducing noise in the data (Van den Bussche, Van den Noortgate, & Reynvoet, 2009), it is likely that electrophysiological studies may have a better chance to detect priming effects. However, there is no way to test this hypothesis using the existing literature because there is only one published study that reported both behavioural and electrophysiological evidence of translation priming (Schoonbaert et al., 2011).

1.4 Translation Activation in Hidden Repetition Priming

Although masked translation priming studies have shed more light on bilingual visual word recognition, a potential limitation with these studies using cross-language priming, as pointed out by Thierry and Wu (2007), is that presenting words of two languages may create a bilingual context (Grosjean, 1998), which triggers non-selective lexical access in bilinguals. Therefore, Thierry and Wu (2007) used a new hidden repetition priming paradigm to investigate whether translation equivalents are activated in a purely monolingual context. In their event-related potentials (ERP) study, Chinese-English bilinguals were presented with English word pairs (e.g., *train* - *ham*) and they had to judge whether the presented word pairs were related in meaning or not. Unknown to the participants, the Chinese translations of the critical English pairs had a repeated Chinese character with identical pronunciation (e.g., the Chinese translation of *train* and *ham* is 火车 and 火腿 respectively, with a repeated character 火). Compared to the non-repeated control condition (e.g., *apple* - *table*, whose Chinese translations are 苹果 and 桌子 respectively), a reduced N400 was found in the repeated character

condition. Because no such difference was observed in the ERP data from a group of native English speakers, the reduced N400 was taken as the evidence for the unconscious activation of the Chinese translation equivalents.

Thierry and Wu (2007) used a unmasked hidden priming paradigm in which the prime words were presented for 500 ms. This is long enough for participants to activate the corresponding Chinese translations (Zhang et al., 2011). To avoid this potential issue, Zhang et al. (2011) used a masked priming paradigm with a lexical decision task to investigate hidden repetition priming among Chinese-English bilinguals. In their study, the position of hidden repeated characters was also manipulated: the target word *thing* [东西] was primed either by *east* [东] with the hidden repetition in the first character of **东**西, or by *west* [西] with the hidden repeated character in the second place. Only when the prime words had hidden repetition with the first characters of the targets (e.g., *east* [**东**] - *thing* [**东**西]) the response times of the target words were facilitated. This facilitation effect of hidden character repetition indicated that the activation of Chinese was fast and automatic when Chinese-English bilinguals read English words.

Wu and Thierry (2010) investigated whether the hidden character repetition effects found in their earlier study (Thierry and Wu, 2007) was driven by the activation of Chinese phonology or orthography. Using the same paradigm during which Chinese-English bilinguals were presented with English word pairs, the Chinese translations of critical English pairs either had the phonology or orthography repeated. For example, in the word pair *experience* [经验, /**Jing1**Yan4/[2]] - *surprise* [惊讶, /**Jing1**Ya4/] only the phonology of the first character is repeated, whereas in the word pair *accountant* [**会**计, /Kuai4Ji4/] - *conference* [**会**议, /Hui4Yi4/] only in the orthographic form is repeated. The ERP data revealed a reduced N400 for the hidden phonological repetition but not for the orthographic repetition, which suggests that Chinese-English bilinguals activate the phonology rather than the orthography of the equivalent Chinese

[2] Pinyin is a Romanisation of the Chinese pronunciation, with numbers indicating tones. More discussion on Chinese phonology and orthography is to be presented in Section 1.8.

translations.

To further explore whether Chinese phonological activation during English word reading is automatic, Wu and Thierry (2012a) investigated the impact of emotional valence on cross-language activation. The rationale is that if Chinese translation activation is automatic, the emotion valence manipulation would not influecne the hidden repetition effects previously observed. Following this logic, Wu and Thierry (2012a) manipulated emotion valence of the prime words which had phonological overlap with the target word when translated into Chinese, namely positive valence prime condition, (e.g., *honesty* [诚实, /**Cheng2**Shi2/] - *program* [程序, /**Cheng2**Xu4/]), negative valence prime condition, (e.g., *failure* [失败, /**Shi1**Bai4/] - *poet* [诗人, /**Shi1**Ren2/]), and neutral prime condition (e.g., *aim* [目标, /**Mu4**Biao1/] - *carpenter* [木匠, /**Mu4**Jiang4/]). Similar to their previous experiments (e.g., Wu & Thierry, 2010), Chinese-English bilinguals performed a semantic relatedness judgment task and the critical manipulations were in the unrelated pairs. Surprisingly, the N400 amplitude was modulated by phonological repetition when prime words were postive or neutral but NOT when the primes were of negative valence (e.g., *failure* - *poet*). These results have led to the conclusion that Chinese translation activation is not fully automatic and can be influenced by emotional valence of the English words.

Wu and Thierry (2011) extended their investigations of cross-language activation in a rhyming judgment experiment. In this experiment, unbalanced Chinese-English bilinguals were presented with two pictures and they had to judge whether the English name of the pictures rhymed or not. The bilingual participants were unaware that some English word pairs had a repeated character when translated into Chinese (e.g., *oven* [烤**箱**, /Kao3Xiang1/**]** - *mailbox* [邮**箱**, /You2Xiang1/]). The findings of their ERP experiment revealed a hidden repetition priming effect in terms of a reduced N400 amplitude compared to the control pairs without overt or covert rhyming (e.g., *slipper* [拖鞋, /Tuo1Xie2/ - *sausage* [香肠, /Xiang1Chang2/]). Because no repetition effects were found with native English speakers, the N400 effect can only be explained by assuming that the Chinese translations are activated when judging rhyme patterns in English. In contrast, when the rhyme decisions were made based on the Chinese name of the pictures, those word pairs whose English translation rhymed (e.g., the

corresponding English translations of 袜子 - 钟表 were st**ock** - cl**ock**) did not modulate N400 component. These findings suggest an asymmetric activation of translation equivalents in unbalanced bilinguals, with strong L1 (Chinese) activations during L2 (English) processing, but not vice versa.

Taken together, there is strong evidence for L1 (Chinese) translation activation when Chinese-English bilinguals process L2 (English) words, which is consistent with the idea of non-selective lexical access (Dijksta, 2005; Dijkstra & van Heuven, 2002; van Heuven et al., 1998). Importantly, using the hidden repetition paradigm, Morford, Wilkinson, Villwock, Piñar, and Kroll (2011) also observed similar findings in a group of deaf bilinguals, who used American Sign Language (ASL) as L1 and English as L2. Analogous to Wu and Thierry (Wu & Thierry, 2010, 2012a), ASL-English bilinguals were presented with pairs of English words and asked to judge whether these word pairs were related in meaning or not. Interestingly, ASL-English bilinguals were slower in rejecting English word pairs whose ASL translations shared formational parameters (i.e., handshape, location, movement and orientation). These results revealed that ASL-English bilinguals activate ASL (L1) translations when reading English (L2) words. Therefore, Morford's et al. (2011) study further confirmed that hidden repetition priming effects as the evidence of translation activation are not limited to Chinese-English bilinguals.

1.5 Translation Activation in Non-linguistic Tasks

1.5.1 Introduction to Translation Activation in Non-linguistic Tasks

One potential drawback of the above-mentioned studies using conventional linguistic tasks (e.g., the semantic judgement task, the lexical decision task, the translation recognition task) is that these tasks may generate in-depth processing of the presented word, which in turn activates their translation equivalents (Wu & Thierry, 2012b). In other words, the co-activation of translation equivalents might be a by-product of experimental tasks. Therefore, it is important to investigate whether co-activation of translation equivalents can be found when no linguistic tasks are used. In the literature, a handful of studies have used eye-tracking or ERP to investigate translation activation

in non-linguistic tasks. Although there is linguistic input in those studies, the tasks generally require visual decisions (e.g., to judge whether the presented stimuli is a circle or a square). Therefore, any evidence of translation activation is believed to indicate automatic language processing, suggesting that co-activation of translations is not task-dependent.

1.5.2 Translation Activation in Non-linguistic Tasks: Eye-tracking Findings

Eye-tracking studies are designed to capture patterns of eye movements, which can be used as an index of linguistic processing. One advantage of eye-tracking studies is that they can measure eye movements without overt responses from participants. Because of this advantage, eye-tracking studies are able to not only examine language processing among toddlers (e.g., Singh, 2013; Von Holzen & Mani, 2012) but also investigate language processing in adults that is less affected by strategies (Rayner, 1998, 2009). Recently, the eye-tracking technique has been gaining popularity to study bilingual language processing (for a review, see Conklin & Pellicer-Sánchez, 2016; Roberts & Siyanova-Chanturia, 2013).

Marian and colleagues conducted a series of seminal eye-tracking experiments to investigate cross-language activation during spoken word recognition (e.g., Marian & Spivey, 2003a, 2003b; Marian, Spivey, & Hirsch, 2003; Spivey & Marian, 1999). In one of their earliest studies, Spivey and Marian (1999) asked a group of Russian-English bilinguals to pick up one of the four objects according to the spoken instructions in separate Russian and English sessions. In both Russian and English sessions, participants were more likely to look at the distractor objects compared to the control objects when the names of the distractor objects in the irrelevant language shared the initial phonetic features with the names of the target objects. For example, when hearing the Russian instruction "Poloji **marku** nije krestika" ("Put the **stamp** below the cross"), participants made more eye movements (measured by the proportion of eye movements) to the distractor object "**marker**" whose English name had initial phonetic similarities to the Russian name of the target object (**marku**) compared to the control object (e.g., **ruler**) with no such phonetic similarities. This study provides strong evidence for the co-activation of translation equivalents during spoken word recognition in an object search task.

However, one potential issue with the study from Spivey and Marian (1999) is that participants were tested in both Russian and English sessions. Although the order of two sessions were counterbalanced, the experiment still created a bilingual context (Grosjean, 1998) which might trigger co-activation of the two languages. Therefore, in a follow-up study using the same object search task, two different groups of Russian-English bilinguals were instructed in either L2 English or in L1 Russian (Marian & Spivey, 2003b). In this study, the target objects (e.g., English name: **spear**) were accompanied by distractor objects whose names had phonological overlap with the target objects in the same language (within-language competition: e.g., **speaker**) and/or in the other language (between-language competition: e.g., matches, **spichki** in Russian), and control objects (e.g., **toothbrush**) with no phonological similarities to the target objects. The results showed that participants looked more often at the within-language competitor (e.g., **speaker**) compared to control objects regardless of the language of instruction. In contrast, participants looked more often at between-language competitors only when the task was instructed in bilinguals' L2 (English). These results indicate that L1 translation equivalents are activated during L2 language processing and L2 translation equivalents did not get activated during L1 processing.

This asymmetry was further investigated in a similar experiment (Blumenfeld & Marian, 2007) with German-English and English-German bilinguals. The bilinguals were instructed in English to identify a picture. Similar to previous studies (e.g., Marian & Spivey, 2003b; Spivey & Marian, 1999), German names of picture distractors had phonological overlap with the English name of target pictures and the names of these distractors were either cross-language cognates (e.g., *arm*, *Arm* in German) or non-cognates (e.g., *bike*, *Fahrrad* in German). Whereas English-German bilinguals looked more often at cognates (tested in their L1), German-English bilinguals looked more at both kinds of distractors (tested in their L2). These findings suggest that cross-language activation asymmetry is influenced by cognate status.

Recent eye-tracking studies also observed automatic L1 activation during L2 processing in cross-script bilinguals (Mishra & Singh, 2014; Singh & Mishra, 2015; Wu, Cristino, Leek, & Thierry, 2013). For example, Wu et al. (2013) asked a group of Chinese-English bilinguals to perform a visual search task in which circles or squares were presented together with English words in four corners of the screen. In critical

trials in which all the four corners were occupied by words, the Chinese translation of one English word had phonological overlap with the Chinese translation of circle or square (e.g., the Chinese translation of *circle* is 圆 pronounced as **/Yuan2/** and the Chinese translation of *reason* is 原因 pronounced as /**Yuan2**Yin1/). The eye-tracking data revealed that the English words with phonological overlap were gazed on more frequently (as measured by the number of fixations) and longer (as measured by the mean fixation duration) compared to those words which had no phonological overlap. These findings suggest that L1 Chinese equivalents are activated during L2 English processing in a purely visual non-linguistic task.

In sum, there is convincing evidence from eye-tracking studies which suggest that translation equivalents of task-irrelevant languages are activated in non-linguistic tasks. Only one ERP study has been published in the literature that investigated translation activation in a non-linguistic task. I will describe this study in the next section.

1.5.3 Translation Activation in Non-linguistic Tasks: ERP Findings

To investigate cross-language activation, Wu and Thierry (2012b) used a similar design as in their eye-tracking study (Wu et al., 2013). In this go/no-go study, Chinese-English bilinguals were presented with shapes (e.g., □ or ○ namely, square or circle) mixed with English words. Participants were asked to differentiate circles and squares by pressing buttons and withhold responses when presented with English words. Unknown to the participants, some of the English words had phonological overlap with the word *circle* or *square* when translated into Chinese (e.g., the Chinese translation of *circle* is 圆 pronounced as **/Yuan2/** and the Chinese translation of *reason* is 原因 pronounced as /**Yuan2**Yin1/), thus using as distractor words. When comparing these distractor words to the control words with no phonological overlap, an increased N200 amplitude was observed. Because the N200 is assumed to be an indicator of cognitive control (Folstein & Van Petten, 2008; Rodriguez-Fornells, De Diego Balaguer, & Münte, 2006; Rodriguez-Fornells, Schmitt, Kutas, & Munte, 2002), a more negative-going N200 indexed increased inhibition for distractor words, revealing that those words caused interference. Critically, the N200 effect was not found in a native English control group. Therefore, these electrophysiological results also indicate that cross-language activation of translation equivalents occurs in a non-linguistic task.

1.6 Summary of Empirical Studies Investigating Translation Processing

In Sections 1.2 to 1.5, I have provided a comprehensive review of empirical studies that used translation equivalents to investigate cross-language activation. In the translation recognition studies, the through-translation form interference effect has been found, indicating that translation equivalents are activated. Word translation activation in proficient bilinguals is generally assumed to occur after the word is processed up to the semantic level. Therefore, translation activation is only possible when sufficient processing time is provided (Guo et al., 2012; Moldovan et al., 2016). This interpretation is consistent with current views of masked translation priming studies. In masked translation priming experiments with a lexical decision task, the most important finding is the translation asymmetry: L1-L2 translation priming is larger than L2-L1 translation priming. Masked translation priming is generally attributed to semantic facilitation, and word translation activation is also possible via semantic feedback when processing time is sufficient (Hoshino et al., 2010; Midgley et al., 2009). In addition, recent evidence of translation activation comes from studies using hidden repetition paradigm, which revealed L1 translation activation in a pure L2 context. Further evidence of translation activation comes from experiments using non-linguistic tasks, suggesting that the activation of translation equivalents is task-independent. Taken together, all the studies provide further support for the idea of non-selective lexical access (e.g., Dijkstra, 2005; Dijkstra & van Heuven, 2002; Grainger & Dijkstra, 1992; van Heuven et al., 1998), which assumes that bilinguals activate both languages when processing words presented in one of their two languages. In light of these findings, the activation of word translations seems to be a fundamental feature of bilingual language processing. In the next section I will focus on models of bilingual word recognition that potentially can explain the activation of translations.

1.7 Models on Bilingual Word Recognition and Lexical Representation

Numerous theoretical frameworks have been proposed in the literature to account for bilingual language processing in terms of how words of two languages are represented and accessed (for a review, see Li & Zhao, 2015; Thomas & van Heuven, 2005; van

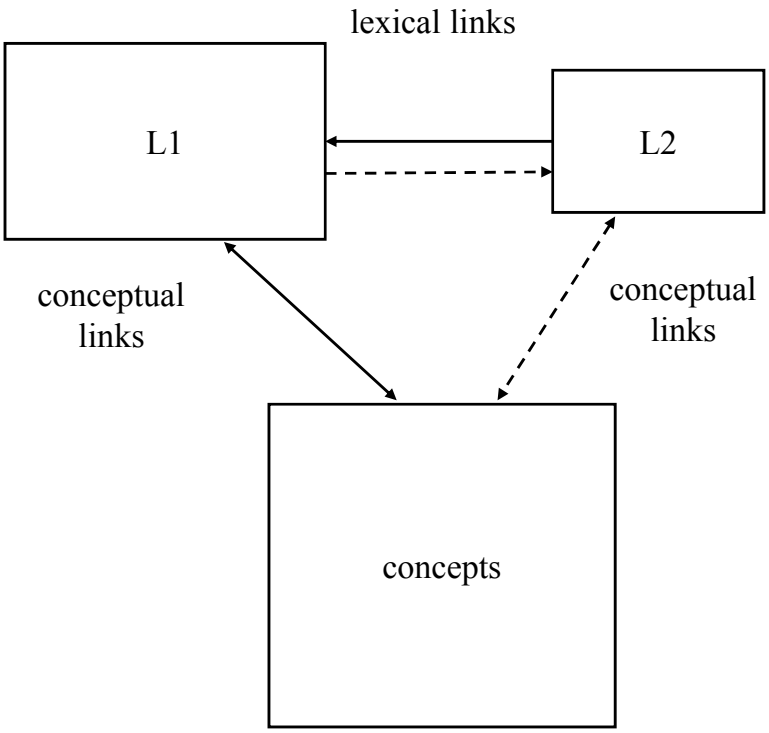
Heuven & Dijkstra, 2010). For example, the Language Mode framework (Grosjean, 1998), the distributed model (de Groot, 1992), the Inhibitory Control Model (Green, 1998; Green, 1986), the L2 Episodic Hypothesis (Jiang & Forster, 2001; Witzel & Forster, 2012), the Semantic, Orthographic, and Phonological Interactive Activation (SOPHIA) model (van Heuven, 2005), BIA-d Model (Grainger, Midgley, & Holcomb, 2010) and Self-organizing Connectionist Model of Bilingual Processing (Li & Farkas, 2002). In this section, I will focus on six models and evaluate their strength and limitations in terms of empirical evidence from studies using translation equivalents. First, the Revised Hierarchical Model and the BIA+ Model are discussed because these are the most frequently mentioned models in the literature. Next, I review the Multilink Model, which incorporates some basic assumptions of the Revised Hierarchical Model and the BIA+ Model. Finally, the Sense Model, the Distributed Representation Model and the DevLex-II model are discussed. The Sense Model and the Distributed Representation Model are specially put forward to explain the masked translation priming asymmetry, and the DevLex-II model uses a computational approach to simulate translation priming asymmetry.

1.7.1 The Revised Hierarchical Model

The Revised Hierarchical Model (RHM) was originally proposed by Kroll and Stewart (1994) to explain the findings of word production experiments with bilinguals. The RHM has been an influential theoretical model that has been widely used to explain findings from bilingual word recognition, production, naming and translation (for a review, see Brysbaert, Verreyt, & Duyck, 2010). As shown in Figure 3, the RHM distinguishes between the lexical representations (i.e., words) and conceptual representations (i.e., meanings of words). An important assumption of the RHM is that words of two languages of bilinguals are represented in separate lexicons. Because bilinguals know more L1 words than L2 words, the RHM assumes that bilinguals have a larger L1 lexicon than L2 lexicon (indicated in Figure 3 by bigger square for L1 than for L2 in). Regardless of the difference in sizes, the words of both languages are connected to a shared concept store. However, the conceptual link between L1 words and concepts is strong (solid line) in contrast to a weak conceptual link between L2 words and concepts (dotted line). These conceptual links enables L1 and L2 words to be connected via a semantic route, e.g., L1 words → concepts → L2 words. Apart from

these indirect links via concepts mediation, there also exists direct lexical links between L1 and L2 words, with a stronger lexical link from L2 to L1 (solid line) than from L1 to L2 (dotted line). Importantly, the RHM captures a dynamic changes of the bidirectional conceptual link between L2 words and concepts as the L2 proficiency increases. It is assumed that the conceptual link between L2 words and concepts is very weak for novice bilinguals, whereas for proficient bilinguals the conceptual link to L2 words becomes stronger in strength. Therefore, the RHM predicts that when accessing meaning of L2 words, novice bilinguals have to rely on L1 translations (i.e., L2 words → L1 words → concepts), whereas proficient bilinguals do not need to activate L1 translations (i.e., L2 words → concepts).

Figure 3. The Revised Hierarchical Model (figure adapted from Kroll & Stewart, 1994)



Initial support for the RHM comes from two findings from Kroll and Stewart (1994). In their study bilinguals were asked to verbally translate words from one language to another language (i.e., word translation). The first key finding was that L2-L1 word translation was faster than L1-L2 translation. The second finding was that when presenting words within the same category (e.g., dress, suit, sweater), which is a semantic manipulation, an interference effect was only observed for L1-L2 word

translation. These results are consistent with the assumption that L1-L2 translation involves concept mediation (i.e., L1 words → concepts → L2 words), so it is slower and more prone to semantic manipulations. In contrast, L2-L1 word translation is completed via a lexical mediation (i.e., L2 words → L1 words), therefore it is faster and less affected by semantic manipulations. Despite these seemingly convincing results, it is likely that the word translation asymmetry found in Kroll and Stewart (1994) is caused by a slower L2 production than L1 production (de Groot, 2011; Snodgrass, 1993). In addition, contradictory findings have been reported in the literature. First, in contrast to the word translation asymmetry found in Kroll and Stewart (1994), some studies have reported equally fast word translation in both directions (e.g., de Groot et al., 1994; de Groot & Poot, 1997; van Hell & De Groot, 1998a). Second, mixed findings have been found in terms of whether L2-L1 word translation involves a direct lexical mediation (Sholl et al., 1995) or a concept mediation (de Groot & Poot, 1997; Duyck & Brysbaert, 2004, 2008; La Heij et al., 1996; Potter et al., 1984).

Furthermore, the RHM has been challenged by empirical findings of bilingual word recognition studies. As Brysbaert and Duyck (2010) pointed out, although it is not explicitly mentioned in the RHM paper (Kroll & Stewart, 1994), the assumption of separate lexicons predicts selective lexical access, which is contradictory to the majority of findings in word recognition studies. However, Kroll, van Hell, Tokowicz, and Green (2010) argued that "the two lexicons are functionally separate but with parallel access and sublexical activation that creates resonance among shared lexical features" (p. 374). Kroll et al. (2010) further argued that the RHM is predominantly a model of bilingual word production and not bilingual word recognition. This claim, as Dijkstra and Rekké (2010) explained, may be problematic because it is assumed that word translation does not involve word recognition at all.

In addition, following the rationale of a stronger lexical link from L2 to L1, it makes perfect sense for the RHM to predict stronger masked translation priming effects from L2 to L1 than vice versa (Brysbaert & Duyck, 2010; Grainger et al., 2010; Schoonbaert et al., 2009), which is clearly inconsistent with the empirical findings (see Section 1.3). In response to this challenge, Kroll and Tokowicz (2001) explained the translation priming asymmetry in terms of different strength of conceptual links. Following this

rationale, it would predict a stronger semantic activation by L1 primes than L2 primes, thus producing a stronger priming in L1-L2 direction than L2-L1 direction (Kroll & Tokowicz, 2001). However, it is very puzzling why the lexical links in the RHM do not impact the bilinguals' performance in masked translation priming experiments if the lexical links exist.

Another issue for the RHM is that the assumption that only novice bilinguals would activate L1 translations during L2 word processing is at odds with numerous studies which observe L1 translation activation during L2 processing in proficient bilinguals (e.g., findings of studies using hidden repetition priming paradigm reviewed in Section 1.5). Kroll and colleagues (Guo et al., 2012; Morford et al., 2011) argued that L1 translation activation is a route to access L2 meaning in novice bilinguals, whereas in proficient bilinguals L1 translations are activated after meaning is accessed. However, if this explanation is correct, it is unclear whether L1 translations are activated through the lexical link from L2 or via a conceptual link from the activated concepts (i.e., L2 words → concepts → L1 words).

Finally, the original framework of the RHM cannot capture the existence of translation ambiguity between translation equivalents. Obviously, translation equivalents do not always have one-to-one mappings. For example, the English word *surprise* can be translated into two Chinese words that are synonyms, namely 惊讶 and 惊喜, and the English word *right* can be translated into two Chinese words with different meanings, namely 右边 (the antonym of *left*) and 正确 (the antonym of *wrong*). If a word in one language has more than one translation equivalent in another language, it is considered to be translation ambiguous. Because recent studies have revealed that translation ambiguity does indeed influence bilingual language processing (see Chapter 2), Eddington and Tokowicz (2013) put forward the Revised Hierarchical Model of Translation Ambiguity (RHM-TA) to account for the representation of translation-ambiguous pairs. In Figure 4, one L1 word corresponds to two L2 translations which have similar meanings (one unique concept representation), whereas in Figure 5, one L1 word corresponds to two L2 translations which have distinctive meanings (two different concept representations and two conceptual links to L1). Because of two lexical links from L1 to L2 in translation-ambiguous pairs, the RHM-TA framework

would predict that both types of translation-ambiguous pairs would be processed slower than translation-unambiguous pairs (e.g., *tiger* - 老虎). In addition, as Eddington and Tokowicz (2013) explained, the RHM-TA framework would also predict that presenting a semantic related prime would show stronger facilitation in meaning translation-ambiguous pairs (e.g., *side* - *right* - 右边) than in synonym translation-ambiguous pairs (e.g., *shock* - *surprise* - 惊喜). The logic is that the related prime can disambiguate the meaning translation-ambiguous pairs, whereas the related prime cannot disambiguate synonym translation-ambiguous pairs. Although the first assumption has been supported by some studies (Boada, Sanchez-Casas, Gavilán, García-Albea, & Tokowicz, 2013; Eddington & Tokowicz, 2013; Laxén & Lavaur, 2010; Tokowicz & Kroll, 2007), the evidence of the second assumption has not been found (Eddington & Tokowicz, 2013).

Figure 4. The Revised Hierarchical Model of Translation Ambiguity adapted to account for synonym translation-ambiguous pairs (figure adapted from Eddington & Tokowicz, 2013)

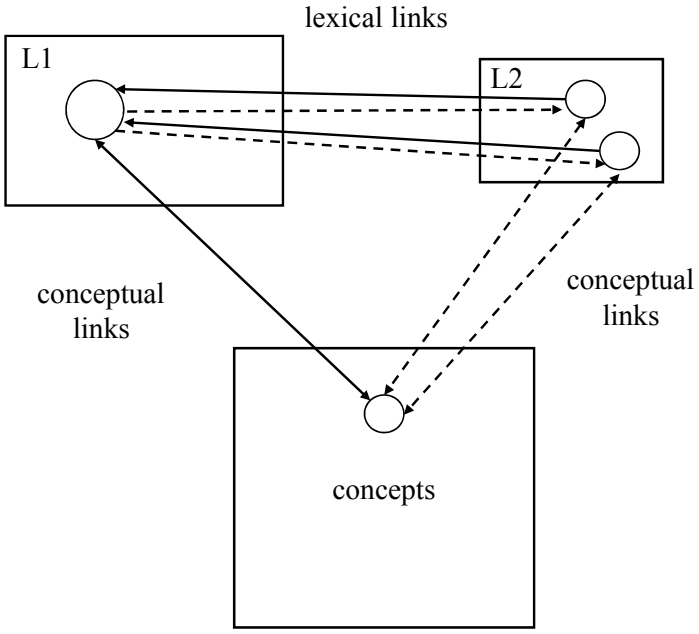
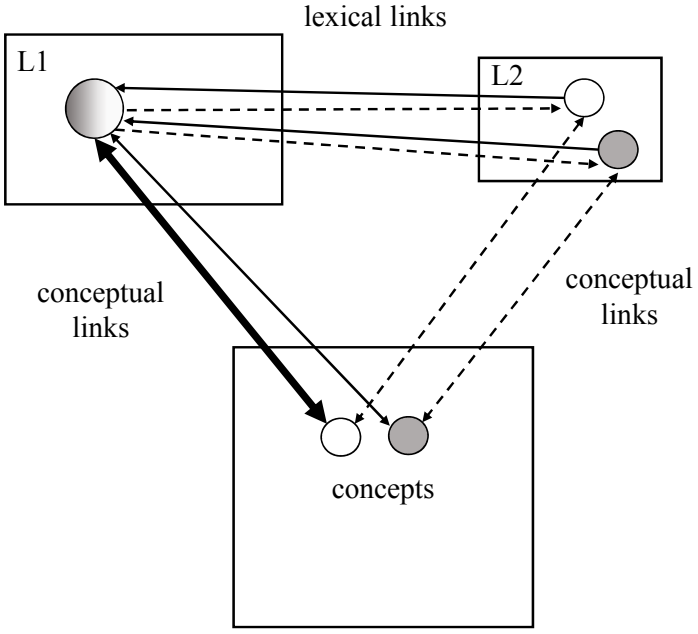


Figure 5. The Revised Hierarchical Model of Translation Ambiguity adapted to account for meaning translation-ambiguous pairs (figure adapted from Eddington & Tokowicz, 2013)



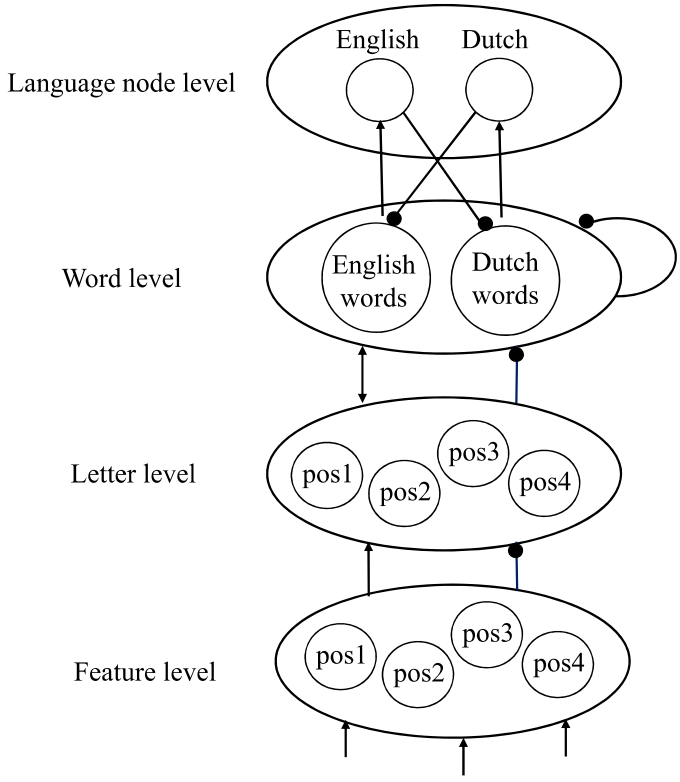
In sum, although the RHM-TA framework is a new development of the RHM, it is still in its infancy. It cannot resolve the challenges that the RHM has been facing. Furthermore, it is obvious that the current RHM-TA framework only captures the situation where one L1 word maps onto two L2 translation equivalents. Given that the original RHM assumes that L1 lexicon is larger than L2 lexicon, it is very possible that translation ambiguity occurs more often in the opposite direction, namely, one L2 word with multiple L1 translation equivalents. It is unclear what the RHM-TA framework would predict in this situation. Finally, the RHM-TA framework, as well as the RHM, seem to assume that translation equivalents are concept equivalents, which neglects the fact that translation equivalents may have shared and language-specific concepts (de Groot, 1992; Dong, Gui, & MacWhinney, 2005; Finkbeiner et al., 2004; Schoonbaert et al., 2009).

1.7.2 The BIA+ and BIA Models

The Bilingual Interactive Activation Plus (BIA+) model (Dijkstra & van Heuven, 2002) is an influential theoretical model of bilingual word recognition. The BIA+ model is a successor of the localist-connectionist BIA model (van Heuven et al., 1998). The BIA model is essentially the bilingual version of the interactive activation model (McClelland & Rumelhart, 1981) with an extra layer of language nodes. Thus, it consists of in total four hierarchical layers: feature, letter, word and language nodes (see Figure 6, arrows with triangle heads indicating activation and arrows with circular heads indicating inhibition). During the word recognition process, the features of each letter at each position are activated, followed by the activation of letters. Because the activated letters are position-specific, each of these activated letters subsequently activate the word candidates which share the same letters at the same position and inhibit the word candidates that have different letters at that particular position. The activated words then inhibits all the other words within the word level (known as lateral inhibition, depicted as the external arrow at the word level in Figure 6), and activate their corresponding language node. When the corresponding language node is activated, it produces a top-down effect by inhibiting word candidates from the other language (the arrow with a circular head between language node level and word level in Figure 6). The BIA model assumes an integrated lexicon of both languages and the idea of non-selective lexical access. In addition, another important assumption of the BIA

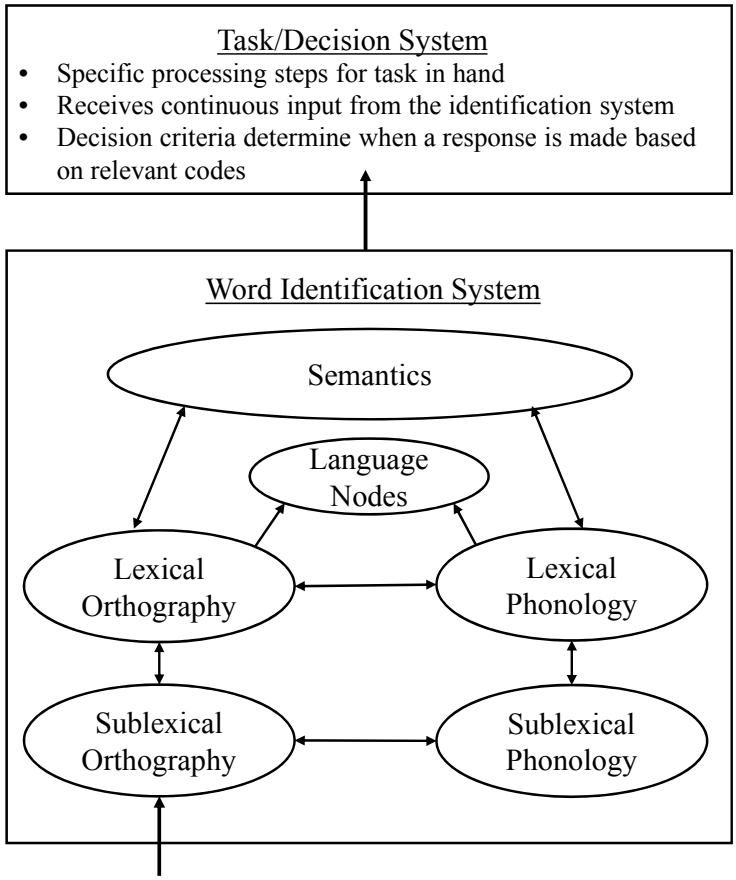
model is that language activation is determined by the resting-level activation of words. This resting-level activation of words depends on the word frequency. For unbalanced bilinguals, L1 words have a higher subjective frequency than L2 words because they are more proficient in L1 than L2. Consequently, the resting-level activation of L1 words is higher than L2 words so L1 words can be quickly activated. In a similar vein, L2 words are activated more slowly because of their lower resting-level activation. However, as L2 proficiency of bilinguals increases, the difference of resting-level activations between L1 and L2 will become smaller. Although the BIA model has successfully simulated bilingual orthographic processing (for a review, see Thomas & van Heuven, 2005), it cannot capture phonological and semantic activation during bilingual word recognition.

Figure 6. The BIA mode (figure adapted from van Heuven, Dijkstra, & Grainger, 1998)



To address this limitation and to account for new empirical findings, the BIA+ model was proposed that extends its predecessor in several aspects, e.g., by incorporating semantic and phonology levels in the model and by distinguishing between a word identification system and a task/decision system (see Figure 7). Within the word identification system of the BIA+ model, sublexical phonological and orthographical codes activate the phonology and orthography of words. The activated lexical information (phonology and orthography) then activates the corresponding language node. All the activated representations in the word identification system then flow to the task/decision system, where a response relevant for a specific task is made. It is obvious that the BIA+ model shares the basic mechanisms of bilingual word recognition implemented in the BIA model. However, a crucial difference between the BIA+ model and the BIA model lies in the role of language nodes. In the BIA+ model, there is no top-down influence from the language nodes to the phonology and

Figure 7. The BIA+ Model (figure adapted from Dijkstra & van Heuven, 2002; van Heuven & Dijkstra, 2010)



orthography. Despite this difference, the BIA+ model shares the underlying assumptions of the BIA model, such as the assumption of an integrated lexicon of both languages, the idea of non-selective lexical access, and the resting-level activation of L1 and L2 words determined by word frequency and language proficiency of bilinguals (van Heuven & Dijkstra, 2010).

The assumptions of the BIA+ model have been supported by a large body of behavioural and neuroimaging studies (van Heuven & Dijkstra, 2010), and it is especially consistent with studies on translation processing reviewed in this chapter. Obviously, the idea of non-selective lexical access is compatible with translation co-activation observed in the translation recognition task (Section 1.2), the hidden repetition paradigm (Section 1.4) and the studies using non-linguistic tasks (Section 1.5). In addition, the difference between the resting-level activations of L1 and L2 words can account for the translation priming asymmetry found in masked priming studies (Section 1.3). That is, because the resting-level activation of L2 words is lower, the representations of L2 primes are activated relatively slowly, resulting in a weak facilitation effect on L1 targets. On the contrary, the resting-level activations of L1 words are high, so L1 primes can be quickly activate L1 words, thus generating a strong facilitation effect on L2 targets.

However, although the assumptions of the BIA+ model are compatible with studies reviewed in this chapter, one potential issue is that the BIA+ model is originally proposed based on studies with two alphabetic languages (van Heuven & Coderre, 2015b). It is unclear whether the BIA+ model can account for the mechanisms of cross-language findings with different-script languages (e.g., Chinese and English). In particular, it is important to consider how hidden repetition priming effects in Chinese-English bilinguals (Section 1.4) could occur in light of the BIA+ model. Given that the logographical Chinese does not use letters for writing, Chinese and English have no orthographic overlap. In addition, the Chinese-English translations involved are non-cognates, so there is no phonological overlap as well. Therefore, the activation of Chinese translations during English words reading cannot be a result of lexical activation from the sublexical levels, and the only possibility is that Chinese translations are activated by the feedback from semantics. However, because Chinese phonology is believed to be the source of hidden repetition priming effects, it is puzzling why the

feedback from semantics only activates phonology but not orthography. One possible explanation is that scripts has been used as a cue for bottom-up inhibition of the orthography, so Chinese orthography cannot be activated. Another possible explanation is that Chinese-English bilinguals automatically translate English words into Chinese through unconsciously and covertly articulating the Chinese translations so the Chinese phonology is activated but not the Chinese orthography of the translation equivalent. Furthermore, another puzzle is whether the feedback from semantics activate lexical Chinese phonology (a combination of segmental and tonal information) or/and sublexical Chinese phonology (separate units of segmental and tonal information). Taken together, although the idea of non-selective lexical access can explain hidden repetition priming effects, the exact mechanism of the findings is underspecified.

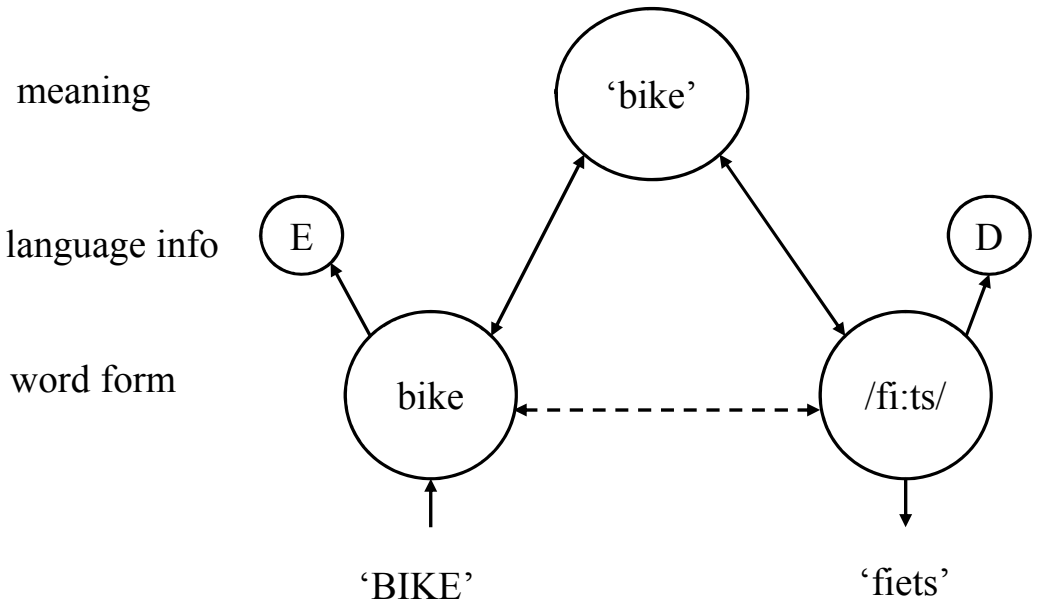
1.7.3 The Multilink Model

Dijkstra and Rekké (2010) developed recently the Multilink model, which is based on the BIA+ model (Dijkstra & van Heuven, 2002) and the RHM (Kroll & Stewart, 1994). Similar to the BIA+ model, the Multilink is a localist-connectionist model, which consists of both a word identification system and a task/decision system. The Multilink model captures the RHM's assumption of different strengths of conceptual links to L1 and L2 words by manipulating the resting-level activation of words which depends on word frequency and the bilinguals' language proficiency. Importantly, the Multilink model improves upon the two influential models discussed above in two major aspects.

First, unlike the BIA+ model, the Multilink model does not have sublexical levels of phonology and orthography. This is because the coding of letter position implemented in the BIA/BIA+ model is position-specific, which are at odds with new empirical evidence (e.g., transposition priming, *desgin* - *design*) and thus requires new coding approach (for more discussion, see Grainger, 2008; van Heuven & Coderre, 2015b). Omitting the sublexical levels has an advantage because it enables Multilink to simulate word length effects (i.e., quicker processing for shorter words) and orthographic similarity effects (i.e., the word *ice* activate the word *price*).

Another important improvement is that Multilink can simulate word translation as well as word recognition. Multilink's theoretical framework for word translation is depicted in Figure 8. It is important to note that Multilink implements BIA+'s lateral inhibition at word level (i.e., the activated word candidates inhibits all the other word), which stands in clear contradiction to the facilitatory lexical links between L1 and L2 words in the RHM. Therefore, Multilink assumes that word translation is only driven by conceptual links, e.g., L1 words → concepts → L2 words. Importantly, if the results of word translation simulation are in line with empirical data (e.g., unbalance bilinguals translate faster from L2 to L1 than vice versa), it would provide support for the BIA+'s lateral inhibition mechanism. Otherwise, it would suggest the possible existence of the direct lexical links. Interestingly, the Multilink successfully simulated faster L2-L1 word translation than L1-L2 word translation in unbalanced bilinguals and equally fast word translation from both directions in balanced bilinguals.

Figure 8. The theoratical framework of the Multilink Model for word translation (figure adapted from Dijkstra & Rekké, 2010)



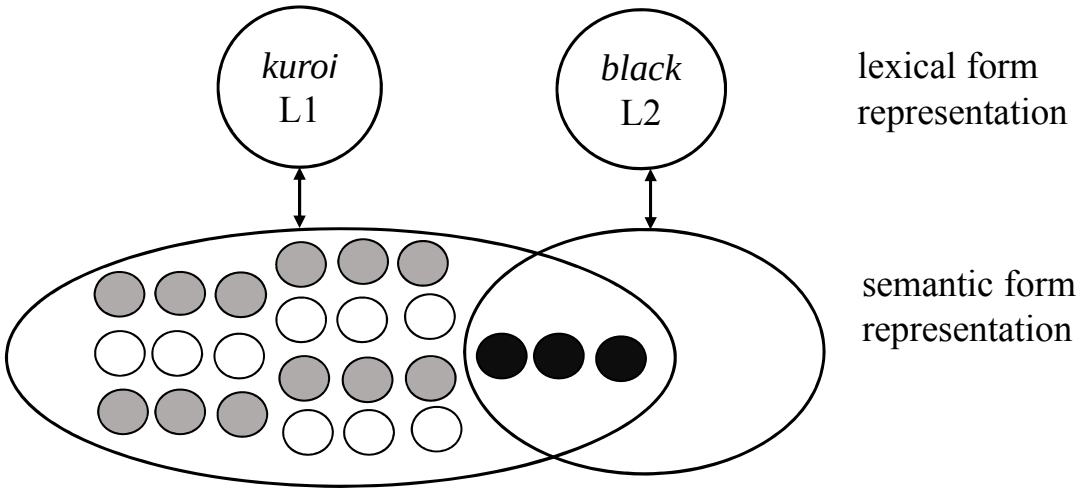
Taken together, Multilink is capable of stimulating not only bilingual word recognition at various levels (e.g., orthography, semantic) but also bilingual word translation from both L1-L2 and L2-L1 direction in unbalanced and balanced bilinguals. In addition, it directly tests the contrast between the RHM and the BIA+ model in terms of how

translation equivalents are interconnected, thus throwing new light on understanding bilingual language processing. However, Multilink is limited to simulating bilingual word recognition and production in Dutch-English bilinguals. It is unclear whether it is capable of stimulating other language pairs, especially different-script language pairs (e.g., Chinese and English). Secondly, it may be necessary to implement the sublexical levels in the Multilink model in the future. With sublexical levels the model might be able to simulate the through-translation form interference effects (Section 1.2). Also the sublexical levels may be crucial for simulating different-script language pairs because the dramatic differences in orthography may impact bilingual language processing. Furthermore, although differences between resting-level activations of L1 and L2 words are compatible with the translation priming asymmetry found in masked priming studies (Section 1.3), Dijkstra and Rekké (2010) have not demonstrated whether the Multilink could simulate the same patterns. Finally, like the BIA+ model, the Multilink also cannot reveal the underlying mechanism of hidden repetition priming effects in Chinese-English bilinguals (Section 1.4).

1.7.4 The Sense Model

The Sense Model (Finkbeiner et al., 2004) is a theoretical model for lexical representations which proposes that translation priming asymmetry in a lexical decision task is attributed to a representational asymmetry between L1 and L2 words. Thus, L1 words are associated with more semantic senses than L2 words. To be more specific, in the Sense Model, semantic senses are referred to as bundles of semantic features. As depicted in Figure 9, circles are symbols of semantic features and three circles are considered as one semantic sense. In the Japanese-English translation pair *kuroi* - *black*, the core sense of colour is shared (dark grey circles in the figure). It is obviously that the L1 word *kuroi* contains almost all the semantic senses of the L2 word *black*, but not vice versa. As a consequence, L1 primes activate almost the complete set of semantic senses of L2 targets, so L1 primes can facilitate L2 targets in a lexical decision task. In contrast, L2 primes only activate a subset of semantic senses of L1 targets, which are insufficient to facilitate the recognition of L1 targets. In sum, this theoretical account predicts that there is no L2-L1 translation priming in a lexical decision task. Another important prediction of this insightful theoretical account is that L2-L1 translation

Figure 9. The Sense Model (figure adapted from Finkbeiner, Forster, Nicol, & Nakamura, 2004)

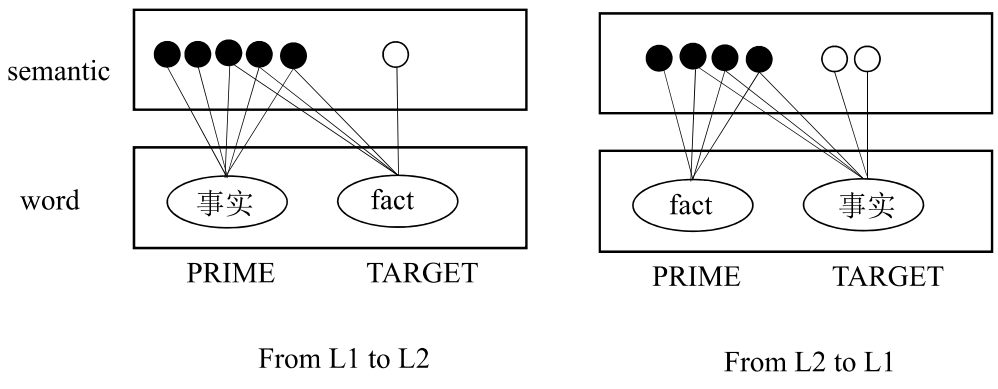


priming could be obtained when words are exemplars in a semantic categorization task. For example, in a colour categorization task, the core sense of *kuroi* is preserved while its other senses are filtered out. Benefiting from the filter effect of the task, the core sense pre-activated by L2 primes is now sufficient to generate a priming effect on L1 targets, and vice versa. So far, the Sense Model has supported by some empirical studies (Finkbeiner et al., 2004; Luo et al., 2013; Wang & Forster, 2010). However, it cannot account for the recent studies that found significant translation priming effects in both directions in lexical decision task (e.g., Greek-English: Dimitropoulou et al., 2011b; Dutch-English: Duyck & Warlop, 2009; Japanses-English: Nakayama et al., 2016). It also cannot explain why translation priming is independent of semantic senses (Allen, Conklin, & van Heuven, 2015) and why translation priming is still stronger in L1-L2 than L2-L1 even in exemplars (Xia & Andrews, 2015).

1.7.5 The Distributed Representation Model

In light of the recent findings that reported significant translation priming effects in both directions (e.g., Greek-English: Dimitropoulou et al., 2011b; Dutch-English: Duyck & Warlop, 2009; Japanses-English: Nakayama et al., 2016), Schoonbaert et al. (2009) proposed a refined Distributed Representation Model (DRM), which was based on the distributed model of de Groot (1992). According to the DRM, L1 words are connected to more semantic nodes than L2 words, so L1 primes activate a larger proportion of semantic nodes of L2 targets than vice versa. For example, a L1 word is connected to 5 semantic features and its L2 translation equivalent is connected to 4 semantic features. Importantly, semantic features could be language-specific or identical across languages. In this example, the L1 word and its L2 translation equivalent share 3 semantic features. As illustrated in Figure 10, these 3 identical semantic features are activated by the primes and these drive the priming effect. Consequently, 75% (3 out of 4) of semantic features in the L2 target are pre-activated, whereas only 60% (3 out of 5) of semantic features in the L1 target are pre-activated. Therefore, priming effects are stronger in L1-L2 direction than L2-L1 direction. Critically, within the DRM differences between L1-L2 and L2-L1 translation priming are quantitatively different rather than qualitatively as implied in other models (Jiang & Forster, 2001; Kroll & Stewart, 1994; Witzel & Forster, 2012). Thus, the model predicts significant L2-L1 translation priming effects, but smaller than L1-L2 translation priming effects. Arguably, a potential limitation of the DRM is that it only explains translation priming effects in terms of semantic overlaps between translation

Figure 10. The Distributed Representation Model (figure adapted from Schoonbaert, Duyck, Brysbaert, & Hartsuiker, 2009)



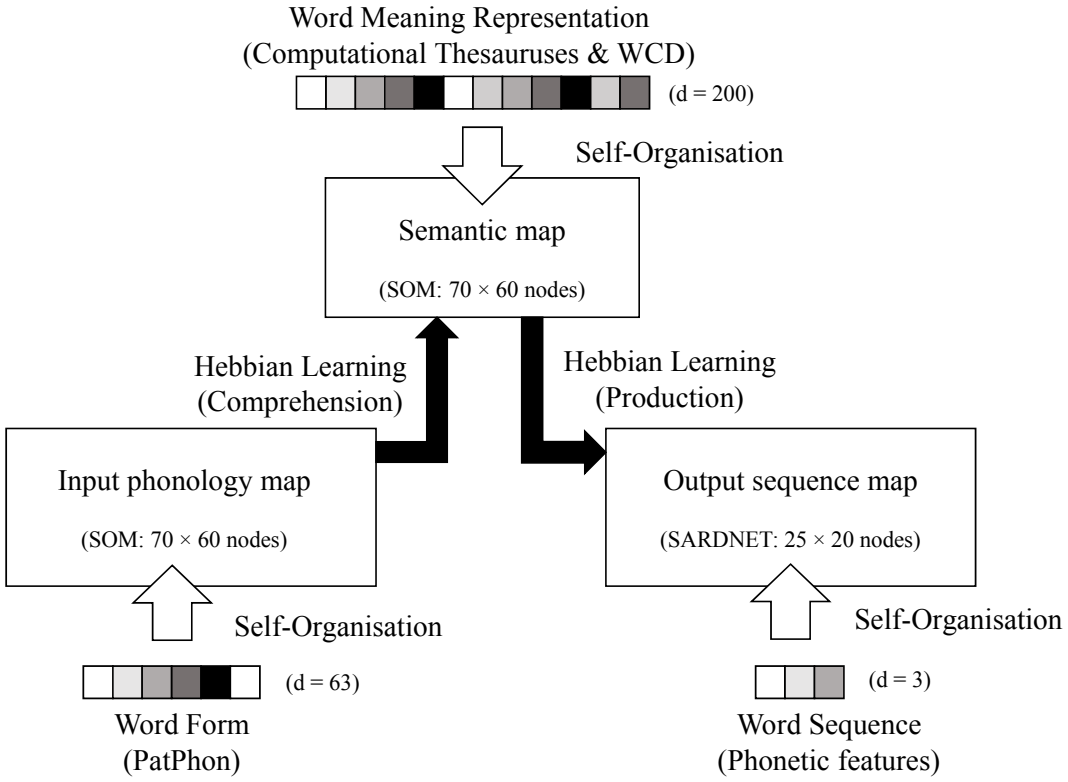
equivalents. Given that the semantic features activated by the prime may also activate its translation equivalent (the target), lexical form overlaps may also contribute to translation priming effects (Hoshino et al., 2010; Midgley et al., 2009).

1.7.6 The DevLex-II Model

The DevLex-II model is a self-organising neural network developed by Zhao and Li (Zhao & Li, 2010, 2013). In contrast to the localist-connectionist BIA model (van Heuven et al., 1998) and the Multilink model (Dijkstra & Rekké, 2010), DevLex-II is a distributed-connectionist model which focuses on dynamic changes in the bilingual mental lexicon of Chinese-English bilinguals with different onsets of L2 (English) acquisition (e.g., early vs. late bilinguals). As depicted in Figure 11, DevLex-II consists of three layers. At the centre of the model is the semantic map, which stores semantic/conceptual information and connects the other two layers, namely the input phonology map and the output sequence map. In the latest version of DevLex-II, Zhao and Li (2013) included 500 words from both English and Chinese. During the training stage which models the word acquisition in both languages, the Hebbian learning and spreading activation mechanisms were implemented to map the input phonology onto the output phonology via the semantic map. It is important to note that the structure and the learning mechanisms of the DevLex-II model are based on the DevLex model (Li, Farkas, & MacWhinney, 2004), which aimed to simulate first language acquisition in monolingual children. After training, the final model is assumed to represent a network of bilingual lexical representation and can be used for simulating cross-language priming effects. Zhao and Li (2013) have demonstrated that DevLex-II is capable of successfully stimulating translation priming and cross-language semantic priming in both directions (L1-L2 and L2-L1), in short and long SOAs (50 vs. 150 units) and in early and late Chinese-English bilinguals. The results of their simulations remarkably revealed several patterns in line with the empirical observations. In particular, I want to highlight three interesting results of the simulated translation priming effects. First of all, the L1-L2 translation priming is larger than the L2-L1 translation priming, and this translation priming asymmetry is independent of the onsets of L2 learning and SOAs. Secondly, translation priming asymmetry is smaller in early bilinguals than late

bilinguals. Finally, for both early and late bilinguals, translation priming effects in both directions are larger in the long SOA than in the short SOA.

Figure 11. The structure of the DevLex-II Model (figure adpated from Zhao & Li, 2010, 2013)



DevLex-II is undoubtedly an ambitious model and it is one of the few models built for processing Chinese and English words. However, an obvious limitation of the DevLex-II model is that orthography is not included (van Heuven & Coderre, 2015b). Therefore, it is unclear how the orthography of two languages that differ in script is represented. Importantly, the role of L2 orthography may be crucial for late bilinguals, especially for late bilinguals who learnt the L2 in a class-room instruction method. In addition, Zhao and Li (2013) have not demonstrated whether DevLex-II could simulate hidden repetition priming effects in Chinese-English bilinguals (Section 1.4). Nevertheless, an important implication drawn from the model is that segment and tone are separately coded in Chinese (for details, see Zhao & Li, 2009) because the model uses five slots

to represent phonemic units in segments (i.e., consonant + vowel + vowel + vowel + consonant) and an extra slot to represent the tone for each syllable.

1.7.7 Summary

The six models presented in this section help to provide a framework to understand how translation activation occurs. Each of these models has its own strengths and limitations, and they differ in terms of how translation equivalents are interconnected and how cross-language activation occurs. For example, in the RHM, translation equivalents are connected by lexical links and conceptual links, whereas translation equivalents are only connected by semantic links in the other five models reviewed above. Interestingly, when applying the assumption of the RHM to explain co-activation of translation equivalents, Morford et al. (2011) argued that words are processed up to the semantic level and then their translation equivalents in the other language are activated as a way to boost semantic processing. In contrast, co-activation of translation equivalents is interpreted as a result of non-selective lexical access in the BIA+ and Multilink models. Overall, one common limitation of these models is that they are not suitable to explain the empirical data from Chinese-English bilinguals, so they cannot explain the underlying mechanism of hidden repetition priming effects (Section 1.4). Although there are a few models based on Chinese-English bilinguals, for example, Self-organizing Connectionist Model of Bilingual Processing (Li & Farkas, 2002) and DevLex-II (Zhao & Li, 2010, 2013), these two distributed-connectionist models focus on the developmental change in bilingual language processing. Therefore, they do not make explicit predictions on how cross-language activation occurs in Chinese-English bilinguals, and thus cannot account for the mechanism of hidden repetition priming effects as well.

1.8 Language Processing in Chinese-English Bilinguals

The present thesis focuses on Chinese-English bilinguals to investigate the co-activation of translation equivalents. Therefore, this section will present a selective review of the relevant studies on Chinese visual word recognition and cross-language phonological activations. These studies are crucial for the interpretation of the findings

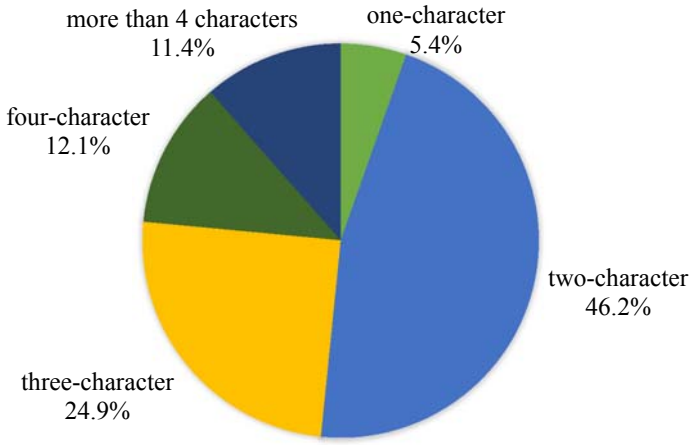
of cross-language activation in Chinese-English bilinguals, in particular, the hidden repetition priming effects. I will begin with a brief introduction to the Chinese language (section 1.8.1). Then I will present a selective review of behavioural and electrophysiological studies on Chinese word reading (Section 1.8.2 & 1.8.3). Next, Section 1.8.4 reviews cross-language phonological activations between Chinese and English in Chinese-English bilinguals.

1.8.1 A Brief Introduction to the Chinese Language

Mandarin Chinese (henceforth Chinese, unless indicated otherwise) differs from English in a number of linguistic aspects (for more detailed introduction to Chinese, see DeFrancis, 1984; Huang, Li, & Simpson, 2014). First and foremost, the writing systems of Chinese and English are very different. English is an alphabetic language that uses letters as the basic writing unit, whereas Chinese is a meaning-based logography using characters as the basic writing unit. It is worth emphasising that the role of Chinese characters does not equal the role of letters (or bigrams, e.g., *th*) in English because Chinese characters have meanings and they are not part of a sublexical level (unlike letters in English). However, whether characters should be considered as words and whether Chinese ever has the words corresponding to words of alphabetic languages it is an controversial issue (for more discussion, see Chapter 11 in DeFrancis, 1984; Myers, 2006). In other words, the notion of characters and words is neither completely inclusive nor mutually exclusive in Chinese. This notorious confusion issue is probably caused by the fact that written Chinese does not use spaces between words (e.g., a Chinses sentence 此时无声胜有声。[Silence speaks louder than words.]). Interestingly, Chinese has dictionaries for both 字 [characters] and 词 [more-than-one-character words], known as 字典 and 词典 respectively. Despite of all the confusing facts, the general consensus is that the majority of Chinese words consist of more than one character (Cai & Brysbaert, 2010; Tse et al., 2016). An analysis that I conducted with the 99,121 Chinese words in the Chinese subtitle corpus (SUBTLEX-CH, Cai & Brysbaert, 2010) revealed that 5.4% (5321) of the words consist of one character, 46.2% (45,871) of the words contain two characters, 24.7% (24,716) have three characters, 12.0% (11,972) are four-character words and 11.4% (11,341) are multiple-character words (Figure 12). However, it is important to note that some of "the one character word" in SUBTLEX-CH are actually bound morphemes (e.g., character 们 similarly to

the plural morpheme "-s" in English). To avoid confusion, the term *character* and *two-character words* are used when discussing relevant studies.[3]

Figure 12. Overview of number of characters in Chinese words based on SUBTLEX-CH (Cai & Brysbaert, 2010)



Chinese characters are generally regarded as a morpheme. For example, the character 我, which means "me/I", is a free morpheme so it can be considered as a word by itself. The character 我 can be combined with a bound morpheme 们 (similarly to the plural morpheme "-s" in English) to form the two-character word 我们, which means "we/us". Also, it can combine with another word 自己 [self] to create a three-character word 我自己, which means "myself". From these examples, it is clear that morpheme information is salient in Chinese because the boundary of morphemes is marked by the boundary between characters. An interesting fact about Chinese characters is that many Chinese characters can appear in more than one Chinese word. For example, the character 事 can occurs in 事业 [career], 事情 [thing], 事件 [incident], 事实 [fact], 事故 [accident], 事迹 [deeds], 事物 [object], 事务 [routines, same pronunciation as 事物], 故事 [story], 公事 [public business], 私事 [personal issue, the antonym of 公事], 失事 [to have a car/plane crash], 时事 [current affairs], 心事 [something haunting your

[3] Sometimes more-than-one-character words are referred to as compound words in the literature. This term is confusing because compound words in English consist of individual words (free morphemes), which is not the case in Chinese.

mind], 小事 [a trivial matter], 大事 [a big event], 相安无事 [to live in peace with each other], as well as in 事由，事宜，事项, which may be difficult to translated into English. It is important to note that the above examples are incomplete because there are many more Chinese words containing the character 事. This phenomenon is very common in the Chinese language, which is illustrated by the fact that the 99,121 Chinese words in the Chinese subtitle corpus (SUBTLEX-CH, Cai & Brysbaert, 2010) contain only 5,936 unique Chinese characters.

Orthographically, Chinese characters consist of spatially marked combinations of strokes. Broadly speaking, simplified Chinese characters now used in Mainland China can be categorized into two principal types, namely simple characters and compound characters. Simple characters cannot be decomposed into sub-components (e.g., 五 /Wu3/, meaning "five"), whereas compound characters consist of sub-components (e.g., 理 /Li3/ meaning "reason" can be decomposed into 王 and 里). It is important to note that this is a very simplified categorization. According to Shuowen Jiezi (Xu, 1985) compiled by Shen Xu during Han dynasty (206 BC–220 AD), there are six types of Chinese characters (see Table 5), which is known as 六书 (/Liu6Shu1/, meaning "six principals of characters formation"). This is a well-known and authoritative classification based on etymological and structural information of Chinese characters. However, there is some debate about whether derivative characters and phonetic borrowing characters exist in modern Chinese (Gu, 2006). Despite this controversy, it is generally accepted that most Chinese characters belong to compound characters and the majority of those compound characters consist of a semantic and phonetic component (Gu, 2006). For example, the character 湖 /Hu2/ [lake] consists of a semantic component (形旁 /Xing2Pang2/) 氵 which means "water", and a phonetic

Table 5. Six types of Chinese characters according to Shuowen Jiezi (Xu, 1985)

	Examples
象形 /Xianga Xing2/ pictograms visualising the shape	山 /Shan1/ meaning "mountain"
指事 /Zhi3 Shi4/ ideograms representing an idea	上 /Shang4/ meaning "up" 下 /Xia4/ meaning "down"
会意 /Hui4 Yi4/ a combination of two or more pictograms or ideograms to assemble meanings	信 [letter] consists of 亻 meaning "people" and 言 /Yan2/ meaning "words".
形声 /Xing2 Sheng1/ semantic-phonetic compound characters	忘 /Wang4/ [forget] consists of a semantic component 心 /Xin1/ meaning "heart" and a phonetic component 亡 /Wang2/ meaning "die"
转注 /Zhuan3 Zhu4/ derivative characters	考 /Kao3/ and 老 /Lao3/ are a pair of derivative characters, with similar orthography, pronunciations and originally both meaning "aged".
假借 /Jia3 Jie4/ phonetic borrowing characters	我 /Wo3/ [I/me] is created by borrowing the character from its homophone, which originally meant weapons.

component (声旁 /Sheng1Pang2/) 胡 which pronounces as /Hu2/.[4] In this example, the radical (部首 /Bu4Shou3/ could be understood as "header of the components") of 湖 is 氵 but not 胡. According to the Contemporary Chinese Dictionary (Institute of Linguistic from Chinese Academy of Social Science, 2005), there are 201 types of

[4] In the literature, the semantic or phonetic component is sometimes referred to as semantic or phonetic radicals. The word *radical* in semantic/phonetic radicals could be understood as 部件 /Bu4Jian4/ which is not the same as 部首 /Bu4Shou3/. Because I used the word *radical* to refer to 部首 in Chinese, I would like to use the term *component* to refer to 部件 in order to avoid confusion.

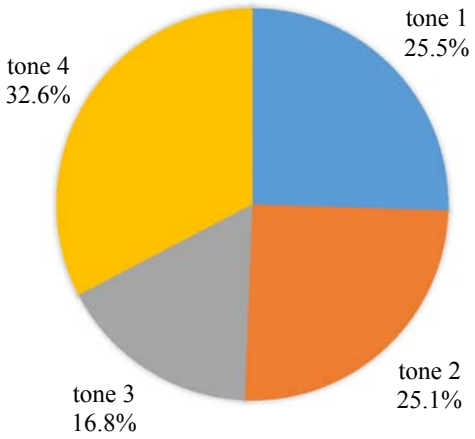
radicals in simplified Chinese. It is crucial to point out that all the Chinese dictionaries provide a radical index for readers to look up Chinese characters or words. This radical index, which lists all the radicals based on the number of strokes in the ascending order, is very useful when readers do not know the pronunciation of a character. Therefore, correctly identifying the radical of the given character is an important part of learning Chinese.

Phonologically, each Chinese character corresponds to a full syllable, which consists of segments (consonants and vowels) and a lexical tone (for a complete introduction to Chinese phonology, see Duanmu, 2006; Duanmu, 2007; Lin, 2007). Mandarin Chinese, unlike English, is a tone language that uses four tones for meaning discrimination. For example, the segment *wu* (written in Pinyin, which is a Romanisation of the Chinese pronunciation) with a high level tone (tone 1, denoted as wu1 or wū), a rising tone (tone 2, denoted as wu2 or wú), a falling-rising tone (tone 3, denoted as wu3 or wǔ), and a falling tone (tone 4, denoted as wu4 or wù) can mean "house" [屋], "nothing" [无], " five" [五], and "fog" [雾] respectively. As illustrated in these examples, tone could be marked with numbers or diacritics in Pinyin, whereas tone is not marked in Chinese characters. Based on my analysis of the 5,910 Chinese characters in SUBTLEX-CH (Cai & Brysbaert, 2010), 25.5% (1506) of characters have a high level tone (tone 1), 25.2% (1485) a rising tone (tone 2), 16.8% (990) a falling-rising tone (tone 3) and 32.7% (1929) a falling tone (tone 4) (see Figure 13).[5] However, this estimation may not reflect the tonal patterns occurring in actual language use for two reasons. First, tone sandhi occurs in Chinese when two characters in a sequence both have the falling-rising tone (tone 3). For example, the first character in the two-character word 总统 [president] changes from tone 3 to tone 2, so the word will be pronounced as /Zong2Tong3/ instead of /Zong3Tong3/. Second, my analysis only extracted one pronunciation for each character, but some Chinese characters have more than one pronunciation. For example,

[5] Pinyin and tone information were obtained from http://www.sinoca.com/web/tool/zhushipinyinshengdiao.html. Twenty-six characters that have a neutral tone (often occurring in an unstressed syllable with no clear rules) in the actual language were excluded from the calculation of the frequencies of each tonal syllables.

the character 会 pronounces as /Kuai4/ in 会计 [accountant] and /Hui4/ in 会议 [conference].

Figure 13. Distribution of tones in Chinese charaters based on SUBTLEX-CH (Cai & Brysbaert, 2010)

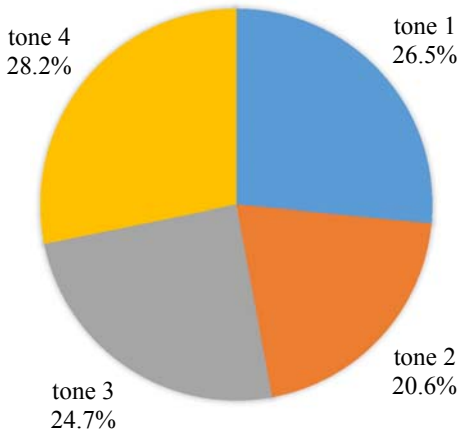


There are 404 different syllables in Chinese (Duanmu, 2007), or approximately 1200 types of syllables when combined with tones.[6] Based on Duanmu's (2006) analysis of tone distribution in 1255 syllables, 26.9% (337) of syllables have a high level tone (tone 1), 20.3% (255) a rising tone (tone 2), 25.2% (316) a falling-rising tone (tone 3) and 27.6% (347) a falling tone (tone 4). The same pattern was also found in my analysis of SUBTLEX-CH (Cai & Brysbaert, 2010), with 26.5% (313) of syllables having a tone 1, 20.6% (243) having a tone 2, 24.7% (292) a falling-rising having a tone 3 and 28.2% (333) having a tone 4 (see Figure 14). It is important to note that these approximate 1200 types of tonal syllables correspond to at least 5,936 unique Chinese characters (Cai & Brysbaert, 2010). Therefore, it is not surprising that in Chinese there are many characters that share the same pronunciations with other characters, and therefore these

[6] According to DeFrancis (1984) and Duanmu (2006), the exact number is 1,277 and 1,255 (excluding syllables with a neutral tone) respectively. An analysis of the 5,840 Chinese characters in SUBTLEX-CH (Cai & Brysbaert, 2010) revealed 1,201 types of tonal syllables including 20 syllables with a neutral tone .

could be considered as homophones, e.g., the homophones of the character 事 (/Shi4/) include 是, 式, 试, 轼, 弒, 室, 市, 视, 似, 氏, 士, 世, 示, 势, 仕, 释, 逝, 誓, 侍, 饰, 适, 嗜, 柿, 噬, 恃, 峙, 谥, 昰, 筮 (Institute of Linguistic from Chinese Academy of Social Science, 2005). As illustrated in these examples, Chinese is considered to have a deep orthography (Frost, Katz, & Bentin, 1987) because the spelling-sound correspondences are opaque. Compared to English, which has approximately 15,831 syllable types (Barker, 2016), Chinese has indeed much fewer syllable types. One possible explanation of the prevalence of homophony lies in the structure of Chinese syllables. Unlike English, Chinese words do not contain consonant clusters and the coda position of the syllable only allows two nasal consonants (i.e., /n/ and /ŋ/). Based on the full lists of syllables provided by Duanmu (2007), 57.2% (231) are open syllables (e.g., /mi/), and 42.8% (173) are closed syllables (e.g., /miŋ/). In terms of phonemes, Chinese has 34 phonemes, including 12 vowels and 22 consonants (Zhao & Li, 2009). The phonemes of Chinese are very different from those of English. Based on the lists of Chinese phonemes provided by Zhao and Li (2009), Chinese and English only share 9 consonants (i.e., /p/, /t/, /k/, /s/, /f/, /l/, /m/, /n/, /ŋ/) and 6 vowels (i.e., /a/, /e/, /o/, /ə/, /u/, /i/).

Figure 14. Distribution of tones in 1,181 syllables based on SUBTLEX-CH (Cai & Brysbaert, 2010)



Because there are no phonemes-grapheme correspondences in Chinese characters, a crucial question is whether the smallest unit of phonological encoding in Chinese production is a segment or a syllable. Up till now, however, there is still no consensus

in the literature. Early behavioural studies suggested that Chinese sound is syllable-sized (Chen & Chen, 2007; O'Seaghdha, Chen, & Chen, 2010), whereas recent neuroimaging studies revealed that Chinese sound is segment-sized (Qu, Damian, & Kazanina, 2012; Yu, Mo, Li, & Mo, 2015; Yu, Mo, & Mo, 2014).The inconsistent findings between behavioural and neuroimaging studies are still debated in the literature (O'Seaghdha, Chen, & Chen, 2013; Qu, Damian, & Kazanina, 2013).

From the examples above, it is clear that Chinese characters have an opaque phoneme-grapheme correspondence. However, as previously noted, the pronunciation of Chinese characters can be transcribed into an alphabetic system, known as Pinyin, which has a transparent phoneme-grapheme correspondence. In Mainland China, all children in grade one in primary schools learn to read and write Pinyin before learning Chinese characters. As far as I am aware, both Pinyin and characters appear in the textbooks of primary schools until grade four (see example below of the sentence "Man proposes, God disposes").

móu shì zài rén chéng shì zài tiān
谋 事 在 人, 成 事 在 天。

Although skilled adult Chinese readers seldom read or write Pinyin, Pinyin is still a part of their daily life. For example, native Chinese speakers often rely on Pinyin when looking up words or characters in Chinese dictionaries. Furthermore, most Chinese speakers use Pinyin to type Chinese characters on the computer or mobile phone. However, people do not read Pinyin in daily life. Therefore, compared to native English speakers, native Chinese speakers are less sensitive in detecting letters (Wong, Qu, McGugin, & Gauthier, 2011) and proficient Chinese-English bilinguals are not good at processing the mapping of grapheme (letter) and phoneme (sound) in reading English words (Yang, Wang, Feng, Li, & Tao, 2016).

To conclude, as discussed above, Chinese and English differ dramatically in terms of orthography and phonology. These cross-linguistic differences have led to a growing interest in Chinese language processing in an attempt to explore the similarities and differences of the cognitive and neural mechanisms in reading Chinese and English. So far, the findings from fMRI studies suggest that similar brain regions are active during Chinese and English word reading (Rueckl et al., 2015; Tan et al., 2001). However, cross-language dissimilarity in active brain regions seem to be more pronounced in

story reading than isolated word reading (Wang et al., 2015b). Furthermore, eye-tracking studies also suggest there is a universal and language-specific pattern in eye-movement during Chinese and English sentence reading (for a review, see Tsang & Chen, 2012). In addition, a substantial number of behavioural and electrophysiological studies on Chinese reading mainly investigated two issues, 1) the role of characters in the recognition of two-character Chinese words; 2) the role of phonology in Chinese single character and two-character words. In the following sections, a general overview of these two research lines is provided.

1.8.2 The Role of Characters in the Recognition of Two-character Chinese Words

As mentioned above, the majority of Chinese words consist of two-character words. To understand the how people recognize two-character words, researchers often used the character repetition paradigm to explore the interaction of semantics, orthography and phonology. Importantly, because Chinese characters are generally considered to be morphemes, character repetition can be considered as morpheme repetition as well. Therefore, studies using the character repetition paradigm also contribute to the investigation of morphological processing. For example, prime-target pairs like **信**封 [envelope] - **信**件 [letter] focus on morpho-semantic priming because these word pairs with an identical character/morpheme are semantically related. In contrast, prime-target pairs like **月**薪 [salary] - **月**亮 [moon] focus on morpho-orthographic priming because they are not related in meaning, similarly to the English words ***depart****ment* - ***depart****ure*.

As far as I am aware, Zhou, Marslen-Wilson, Taft, and Shu (1999a) were the first to systematically investigate morpho-semantic priming (e.g., 华丽 [magnificent] - 华贵 [luxurious]) and morpho-orthographic priming (e.g., 华侨 [overseas Chinese] - 华贵 [luxurious]) in Chinese word recognition. In addition to manipulating the semantic relationship between primes and targets, they also manipulated the position of the repeated characters (i.e., the first character was repeated **华**丽 [magnificent] - **华**贵 [luxurious], the second character was repeated 轻**易** [easily]- 简**易** [simple] or the repeated character was in different positions 宽**容** [lenient]- **容**忍 [tolerate]). They used a masked priming paradigm in which the prime was presented for 57 ms. The results revealed that participants' responses in the lexical decision task were facilitated in both

morpho-semantic priming and morpho-orthographic priming conditions regardless of the position of the repeated characters. Importantly, morpho-semantic priming was always significantly stronger than morpho-orthographic priming. Another masked priming study with a lexical decision task by Peng, Liu, and Wang (1999) focused on the impact of semantic transparency in morpho-orthographic priming. Their results revealed that morpho-orthographic priming only occurred in prime words that were transparent (i.e., the meanings of the morphemes/characters are related to the meaning of the Chinese words, e.g., the two characters in 安宁 [quiet] mean "quite" and "peace" respectively) but not in prime words that were opaque (i.e., the meanings of the morphemes/characters are not related to the meaning of the Chinese words, e.g., the two characters in 快活 [happy] mean "quick" and "live" respectively). Also, using a masked priming technique with a lexical decision task but with 100-ms prime duration, Wang and Peng (2000) found that morpho-orthographic priming could be observed in opaque targets words when preceded by single-character primes (e.g., 小 [small] - 小心 [careful]).

In addition to behavioural investigations, a handful of electrophysiological studies also used the character repetition paradigm to investigate Chinese word recognition. In a semantic relatedness judgment task, Thierry and Wu (2007) employed a 2×2 factorial design to dissociate the impact of semantic and character repetition, e.g., +semantic +character repetition: 邮政 [post] - 邮件 [mail], +semantic -character repetition: 妻子 [wife] - 丈夫 [husband], -semantic +character repetition: 火车 [train] -火腿 [ham], -semantic -character repetition: 苹果 [apple] - 桌子 [table]. Critically, the main effect of the character repetition priming reduced the P200 and N400 amplitudes. However, no interactions were found, which suggests that the impact of character repetition is independent of semantic relationship.

In another recent ERP study, Jia, Wang, Zhang, and Zhang (2013) used a cross-trial priming paradigm to investigate morpho-semantic priming (e.g., 问句 [questioning sentence] - 问号 [question mark]) and morpho-orthographic priming effects (e.g., 问卷 [questionnaire] - 问号 [question mark]). They found that both priming conditions reduced the N200 and N400 component, which index Chinese orthographic processing

(Zhang et al., 2012) and semantic integration processing (Kutas & Federmeier, 2011) respectively. Interestingly, the morpho-semantic priming condition further reduced the N400 component but not the N200 component compared to the morpho-orthographic priming condition. In light of these electrophysiological patterns, Jia et al. (2013) argued that Chinese word recognition is a serial process with orthographic processing proceeding semantic processing.

However, it is important to note that the repeated characters in the studies mentioned above had the same pronunciation. As a result, the character repetition priming effect may be also attributed to phonology. To remove the potential impact from phonology, Zhou et al. (1999a) used homographs to create the character repetition priming in a lexical decision study. For example, in the word pair 重复 [repetition], /Chong2Fu4/, -重量 [weight], /Zhong4Laing4/, the repeated character 重 has different pronunciations. Zhou et al. (1999a) failed to observe a priming effect in a masked priming experiment. However, an inhibition effect was obtained when an unmasked prime was presented for 200 ms. In an ERP study by Wu and Thierry (2010), the character repetition priming with homographs was combined with the semantic relatedness judgment task, which reduced the ERP amplitude from 250 to 660 ms.

To summarize, several studies in the literature have investigated the role of characters in Chinese visual word recognition using the character repetition paradigm. Overall, the most important finding is that the character repetition effect is modulated by the semantic relationship of prime and target words, revealing the top-down impact of words on their constituent characters. In addition to studies with Mandarin speakers, similar investigations have also been conducted with Cantonese speakers (e.g., Liu & McBride-Chang, 2010; Tsang & Chen, 2010, 2013a, 2013b, 2014; Wong, Wu, & Chen, 2014). However, it is obvious that more studies are needed to fully understand the role of Chinese characters in two-character visual word recognition. First and foremost, the studies mentioned above often failed to consider the impact of the neighbour density (i.e., the number of words sharing the same characters). This limitation is probably due to the fact that no freely available database exists that provides neighbourhood information of Chinese words. It is therefore unclear whether the lexical competition of neighbours would affect the character repetition effect.

1.8.3 The Role of Phonology in Chinese Reading

1.8.3.1 Phonological Activation in Chinese Reading

It has been well established that when silently reading an alphabetic language, the phonological information of words is activated (for a review, see Frost, 1998). However, as previously noted, Chinese is a logographic language which does not have phonemes-grapheme correspondences. Therefore, it is important to investigate the role of phonology in Chinese reading (e.g., whether phonological activation is a route to semantic activation or it occurs after semantic activation, to what extent Chinese phonological activation is automatic). The notion of reading in studies on Chinese phonological activation has been defined in a broad sense, which includes silent reading as well as reading aloud (e.g., naming). The investigation into phonological activation during Chinese reading has generated a large number of studies in the literature (e.g., Chen & Shu, 2001; Liu et al., 2003; Perfetti, Liu, & Tan, 2005; Perfetti & Tan, 1998; Perfetti & Zhang, 1991, 1995; Shen & Forster, 1999; Spinks, Liu, Perfetti, & Tan, 2000; Tan & Perfetti, 1997; Wang, Mecklinger, Hofmann, & Weng, 2010; Xu, Pollatsek, & Potter, 1999; Zhou & Marslen-Wilson, 1999a, 1999b, 2000; Zhou, Shu, Bi, & Shi, 1999b). However, researchers are far from reaching an agreement about the role of Chinese phonological activation in Chinese reading.

In an unmasked priming study, Perfetti and Tan (1998) systematically manipulated the SOA (43/57/85 ms) to investigate the time course of orthographic, phonological and sematic activation in Chinese character naming. In this seminal study, the priming effect occurred first in the orthographic condition (e.g., 司 - 可, /Si1/ - /Ke3/) with an SOA of 43 ms, whereas the phonological priming effect (e.g., 其 - 齐, /Qi2/ - /Qi2/) emerged at an SOA of 57 ms, followed by the semantic priming effect (e.g., 究 - 查, /Jiu1/ - /Cha2/, meaning “check’) appearing at an SOA of 85 ms. Based on these results, Perfetti and Tan (1998) concluded that in Chinese character recognition semantic access happens via orthography to phonology and then to semantics. However, Chen and Shu (2001) failed to replicate Perfetti and Tan’s (1998) findings with identical stimuli and the same procedure. The main difference is that in Chen and Shu (2001) the semantic priming effect started earlier at an SOA of 43 ms, followed by the phonological priming effect at an SOA of 57 ms. Therefore, Chen and Shu (2001) argued that the semantics of Chinese characters are activated via a direct route from orthography to semantics and

the phonological activation occurs after semantic activation.

However, the speed of the phonological activation in Chinese character recognition seems to be task-dependent. In an unmasked priming study, Zhou and Marslen-Wilson (2000) revealed that the phonological priming effect (e.g., 独 [alone] - 读 [read], /Du2/ -/Du2/) appeared at an SOA of 57 ms and 200 ms in a naming task. However, in a lexical decision task, phonological priming only occurred with an SOA of 200 ms. These findings were replicated in a masked priming study by Shen and Forster (1999). Using a prime duration of 50 ms, phonological priming was found in naming but only with simple characters (i,e., characters that cannot be decomposed into sub-components, e.g., 五 [five] - 午 [noon], /Wu3/ - /Wu3/), and not in compound characters (i,e., characters consist of sub-components, e.g., in this pair 礼[gift] - 理[reason], /Li3/ - /Li3/, characters can be decomposed into 礻, 乚, and 王， 里 respectively), whereas no phonological priming effect was found in the lexical decision task.

The mediated priming paradigm has also been used to investigate phonological activation during Chinese character recognition. For example, Tan and Perfetti (1997) created a phonologically mediated priming condition by using a homophone that is semantically related to the target (e.g., 括 - 宽, /Kuo4/ - /Kuan1/, the prime 括 is the homophone of 阔 [broad], which is semantically related to the target 宽 [wide]). Critically, the homophone density (i.e., number of homophones) of primes was manipulated. With 129 ms and 250 ms SOAs, the naming latencies of the targets were facilitated by phonological-mediated primes when homophone density of the primes was low (≤5) or medium (<7 and ≤ 15), but not when homophone density of the primes was high (≥20). It is important to note that mediated priming effects were obtained because primes activated their homophones, which facilitated the recognition of the targets. Therefore, it seems to indicate that phonological activation constrains the lexical processing of Chinese characters. But this account has been challenged by Zhou and Marslen-Wilson (1999a), who failed to observe any mediated priming effects or the impact of homophone density on mediated priming effects using the identical stimuli with the same procedures (Experiment 1 in Zhou and Marslen-Wilson, 1999a). Zhou and Marslen-Wilson (1999a) further revealed that mediated priming effects could be obtained with a special group of primes, namely, compound characters whose

pronunciation is identical to one of its sub-components. For example, the character 滤 contains a character 虑 (known as the phonetic component of 滤), both of which are pronounced as /lü4/. These results reveal a complex interaction between orthography, phonology and semantics, and no predominant role of phonology in character recognition. Zhou and Marslen-Wilson (1999a) argued that Chinese characters only consist of 30% compound characters which have the same pronunciation as its phonetic components. Therefore, the idea that phonological activation constrains the lexical processing is only limited to a small number of Chinese characters.

In addition to priming studies, experiments involving the Stroop task (Stroop, 1935) have also been used to investigate automatic phonological activation with Chinese characters. In a classic Stroop task, colour words (e.g., red, green, blue) are presented in colour ink and participants are asked to name the colour of the ink or press the corresponding buttons in response to the colour of the ink. Critically, colour words presented in the incongruent colour (e.g., the word *red* presented in colour green) will lead to slower responses relative to a control condition (e.g., the word *fact* presented in colour green), which is referred to as Stroop interference. In contrast, colour words presented in the congruent colour (e.g., the word *red* presented in colour red) will speed up participants' responses relative to control condition, which is referred to as Stroop facilitation. To investigate Chinese phonological activation, homophones of colour characters were critical stimuli in the Stroop task. The logic is that if Stroop interference is found in homophones of colour characters presented in the incongruent colour (e.g., 洪, the homophone of 红 [red] presented in colour green) or Stroop facilitation is found in homophones of colour characters presented in the congruent colour (e.g., 洪, the homophone of 红 [red] presented in colour red), it would suggest that the phonology of Chinese character is activated. Using this paradigm, Guo, Peng, and Liu (2005) found Stroop interference and facilitation effects in naming latencies with colour homophones in children and skilled adult readers. Using a similar design in an ERP study, Wang et al. (2010) reported that homophones of colour characters presented in the incongruent colour (e.g., 洪, the homophone of 红 [red] presented in colour green) modulated the ERP amplitude compared to homophones of colour characters presented in the congruent colour. However, the modulation of ERP amplitude was found in the late time window between 600 and 800 ms but not in the early time window between 400

and 500 ms, suggesting that homophones of colour characters cause response conflicts in later stages of processing rather than semantic conflicts in the early processing. Therefore, Wang et al. (2010) concluded that phonological activation of Chinese characters occurs after semantic activation.

Another paradigm has been used to explore phonological activation of Chinese characters is the interference paradigm. For example, participants are presented a pair of characters in succession and decide whether the presented pair is related in meaning or not (i.e., the semantic relatedness judgment task). Critically, homophone pairs (e.g., 事 [matter] - 视 [see], /Shi4/ - /Shi4/) are not related in meanings and belongs to the "No" trials. If homophone pairs resulted in longer response times compared to semantically and phonologically unrelated controls (e.g., 清 [clear] - 视[see], /Qing1/ - /Shi4/), this phonological interference is taken as the evidence of phonological activation and its consequent impact on semantic processing. In a similar vein, the semantic interference is created by using semantic related pairs (e.g., 看 [look at] - 视 [see], /Kan4/ - /Shi4/) when participants are asked to decide whether a pair of characters has the same pronunciation (i.e., the homophone judgment task). By comparing the phonological and semantic interference at different SOAs (i.e., the interval between the onsets of two characters presented in succession), it may shed light on the time course of phonological and semantic activation. Following this approach, Perfetti and Zhang (1995) used homophone pairs in a semantic relatedness judgment task to create phonological interference and semantic related pairs in a homophone judgment task to create semantic interference. The phonological interference occurred with an SOA of 90 ms, whereas the semantic interference effect started at the SOA of 140 ms. In light of these findings, Perfetti and Zhang (1995) concluded that phonological activation precedes semantic activation and phonological activation is an essential constituent of character recognition. Using the same interference paradigm but with a SOA of 640 ms, a follow-up study by Liu et al. (2003) found a phonological interference effect (i.e., a N400 reduction) but not a semantic interference effect. Using a similar paradigm (SOA: 500 ms, prime duration 140 ms) with the semantic relatedness judgment task, the phonological interference effect was also observed in a recent study by Kong et al. (2010). But instead of a N400 reduction, a larger P200 amplitude was observed when a Chinese character were preceded by their homophone. Furthermore, a behavioral study

by Zhang, Perfetti, and Yang (1999) and ERP study by Zhang, Zhang, and Kong (2009) further indicated that the phonological interference effect was modulated by character frequency.

The studies discussed above involved Chinese single characters, which may not represent the whole picture of Chinese word recognition. First of all, it is difficult to tell whether or not all the characters used in the experiments could be considered as words. In fact, for some it is difficult to provide an English translation, and this is why I did not provide the English translation for some examples of Chinese characters in the thesis. Secondly, there are much fewer one-character than two-character Chinese words (e.g., 5,321 vs. 45,871 in SUBTLEX-CH). In addition, it is common for Chinese characters to have a large number of homophones, whereas most two-character Chinese words do not have a homophone (e.g., the word 美丽 [pretty], pronounced as /Mei3Li4/, has no homophones) and some have only one homophone (e.g., the word 警官 [police] and 景观 [scenery] are homophones with the same pronunciation of /Jing3Guan1/. Finally, single Chinese characters are more likely to be semantically ambiguous than two-character Chinese words. For example, the Chinese character 安 has 11 different meanings according to the Contemporary Chinese Dictionary (Institute of Linguistic from Chinese Academy of Social Science, 2005). Furthermore, this character appears in the words 安全 [safety], 安静 [quiet], 安装 [install], to name just a few. Therefore, when reading a single character like 安, ambiguity might occur and this could impact the character recognition process. In fact, Ma, Wang, and Li (2015) found that the reaction times for single characters in a semantic categorization task were slower than for two-character words, indicating that processing one-character words indeed requires more cognitive resources than two-character words. Taken together, it might be more informative to use two-character words to investigate phonological activation because two-character words are the most common Chinese words, and they have less potential confounds (e.g., homophone density, semantic ambiguity).

Interestingly, only a handful number of studies in the literature have focused on the role of phonology in the recognition of two-character Chinese words. In an early study using homophonic word pairs like 捷径 [shortcut] - 洁净 [clean] , /Jie2Jing4/ - /Jie2Jing4/, Zhou et al. (1999b) did not find any behavioural priming effects in masked (prime

duration: 57 ms) and unmasked (prime duration: 100 ms) priming lexical decisions experiments. However, the homophone priming effect was found in their naming experiment (unmasked prime duration: 100 ms). Zhou and Marslen-Wilson (2000) also failed to observe homophone priming in lexical decision experiments with an SOA of 57 ms and 200 ms. In contrast to these findings, Ma et al. (2015) found homophone priming in a masked priming experiment with an SOA of 47 ms but in a semantic categorization task. These results seem to suggest that homophonic priming in two-character words is a task-dependent effect, which is similar to the patterns observed in studies using single characters.

In addition to homophones, Zhou et al. (1999a) used two-character word pairs involving a single character that had the same pronunciation in the prime and target to explore the activation of phonology in two-character words. For example, the first characters in the prime-target pair 滑翔 [glide] - 华贵 [luxurious] are pronounced as /Hua2/. The results of their lexical decision experiment indicated no evidence of phonological priming with both masked and unmasked primes presented for 57 ms. Zhou et al. (1999a) presented primes up to 100 ms, but again no priming effect was obtained. However, in a semantic relatedness judgement experiment with a 500-ms prime duration, Wu and Thierry (2010) found effects of phonological repetition. Their experiment also involved two-character words sharing the same pronunciation (e.g., 经验 [experience] - 惊讶 [surprise], /**Jing1**Yan4/ - /**Jing1**Ya/) and the phonological repetition resulted in slower responses in reaction times and a reduced N400 amplitude in the ERP data compared to the control condition. Although these results provide evidence for phonological activation, it is important to point out that the slower behavioural responses indicated an interference effect, whereas a N400 reduction suggests facilitatory phonological priming (Grainger, Kiyonaga, & Holcomb, 2006; Rugg & Barrett, 1987). This contrast between behavioural and ERP data seems to suggest a dissociation between the conflict in responses and the facilitation in lexical processing as a result of phonological priming.

To summarize, a considerable number of studies have explored phonological activation in Chinese character/word reading. Although there is currently no consensus in the field, phonological priming seems to be more salient in semantic tasks (i.e., semantic relatedness/categorization judgement task) and tasks involving phonological encoding

(e.g., naming), and more robust when long SOAs are used.

1.8.3.2 Segmental and Tonal Activation in Chinese Character Recognition

Although there are numerous studies in the literature that have investigated the activation of phonology during Chinese character/word reading, the role of segmental and tonal information is less explored. Taft and Chen (1992) were, as far as I am aware, the first to investigate the role of segment and tone in Chinese reading. In a homophone judgements task, they found that pairs of characters that are different-tone homophones (e.g., 曲 - 去, /Qu3/ - /Qu4/) slowed down reaction times and increased error rates relative to control pairs (e.g., 年 - 去, /Nian2/ - /Qu4/). A similar but less robust inhibition effect was also found with pairs of characters that only differed in vowels (e.g., 气 - 去, /Qi4/ - /Qu4/).

To further explore the role of segment and tone in Chinese reading, Spinks et al. (2000) used a Stroop task in which native Chinese speakers were asked to name the presentation colour of the word. In addition to Chinese colour characters (e.g., 红, /Hong2/, [red]), the experiment included homophones of the colour characters (+*Segment* +*Tone*, e.g., 洪 [flood], /**Hong2**/) and different-tone homophone (+*Segment* -*Tone* e.g., 轰 [boom], /**Hong**1/). In the congruent conditions (e.g., 红, 洪, 轰 presented in red), both kinds of homophones showed a significant facilitation effect with larger facilitation in the identical-tone homophone condition (+*Segment* +*Tone*). In the incongruent conditions (e.g., 红, 洪, 轰 presented in green), only identical-tone homophones (+*Segment* +*Tone*) resulted in a reliable interference effect. These findings indicate that the tonal information constraints lexical access during Chinese reading. However, since this study did not include a -*Segment* +*Tone* condition, the direct comparison between segmental and tonal information is not possible. To investigate this, Li, Lin, Wang, and Jiang (2013) used the design of Spinks et al. (2000) and included a -*Segment* +*Tone* condition (e.g., 瓶 [bottle], /Ping2/). In the congruent conditions (e.g., 红, 洪, 轰, 瓶 presented in red), +*Segment* +*Tone*, +*Segment* -*Tone*, -*Segment* +*Tone* all showed a facilitation effect. However, whereas the magnitude of the facilitation effect in +*Segment* +*Tone* and +*Segment* -*Tone* was comparable, the facilitation effect of +*Segment* -*Tone* was larger than that of -*Segment* +*Tone*. The

authors concluded that even though both segment and tone constrain lexical access, segmental information plays a more important role.

Taken together, the findings of these three studies suggest that segmental and tonal information is activated separately (Li et al., 2013; Spinks et al., 2000; Taft & Chen, 1992). In contrast, a recent priming study suggested that segment and tone are combined in one integrated encoding unit (Wang, Li, & Lin, 2015a). This study investigated the segmental and tonal activation of Chinese characters in a naming task. Independent of prime duration (i.e., 57ms, 100 ms, 200 ms), the naming latencies of target words were faster only when they were preceded by their same-tone homophones (e.g., 连 - 怜, /Lian2/ - /Lian2/, +*Segment* +*Tone*), but not by their different-tone homophones (e.g., 练 - 怜, /Lian4/ - /Lian2/, +*Segment* -*Tone*) or same-tone neighbours (e.g., 成 - 怜, /Cheng2/ - /Lian2/, -*Segment* +*Tone*).

In addition to the behavioural investigation presented above, two ERP studies used a visual oddball paradigm to investigate tonal activation during Chinese characters reading (Wang, Liu, Wu, & Wang, 2013a; Wang, Wu, Liu, & Wang, 2013b). In a version of the visual oddball paradigm, participants are visually presented with sequences of repetitive stimuli (e.g., squares or circles in different colours) and their task was to press the button in response to target stimuli (e.g., red stimuli irrespective of shape). Critically, the frequency of non-target stimuli (e.g., stimuli that are not in red) was manipulated, with the more frequent one referred to the standard stimuli (e.g., 70% of squares) and the less frequent one as the deviant stimuli (e.g., 30% of circles). The difference of ERP responses to the standard stimuli and the deviant stimuli is known as the visual mismatch negativity (vMMN). Because detecting the standard and deviant stimuli is task-irrelevant, the presence of the vMMN is the evidence of automatic and pre-attentive processing of the feature that differentiates the standard from the deviant stimuli (e.g., the shape of the stimuli). In a similar vein, two ERP studies used same-tone characters (e.g., 忆, 议, 译, 抑, all pronounces as /Yi4/) as the standard stimuli and the different-tone characters (e.g., 依, /Yi1/) as the deviant stimuli. Both studies observed a vMMN, suggesting that tonal information of the visual characters was rapidly activated (Wang et al., 2013a; Wang et al., 2013b).

In summary, some studies have explored the segmental and tonal activation in reading Chinese characters. However, the empirical findings are far from conclusive. In addition, although computational models for Chinese characters have generally coded segment and tone in the phonology level separately (Perfetti et al., 2005; Zhao & Li, 2009, 2013), these models did not specify the time course and the exact mechanism of segmental and tonal activation when Chinese phonology is activated. For example, as previously mentioned (Section 1.7.6), the DevLex-II model (Zhao & Li, 2010, 2013) implemented the slot-based phonological representation in which each character (corresponding to a syllable) is represented by five slots of phonemic units in segment and one slot of tonal information (Zhao & Li, 2009). Because the DevLex-II model mainly focused on the bilingual mental representation, it did not make specific predictions on the relationship between tone and segment during phonological activation. Similarly, the lexical constituency model developed by Perfetti and colleagues (Perfetti et al., 2005; Perfetti & Tan, 1999) represented each character in terms of the onset, vowel and tone. The focus of the lexical constituency model is to simulate orthographic, phonological and semantic activation during visual Chinese character recognition (Perfetti et al., 2005). Therefore, it also does not tell us how segmental and tonal information is activated within the phonology level.

1.8.3.3 Summary

Based on the available data on Chinese phonological activation during Chinese visual word processing it is clear that phonological activation (segmental and tonal activation) during Chinese reading is supported by empirical evidence and the supporting evidence seems to be task-dependent. In contrast, it seems that the activation of Chinese phonology is strong during English word reading in Chinese-English bilinguals, as shown by studies using the hidden repetition priming paradigm (Section 1.4). To understand this contrast, it is necessary to investigate whether studies using other paradigms also support Chinese phonological activation when Chinese-English bilinguals process English. Therefore, I will present in the next section studies that investigated cross-language phonological activation between English and Chinese.

1.8.4 Cross-language Phonological Activation

As mentioned in Section 1.8.1, the phonology of Chinese and English differ in many

aspects (e.g., phonemes used in the languages, the structure of the syllables, a tone language vs. a non-tone language). However, several Chinese characters and English words have similar pronunciation, and therefore they could be considered as cross-language homophones. For example, the character 卖 (IPA: /mai/, Pinyin: /Mai4/) which means "sell" is pronounced very similar to the English word *my*. In a masked priming study using both a naming task and a lexical decision task, Zhou, Chen, Yang, and Dunlap (2010) found that Chinese-English bilinguals responded faster to Chinese or English words when they were primed by cross-language homophones. Interestingly, this cross-language homophone priming effects occurred in groups of bilinguals with a high and low English proficiency. These findings indicate that cross-language activation at the phonological level occurs and thus they provide support for the non-selective access assumption of bilingual word recognition (Dijksta, 2005; Dijkstra & van Heuven, 2002; van Heuven et al., 1998).

To further investigate the role of phonology in cross-language activation, cross-script cognate pairs in English and Chinese have been used in two studies (Dong & Lin, 2013; Qi, 2011). Etymologically, Chinese-English cognates are transliterated loan words, which were created to keep pronunciations as similar as possible to the original English words, such as coffee - 咖啡 (IPA: /k^ha fei/, Pinyin: /Ka1Fei1/). Despite that Chinese-English cognates, which are not very common, do not overlap in phonology to the same extent as same-script cognates, studies have found a cognate advantage over non-cognate translation pairs in a translation recognition experiment (Dong & Lin, 2013) and in a masked priming experiment with a naming task (Qi, 2011). Obviously, Chinese-English cognates do not have any orthographic overlap, therefore, the cognate advantage must be driven by phonological similarity. This supports the idea that the cross-language activation can occur at the phonological level.

Because Chinese is a tone language, the impact of tonal information during language processing in bilinguals was also investigated in a recent study. Shook and Marian (2016) asked Chinese-English bilinguals to identify the correct Chinese translations presented on a screen while spoken English words were presented at the same time. In critical trials, the pitch contour of the spoken English words were identical to their Chinese translations. For example, when the English word *tree* is pronounced in a

failing intonation, it shares the same pitch contour with its Chinese translation 树, /Shu4/. Compared to control trials (e.g., when *tree* is pronounced in a rising intonation), a facilitation effect (i.e., faster response times and earlier eye-movements) was found when the pitch contour was shared between English words and their Chinese translations. This finding suggests that Chinese-English bilinguals are highly sensitive to the pitch contour of English words and that the tonal information of Chinese translations is activated quickly.

Recently, Li, Wang, and Lin (2016) investigated the automatic activation of Chinese phonology during English naming in a Stroop paradigm. Similar to their earlier study (Li et al., 2013), they presented Chinese colour characters (e.g., 红 [red], /**Hong2/**), together with homophones of the colour characters (+*Segment* +*Tone*, e.g., 洪 [flood], /**Hong2**/), different-tone homophones (+*Segment* -*Tone* e.g., 轰 [boom], /**Hong**1/), and same-tone characters (-*Segment* +*Tone* e.g., 瓶 [bottle], /Ping**2/**) in different colours. During the experiment, Chinese-English bilinguals were asked to ignore the Chinese characters and to name the colour of the ink (in which the characters were presented) in English. In the congruent condition (e.g., 红, 洪, 轰, 瓶 presented in red), only naming latencies of the Chinese colour characters showed a facilitation effect. Furthermore, the Chinese colour characters also showed an inhibition effect in the incongruent condition (i.e., 红 in green). Li et al. (2016) concluded that there was no evidence of automatic Chinese phonological activation during English word production.

From the above-mentioned studies, only Li et al. (2016) failed to observe Chinese phonological activation. One possible explanation may lie in the stimuli involved. As already discussed earlier in this chapter (Section 1.8.1 & Section 1.8.3.1), it is controversial issue of whether single Chinese characters should be considered as words. More importantly, single Chinese characters often cause semantic ambiguity, which may have an unclear impact on phonological activation during cross-language activation. For example, as far as I know, the character 树 used in Shook and Marian (2016) is a free-morpheme and can be considered as a word, whereas the homophone of the colour character 洪 is not a free-morpheme in Chinese. In addition, the character 洪 has three different meanings according to the Contemporary Chinese Dictionary

(Institute of Linguistic from Chinese Academy of Social Science, 2005), namely "big" (e.g., two-character word 洪亮 [loud]), "flood" (e.g., two-character word 洪水 [flood]), and it can be used as a family name. Therefore, one important methodological implication for the current thesis is that it is better to use two-character Chinese words as stimuli when investigating cross-language phonological activation.

1.9 Overview of the Present Thesis

In this thesis I will present a series of behavioural and electrophysiological studies that investigated Chinese (L1) translation activation during English (L2) word reading in Chinese-English bilinguals, mainly through the use of the hidden repetition priming paradigm. As reviewed in Section 1.5, ERP studies using the *unmasked* hidden repetition priming paradigm consistently reported the activation of Chinese phonological information when Chinese-English bilinguals read their English translations equivalents (Wu & Thierry, 2010, 2012a). However, two aspects of this phonological activation need to be further explored to fully understand its underlying mechanisms. First of all, because Chinese phonology consists of both segmental (consonants and vowels) and tonal information, it remains unclear whether the hidden repetition priming in phonology is driven by segmental and/or tonal information. Another aspect that needs to be further investigated is whether Chinese phonology is the source of the hidden repetition priming effect observed in behavioural data from *masked* priming experiments (Zhang et al., 2011). Therefore, the general aim of the thesis is to specify the source of translation activation (phonology: segment and/or tone, and/or orthography) in the hidden repetition paradigm.

This thesis consists of five experimental chapters. Before the experimental investigations into translation activation, Chapter 2 reports a large-scale norming study which provides Chinese translation norms for English words. This database is crucial because it makes it possible to select carefully controlled stimuli for the experiments reported in the following chapters. In addition, Chapter 2 also investigates the relationship between translation ambiguity and word frequency, concreteness and language proficiency because previous studies suggested that translation ambiguity

interacted with these factors (Prior, MacWhinney, & Kroll, 2007; Tokowicz, Kroll, de Groot, & van Hell, 2002). Using the hidden repetition priming paradigm in unmasked and masked behavioural experiments, Chapter 3 explores whether segmental and/or tonal information of the Chinese translation is activated when Chinese-English bilinguals read English words. As discussed in Section 1.8.3, several studies on Chinese character recognition suggested that segmental and tonal information was activated separately (Li et al., 2013; Spinks et al., 2000; Taft & Chen, 1992). In addition, computational models for Chinese characters have separately coded segment and tone in the phonology level (Perfetti et al., 2005; Zhao & Li, 2009, 2013). Therefore, it is likely that there is a dissociation between segmental and tonal activation of the Chinese translation during English word reading. Chapter 4 further explores the activation of Chinese segmental and tonal information during English and Chinese word reading in the electrophysiological experiments by analysing both the event-related brain potentials and brain oscillations. Analysing the event-related brain oscillations is insightful because it can be used as a frequency-based segregation approach to investigate segmental and tonal activation (more discussions on this approach to be presented in Chapter 4). As far as I am aware, there are no papers in the literature on bilingual word recognition or on Chinese phonological activation that have analysed the event-related brain oscillations. Therefore, Chapter 4 is the first attempt to explore cross-language activation and Chinese phonological activation by analysing the event-related brain oscillations. Chapter 5 continues the behavioural investigations with the masked hidden repetition priming paradigm in an attempt to clarify the source of translation activation (phonology/orthography) observed in Zhang et al. (2011). As discussed previously, electrophysiological studies revealed (Wu & Thierry, 2010, 2012a) that hidden repetition priming effects in terms of the N400 reduction was driven by the phonology of Chinese translation. Therefore, it is likely that the behavioural masked hidden repetition priming effects observed in Zhang et al. (2011) are also driven by Chinese phonology. In Chapter 6, I will present the first meta-analysis of masked priming lexical decision experiments to assess the effect sizes of L1-L2 and L2-L1 non-cognate translation priming effects in bilinguals. As discussed in Section 1.3, studies using the masked translation priming paradigm have reported mixed findings, in particular for L2-L1 translation priming. Therefore, it is crucial to conduct a meta-analysis for the field because it can provide an answer to whether L2-L1 translation priming in the masked translation priming paradigm reported the literature is reliable

and whether the translation priming effects in both direction are modulated by other factors. Following these five experimental chapters, a general discussion is presented in Chapter 7, which considers theoretical and methodological implications of the overall findings reported in the thesis.

Chapter 2: English to Chinese Translation Norms[7]

2.1 Introduction

As discussed in Chapter 1, translation equivalents have been used extensively to investigate bilingual language processing. A potential issue with the translation priming paradigm is that translation equivalents do not always have one-to-one mappings. When a word in one language has more than one translation equivalent in another language, it is considered to be translation ambiguous. To investigate the prevalence of translation ambiguity, Tokowicz et al. (2002) asked Dutch-English bilinguals to write down the first translation that came to their mind (first translation method) of words taken from a number of published translation studies (e.g., de Groot, 1992; de Groot et al., 1994). Surprisingly, more than 25% of the 562 words in their study that received more than one correct translation were classified in the published translation studies as having an unambiguous translation. Thus, translation ambiguity is in fact more prevalent and studies often incorrectly classify words as translation unambiguous. The prevalence of translation ambiguity is not limited to Dutch and English. In a large-scale Spanish-English translation norming study, more than 50% of the words were translation-ambiguous (Prior et al., 2007). Given that translation ambiguity is common, an important question is whether translation ambiguity affects bilingual lexical access.

Recent studies have revealed that translation ambiguity does indeed influence bilingual language processing (for a review, see Tokowicz, 2014). For example, in a translation production task with English-Spanish bilinguals, Tokowicz and Kroll (2007) found that bilinguals were slower and less accurate when presented with translation-ambiguous

[7] This data of this chapter has been published in *Behavior Research Methods* as Wen, Y., and van Heuven, W. J. B. (2016). Chinese translation norms for 1429 English words. *Behavior Research Methods*. doi: 10.3758/s13428-016-0761-x. This chapter is an adapted version of the published paper because slight modifications of the text have been made to the introduction and the discussion in order to integrate it in the thesis.

words. This disadvantage for translation-ambiguous words has also been reported in word recognition tasks (Boada et al., 2013; Laxén & Lavaur, 2010). For example, when French-English bilinguals had to judge whether two words were translation equivalents, they responded faster to translation-unambiguous pairs than to translation-ambiguous pairs, and faster to words with their dominant translations than to words with subordinate translations (Laxén & Lavaur, 2010). More surprisingly, as reviewed by Tokowicz (2014), Degani, Prior, Eddington, Arêas da Luz Fontes, and Tokowicz (2013) found that bilinguals showed a disadvantage for translation-ambiguous words even in a monolingual context. In their English lexical decision task, Spanish-English bilinguals' responses were less accurate for English words with several Spanish translations than for English words with only one Spanish translation. Taken together, these studies not only suggest that translation ambiguity affects language processing but also provide further support for non-selective bilingual lexical access (e.g., Dijkstra, 2005; Dijkstra & van Heuven, 2002; van Heuven et al., 1998).

Because of the extensive use of translation equivalents in bilingual studies and the importance of controlling for translation ambiguity, the primary goal of the present study was to create a large database of English-Chinese translation norms and to make this available for other researchers. This database would directly benefit future research using Chinese-English translation equivalents to investigate bilingual language processing. As far as we are aware, there is only one database reported in the literature that includes English-Chinese translation norms (Tseng, Chang, & Tokowicz, 2014). This translation norming study not only conducted the first in-depth investigation of bilingual participants' translation errors, but also provided a list of alternative Chinese translations for 562 English words together with semantic similarity ratings for all the translation pairs. However, the norming study does not provide all information about the translations and it has some limitations. Most importantly, no information about how many participants provided each alternative translation is available for words with multiple translations (N=378). Although there are 165 translation-unambiguous pairs in their database, it is unclear whether the unique translations were generated by the majority of the participants or just by one individual participant. In addition, the language background of the bilinguals was heterogeneous because not all of the bilinguals came from Mainland China where only simplified Chinese is used.

Without a large comprehensive database of English-Chinese translation norms available in the literature, previous studies using, for example, the hidden translation-priming paradigm (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a, 2012b) repeatedly conducted translation-norming experiments with relatively large groups of bilinguals ($N \geq 10$) in which all the intended translations were first checked. Other researchers either recruited a small group of bilinguals ($N \leq 4$) to perform a translation norming task or they did not conduct a translation norming study at all (e.g., Guo et al., 2012; Jiang, 1999; Jiang & Forster, 2001; Wang & Forster, 2010, 2015). Translation norming experiments were until now often unavoidable because there was no large scale norming database with English-Chinese translations in the literature.

To fill the gap in the literature and help researchers who are investigating Chinese-English bilingual processing, we obtained translation norms from 28 Chinese-English bilinguals using the "first-translation" method employed in previous translation norming studies (Prior et al., 2007; Tokowicz et al., 2002). However, unlike these previous studies, all bilingual participants in the present study were asked to translate the full list of 1429 English words, resulting in a large translation database. For each English word, all the correct translations and the number of participants providing the correct translations were included in the database. The most frequent translation was defined as the dominant translation for that English word. Furthermore, we included comprehensive lexical information of the translation equivalents (e.g., word frequency, part of speech and concreteness, see below for details). In addition to collecting translation norms, the present study also investigated the relationship between translation ambiguity and word frequency, concreteness and language proficiency.

2.1.1 Word Frequency and Concreteness

In the present study, we first focused on the relationship between word frequency, concreteness and translation ambiguity. Contradictory findings have been reported in the literature in terms of the relationship between word frequency and translation ambiguity. In regression analyses, both Prior et al. (2007) and Tseng et al. (2014) found that the word frequency of the source language was a reliable predictor of the number of translations. However, whereas a negative correlation between word frequency and number of translations was found for English-Spanish translation pairs (Prior et al.,

2007), a positive correlation was observed for English-Chinese translation pairs (Tseng et al., 2014). Thus, in the present study we also investigated the relationship between word frequency and translation ambiguity. Unlike word frequency, the relationship between concreteness and translation ambiguity is more consistent. Using the identical set of English words, the English-Dutch (Tokowicz et al., 2002) and English-Chinese (Tseng et al., 2014) translation norming studies both found a negative correlation between the number of translations and the concreteness ratings obtained from the MRC Psycholinguistic Database (Coltheart, 1981), which indicates that more concrete words have fewer translations. This finding was replicated in a Spanish-English translation norming study which showed that imageability, a lexical characteristic highly correlated with concreteness, was negatively correlated with the number of translations (Prior et al., 2007). However, the MRC Psycholinguistic Database did not provide the concreteness ratings for 18.3% (103 out of 562) English words in Tokowicz et al. (2002) and Tseng et al. (2014). Therefore, it is important to explore further whether this relationship remains the same when a much larger set of words is used. To estimate word concreteness, we used the concreteness ratings from Brysbaert et al. (2014). This recent large-scale concreteness database provides concreteness ratings for almost all of the English words used in the present study (99.9%).

2.1.2 Second Language Proficiency

The second focus of our investigation is the role of second language (L2) proficiency on the bilinguals' translations choices. There are inconsistent findings in the literature about whether L2 proficiency impacts bilinguals' translation choices. The first study that examined the relationship between the bilinguals' language proficiency and their translation choices was Prior et al. (2007). In this study the bilinguals' L2 lexical decision accuracy and their self-assessed rating proficiency were used as estimates of the bilinguals' L2 proficiency. They found a significant positive correlation between L2 proficiency and bilinguals' forward translation (L1 to L2) scores, but not their backward translation (L2 to L1) scores. In contrast, Lemhöfer and Broersma (2012) found that bilinguals' L2 proficiency as measured by an English vocabulary test (LexTALE) was positively correlated with their forward and backward translation accuracy for stimuli taken from Tokowicz et al. (2002). One possible explanation for the inconsistent findings might be the use of different measures of language proficiency.

Some studies have suggested that the objective LexTALE vocabulary test is more reliable than self-assessment (Brysbaert, 2013; Khare et al., 2013; Lemhöfer & Broersma, 2012). Although LexTALE has been used as an estimate of L2 proficiency in several studies (e.g., Bultena, Dijkstra, & van Hell, 2014; Diependaele, Lemhöfer, & Brysbaert, 2013; Khare et al., 2013), none of the translation norming studies have so far used LexTALE to measure the participants' language proficiency. Therefore, in the present study the English language proficiency of the bilinguals will be measured using LexTALE (Lemhöfer & Broersma, 2012) as well as self-rated scores of English language skills.

2.1.3 Dominant and Subdominant Translations

Because word frequency plays such an important role in monolingual and bilingual language processing (e.g., Antón-Méndez & Gollan, 2010; Dijkstra et al., 1998; Kuperman & Van Dyke, 2013; Peeters et al., 2013), we expected that word frequency would also impact the bilinguals' translation choices. To examine this, we first determined the dominant and subdominant translations by searching for the first and second most frequent translations in the present database. If word frequency of the target language does influence the bilinguals' translation choice, we expected a significant frequency difference between the dominant and subdominant translations.

2.1.4 Relationship between Source and Target Language

Given that semantic overlap is a key feature of translation equivalents, Tokowicz et al. (2002) and Tseng et al. (2014) used semantic similarity ratings of translation pairs to investigate the relationship between source and target language. With semantic similarity ratings collected from a different group of participants in their translation norming studies, both studies observed that words with more translations were considered less similar to their translation equivalents, indicated by a reliable negative correlation between number of translations and semantic similarity ratings. This observation captures an important fact about translation ambiguity and the bilingual mental lexicon, because the existence of multiple translations may influence bilinguals' judgments on how similar the translation pairs are in meaning. In the present study, we further explored the relationship of source and target language by investigating the correlation between the word frequencies of the English words and the Chinese

translations (dominant translation) and for pairs that are unambiguous and ambiguous. Kondrak (2013) reported a positive correlation between the word frequency of English words and their French translations (r = 0.573), whereas the correlation of word frequency for non-translation pairs was approximately zero. Kondrak pointed out that this positive correction for translation pairs was due to the considerable overlap in semantic concepts, which would occur with similar frequencies in French and English. According to this logic, we expected to find a higher correlation with word frequency for translation-unambiguous pairs than translation-ambiguous pairs. Finally, we examined the correlation of word frequencies among Chinese-English cognate pairs. As mentioned in Chapter 1, Chinese-English cognates are transliterated loan words, which were created to keep pronunciations as similar as possible to the original English. Thus, cognates are defined in the present study as loan words with high similarity in phonology between English and Chinese, such as coffee - 咖啡 [kʰa fei].[8] Because these cross-script cognates can be regarded as semantically unambiguous (Dong & Lin, 2013; Qi, 2011), the correlation of the word frequencies of cognates pairs is likely to be high.

[8] Cognates have generally been defined in same-script language pairs as words that "have (almost) the same form and meaning" (Dijkstra, 2007, p. 373). Recently, cognates have been defined as words that "share all or a large part of their phonological and/or orthographic form with their translation in another language" (de Groot, 2011, p. 449). Studies involving different-script languages have used the term "cognate" to refer to phonologically similar translation pairs (e.g., Hebrew-English: Gollan et al., 1997; Japanese-English: Hoshino & Kroll, 2008; Korean-English: Kim & Davis, 2003; Greek-English: Voga & Grainger, 2007). As pointed out by Voga and Grainger (2007), the notion that cognates are phonologically and orthographically similar translations is "not a viable definition for languages with different scripts" (p. 948). In line with these studies, we defined in this chapter cognates as phonologically similar translations.

2.2 Experiment 1: Translation Norming Study

2.2.1 Method

2.2.1.1 Participants

Thirty (Male: 2) Chinese-English bilinguals participated in the study. All participants received an inconvenience allowance. One participant dropped out and another one was excluded from analyses because she did not receive formal Chinese education. The remaining participants included undergraduate (N = 22) and postgraduate (N = 6) university students studying in Nottingham. They met the minimum English language entry requirements to study at the University of Nottingham (IELTS 6.0 for undergraduates and 6.5 for postgraduates). The bilingual participants were all native Mandarin Chinese speakers who learnt Chinese from birth and learnt English by classroom instructions from primary or secondary schools onwards. All participants grew up in Mainland China and were not immersed in an English-speaking environment before they came to the UK for their undergraduate or postgraduate studies. The language background of the bilinguals was comparable to bilinguals who participated in other studies in the literature (Guo et al., 2012; Wang & Forster, 2010, 2015; Wu et al., 2013; Wu & Thierry, 2010, 2011, 2012a, 2012b; Zhang et al., 2011; Zhou et al., 2010). For example, participants' self-rated proficiency scores were similar to those of Zhang et al. (2011), i.e., 4.70 vs. 4.69 (7-point scale, 1 = very poor, 7 = native-like). Table 6 summarizes the participants' language background data.

Table 6. Summary of language background data

	Mean (SD)
Age (years)	22.5 (2.62)
Age exposed to formal English education	9.5 (1.90)
Time studies English (years)	13.0 (2.40)
English immersion experience (months)	11.0 (12.30)
LexTALE test score	57.4 (8.28)
Subjective English ability assessment	
Speaking	4.5 (0.74)
Listening	4.8 (0.70)
Reading	4.9 (0.76)
Writing	4.6 (0.92)

Note. LexTALE (Lemhöfer & Broersma, 2012); Speaking, listening, reading and writing ability were rated on a 7-point scale (1 = very poor, 7 = native-like)

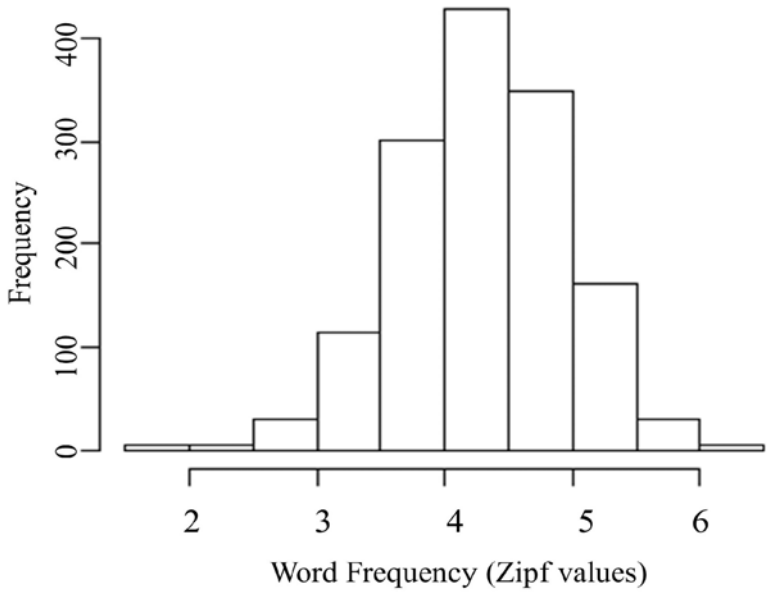
2.2.1.2 Stimuli

Because only a few studies in the literature that used English-Chinese translation equivalents provided their experiment stimuli in an appendix (e.g., Dong & Lin, 2013; Guo et al., 2012; Wang, 2013; Wang & Forster, 2015), it is impossible to generate a large set of English words by simply using the experiment stimuli used in previous studies. Therefore, a large set of English words for the current study was selected by translating two-character Chinese nouns from the Contemporary Chinese Dictionary (Institute of Linguistic from Chinese Academy of Social Science, 2005).[9] Only words that could be translated into a single English word were selected. These Chinese words were considered the expected translations of the English words for the scoring procedures. In total, 1429 English words were selected with a mean Zipf value of 4.30

[9] The Contemporary Chinese Dictionary is known as 现代汉语词典 (Xiandai Hanyu Cidian) and compiled by the Institute of Linguistic from Chinese Academy of Social Science. This is a well-known and authoritative dictionary in Mainland China.

(SD = 0.65) in SUBTLEX-UK (van Heuven et al., 2014). As shown in Figure 15, these 1429 words covered a wide range of frequencies.

Figure 15. Histogram of English word frequencies (N = 1429)



The mean concreteness rating of the English words (Brysbaert et al., 2014) was 3.69 on a 5-point scale (1= abstract and 5 = concrete, N = 1427, SD = 1.06).[10] The average word length was 6.51 letters (SD = 2.18). Based on the dominant part-of-speech information

[10] British English was used in the study, whereas the concreteness ratings are based on American English. Therefore, I obtained the concreteness ratings for 15 British English words using their American counterparts: millimetre (millimeter), centimetre (centimeter), theatre (theater), fibre (fiber), saviour (savior), neighbour (neighbor), behaviour (behavior), flavour (flavor), colour (color), labour (labor), harbour (harbor), humour (humor), pyjamas (pajamas), cheque (check), defence (defense). In addition, the concreteness ratings for 6 words were obtained from their corresponding singular/plural: criterion (criteria), nobles (noble), jungles (jungle), materials (material), maths (math), fans (fan).

in SUBTLEX-UK (van Heuven et al., 2014), most of the words were nouns (N = 1285).

2.2.1.3 Procedure

English words were presented one at a time on a computer screen using E-Prime (Schneider, Eschman, & Zuccolotto, 2002). Participants were asked to type in the first translation that came to their mind and they were instructed to skip the word by pressing the ENTER key if they could not translate the word. All participants used simplified Chinese characters when typing Mandarin Chinese words. The 1429 English words were divided into 14 blocks (about 100 words per block) and the words in each block were matched in frequency. Participants completed the translation task across four separate days within a week. In the first three days, participants completed four test blocks per day. On the final day, after the last two blocks participants completed an English vocabulary test (LexTALE, Lemhöfer & Broersma, 2012) and a language background questionnaire. The 14 blocks were counterbalanced across participants using a Latin Square design. Within each block, the words were presented in a random order. Participants were prompted to take a rest after completing each block. The experiment was approved by the Ethics Committee in the School of Psychology at the University of Nottingham. Participants all signed a consent form before the experiment and were tested individually in a sound-attenuated experimental room.

2.2.1.4 Measures of Word Frequency

As an estimate of word frequency, the present study used the subtitle-based word frequencies from the SUBTLEX-UK (van Heuven et al., 2014) and SUBTLEX-CH (Cai & Brysbaert, 2010). Furthermore, we used the Zipf scale for word frequency. The Zipf scale is a logarithmic scale with values ranging from 1 (very low frequency) to 7 (very high frequency), which are comparable across corpora with different sizes (van Heuven et al., 2014). In addition, the Zipf value for words not available in the databases were calculated using the formula provided by van Heuven et al. (2014), which was 1.47 for SUBTLEX-CH.

2.2.1.5 Language Proficiency Test (LexTALE)

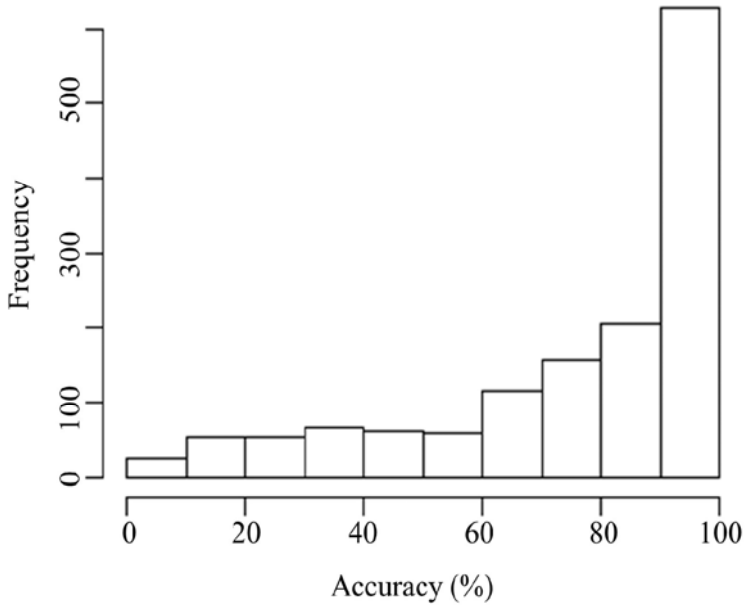
The Lexical Test for Advanced Learners of English (LexTALE) was developed by Lemhöfer and Broersma (2012). In this proficiency test, participants have to decide whether letter strings are real English words or not without any time limit. Although LexTALE only tests English vocabulary knowledge, Lemhöfer and Broersma (2012) demonstrated that the scores are a valid measure of general English proficiency. LexTALE scores range from 0 to 100, and the test is in particular good at differentiating between proficient bilinguals: advanced users (80-100), upper intermediate users (60-80), and lower intermediate users (below 59). LexTALE has become a useful test to quickly measure the participants' language proficiency (e.g., Bultena et al., 2014; Diependaele et al., 2013; Khare et al., 2013) because it takes less than 5 minutes to complete and it is freely available (www.lextale.com).

2.3 Results and Discussion

2.3.1 Translation Norms

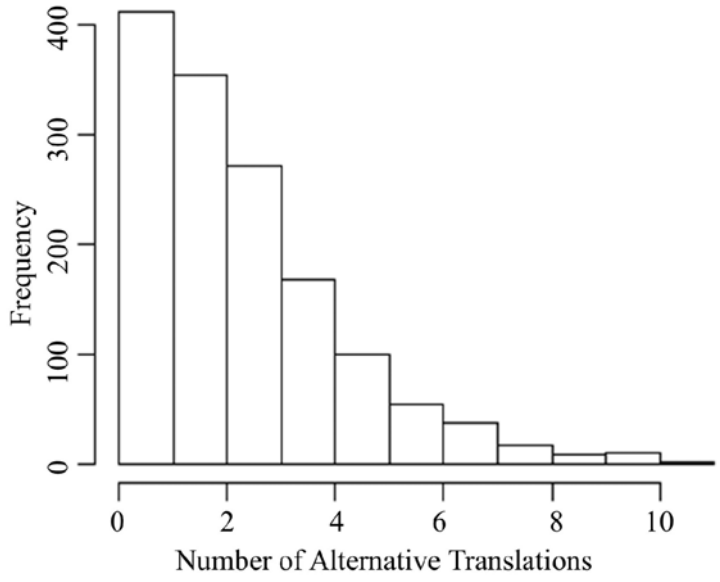
The Chinese translations provided by the participants were automatically compared to the expected translation in E-Prime. All other translations were manually checked using the Oxford Advanced Learner's English-Chinese Dictionary (Hornby, 2010). In total, 28 participants provided 35,262 translations for the 1429 English words. There were 12.0% (4750) omitted responses. Out of the 35,262 translations, 86.7% (30,576) were correct. A large percentage (83.0%) of the English words (1187 words) had an accuracy of 50.0% or more. The histogram of the translation accuracy presented in Figure 16 revealed that the Chinese-English bilinguals knew most of the English. The translation accuracy of the English words was positively correlated with word frequency, $r_s = .601$; $p < .01$. This shows that the more frequent the word is, the more likely the meaning was known by the bilinguals.

Figure 16. Histogram of the translation accuracy



Of the 1429 English words, 0.3% (5 words) received no correct translations, 28.6% (408 words) received one unique correct translation, and 71.2% (1017 words) received more than one correct translation and these were therefore considered translation ambiguous. The histogram of the number of translations is presented in Figure 17.

Figure 17. Histogram of the number of translations



For each English word, the most frequent correct translation was defined as the dominant translation for that English word.[11] The word frequency of the dominant translations also covered a wide range of frequencies (mean = 4.25, SD = 0.74, range 1.47-6.76). Because there were 318 English words in the present study that were also used in Tseng et al. (2014), we further validated the dominant translations by checking whether they are one of the translations provided by participants in Tseng et al. (2014). In total, 96% of the 318 dominant translations provided for the English words in this study were one of the translations provided for the same English words in Tseng et al. (2014).

The proportion of translation ambiguity in the data (71.2%) is very similar to the proportion reported in Tseng et al.'s (2014) English-Chinese translation study (67.3%). Interestingly, compared to same-script language translation norms, the percentage of translation ambiguity is higher for English-Chinese translation pairs than for Dutch-English pairs (Dutch to English: 25.3%; English to Dutch: 30.4%, taken from Tokowicz et al., 2002) and for Spanish-English pairs (Spanish to English: 48.1%; English to Spanish: 57.9%, taken from Prior et al., 2007). One possible explanation lies in the inherent differences in the writing systems. As mentioned in Tseng et al. (2014), most of the Chinese-English translation pairs received a "1" on a 7-point scale in a form similarity rating task. In contrast, same-script language pairs, such as Dutch-English and Spanish-English have larger orthographic and phonological overlap, which impacts bilinguals' translation choices because translations similar in form are translated more often than those different in form (Prior et al., 2007). Furthermore, the direction of the translation task may also have contributed to the high percentage of translation ambiguity. In the present study, unbalanced Chinese-English bilinguals performed a backward translation task. Tokowicz et al. (2002) reported that backward translation (L2 to L1) resulted in more alternative translations than forward translation because of the larger vocabulary size in L1. In addition, Chinese is a language full of synonyms or near-synonyms. For example, the English word "surprise", received three different

[11] 46 items have 2 translations with equal participants. In these case I used the expected translation as the dominant translation. The file ECTN_all_translations.xlsx contains all the correct translations and the number of participants providing each translation.

correct translations in the present study and in Tseng et al. (2014), i.e., “惊喜” (dominant translation), “惊讶”, and “惊奇”. The non-dominant translations could also be translated into synonyms or near-synonyms of its dominant translation, i.e., “诧异”, “惊异”. All these correct translations have very subtle differences in meaning and in the way that they are used in the language so that even native Chinese speakers are sometimes unable to distinguish between them. Given that dialects of Chinese may also vary in vocabulary, the role of synonyms or near-synonyms in Chinese is a complex phenomenon.

2.3.2 English (L2) Word Frequency, Concreteness and Translation Ambiguity

Next, we investigated the relationship between the English word frequency, concreteness, and translation ambiguity. The Spearman’s rank correlation between the number of alternative translations and English word frequency was significant, $r_s = .161$, $p < .01$. This small positive correlation indicated that more frequent words tend to have more translations. This finding is consistent with Tseng et al. (2014), who also reported a small positive correlation between the number of alternative translations and English word frequency ($r = .07$). However, Prior et al. (2007) reported a negative correlation between word frequency and number of translations when Spanish-English bilinguals translated from Spanish or English. It is not clear why the relationship between word frequency and number of translations is different in the Prior et al. (2007) study. Given that Tseng et al. (2014) and the present study used different sets of English words and reported similar findings, one possible explanation lies in the inherent difference between Chinese and Spanish. As discussed above, Chinese has an overall higher level of translation ambiguity, and is full of synonyms or near-synonyms. Potentially, higher-frequency English words have more Chinese translation equivalents than Spanish translation equivalents.

A Spearman’s rank correlation was also calculated between the number of alternative translations and the concreteness ratings of the English words ($N = 1427$) obtained from Brysbaert et al. (2014). Consistent with Tokowicz et al. (2002) and Tseng et al. (2014), a small but significant negative correlation was found, $r_s = -.190$, $p < .01$. This correlation indicated that more concrete words tend to have fewer alternative translations.

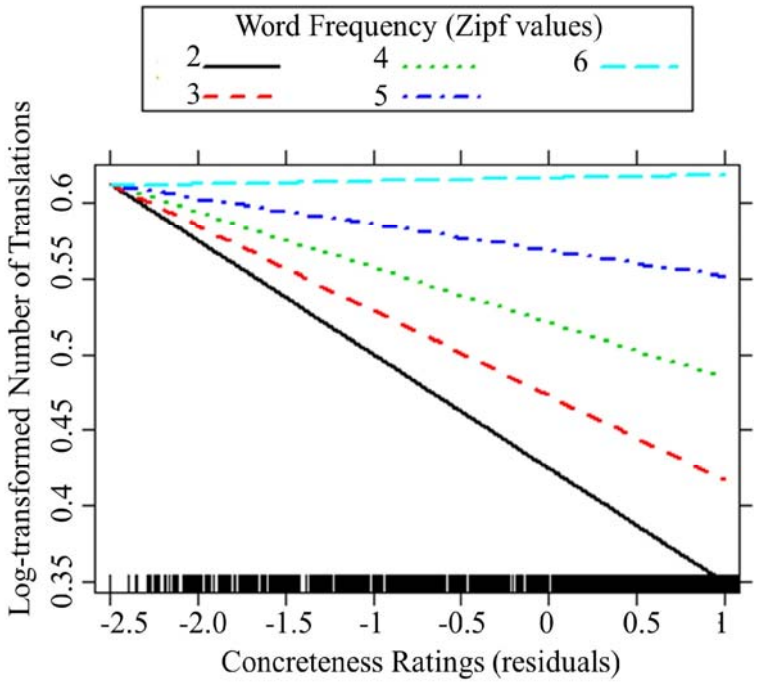
Although correlations analyses suggest a relationship between word frequency, concreteness and translation ambiguity, it is unclear which of these factors predict translation ambiguity. Therefore, a fixed-effects hierarchical regression analysis was conducted using R (version 3.2.4) with the number of translations as the dependent variable and lexical variables of English words as fixed effects. To identify possible predictors, correlation analyses were conducted between the number of translations and other available lexical information: word length and number of part-of-speech categories taken from van Heuven et al. (2014). Both word length and number of part-of-speech categories did not correlate with the number of translations (*p*s> .30). Therefore, the regression analysis considered only word frequency and concreteness ratings as fixed effects. Two English words were excluded from the analysis because they did not have concreteness ratings and five words were excluded because they received no correct translations. The dependent variable (number of translations) was log-transformed (Baayen, 2008). In order to address the issue of the collinearity between word frequency and concreteness ratings, the concreteness ratings were orthogonalised by fitting a linear model in which concreteness ratings were predicted by word frequency (see, for example, Siyanova-Chanturia, Conklin, & van Heuven, 2011, for a similar approach). The residuals of this model were then included in the regression analysis as the concreteness predictor. Word frequency was entered in the first step to predict the number of translations, concreteness ratings were entered in the second step, and their interaction was entered in the final steps. For each step, the model was compared to the previous model using an ANOVA test. We proceeded with the more complex model if the test was significant. In the first step, word frequency (SUBTLEX-UK Zipf value) was a significant predictor. In the second step, word frequency and concreteness ratings were both significant predictors. In the third step, word frequency, concreteness ratings and their interactions were all significant predictors (Table 7). The interaction between word frequency and concreteness ratings was visualized in Figure 18 using the Effects package (Fox, 2003) in R. As can be seen in Figure 18, the word frequency effect on number of translations was more pronounced

for more concrete words than less concrete words.

Table 7. Results of fixed-effects regression analysis

	R^2, adjusted R^2	Estimate (SE)	t value	*p*
Step 1	0.02, 002			
Word frequency		0.05 (0.03)	5.97	< .001
Step 2	0.05, 0.05			
Word frequency		0.05 (0.01)	6.06	< .001
Concreteness		-0.31(0.01)	6.59	< .001
Step 3	0.06, 0.06			
Word frequency		-0.16(0.10)	6.23	< .001
Concreteness		-0.11(0.03)	3.46	< .001
Frequency*Concreteness		0.02(0.01)	2.55	<.05

Figure 18. The interaction between word frequency and concreteness ratings



Overall, the regression analysis showed that both word frequency and concreteness ratings significantly predicted the number of Chinese translations for English words.

This finding is consistent with previous translation norming studies (Prior et al., 2007; Tseng et al., 2014), which reported that word frequency significantly predicted number of translations. As previously discussed, both Tseng et al. (2014) and the present study found that more frequent words tend to have more alternative translations than less frequent words, whereas Prior et al. (2007) observed the opposite pattern. The regression analyses further revealed that the word frequency effect was more pronounced in more concrete words than less concrete words. In addition, more concrete words tended to have fewer alternative translations than less concrete words, which is in line with the findings of Tokowicz et al. (2002) and Tseng et al. (2014). However, it is important to note that the number of translations is not equal to the number of all the possible translations known to all the participants because the present study and all the previous translation norming studies used the "first-translation" method (the first translation that came to the participants' mind). Therefore, this methodological limitation requires further research.

2.3.3 Language Proficiency and Translation Ambiguity

We first obtained the dominant translation scores of each participant by calculating how many times they provided the dominant translations across all the English words. Pearson correlations were then calculated between the participants' dominant translation scores and the various proficiency measures. Participants' dominant translation scores were significantly correlated with their LexTALE scores, $r = .385$, $p < .05$. However, self-rated proficiency scores (listening, speaking, reading and writing) did not correlate with the participants' dominant translation scores (all ps > .10).

To rule out the possibility that more proficient participants were more likely to give any correct translation, the non-dominant translation scores of each participant were calculated by adding up the total number of non-dominant translations provided for all the English words. Participants' non-dominant translation scores did not significantly correlate with their LexTALE scores, $r = .292$, $p = .132$ or self-rated proficiency scores (speaking: $r_s = .362$, $p = .059$, other ps > .10). We also calculated the participants' overall translation scores by adding up the total number of all correct translations (dominant and non-dominant translations) provided for all the English words. As expected, participants' overall translation scores were positively correlated with their LexTALE scores, Pearson's $r = .385$, $p < .05$, but not with self-rated proficiency scores (all

*p*s > .20). Thus, the bilinguals' language proficiency did influence their L2 to L1 translation performance when L2 proficiency was measured using LexTALE (Lemhöfer & Broersma, 2012). The positive correlation between the LexTALE scores and dominant translation scores indicated that more proficient bilinguals are more likely to produce dominant translations than less proficient bilinguals. In line with the previous studies (Brysbaert, 2013; Khare et al., 2013; Lemhöfer & Broersma, 2012), self-rated proficiency scores did not correlate with translation performance. A possible explanation is that self-rated proficiency scores are a less sensitive measure of proficiency than objective proficiency measures such as obtained with LexTALE.

As reported above, translation ambiguity (number of translations) at the item level is predicted by word frequency and concretenes. However, it is unclear whether these factors as well as English proficiency influence whether or not participants provided the dominant translation. To investigate this, a mixed-effects logistic regression was conducted using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R. This regression included lexical variables of English words (word frequency, concreteness ratings) and a subject variable (English proficiency) as fixed effects to predict whether participants provided the dominant translation for an English word or not. This analysis is possible because in the current study all bilingual participants were asked to translate all the English words and dominant translations were obtained. Previous studies could not do this analysis because they did not require all participants to translate all the words or they did not provide the dominant translations (Prior et al., 2007; Tokowicz et al., 2002; Tseng et al., 2014).

Participants' responses were first transformed into a binomial variable: for each English word the response of the participant was coded 1 when their response was the dominant translation and otherwise it was coded as 0. Two English words were excluded from the analysis because they did not have concreteness ratings and five words were excluded from the analysis because they received no correct translations. As before, the collinearity between the word frequency and concreteness ratings was reduced by using the residuals from the linear model. Subjects and items were included in the model as random effects (Baayen, Davidson, & Bates, 2008) and the fixed effects were word frequency, concreteness, and language proficiency (LexTALE scores). A forward model selection procedure was used. For each step, the model was compared to the

simpler model using a Chi-squared test. All models that included fixed effects differed significantly from the null mode with only the random effects ($p < .001$). The final model is presented in Table 8. English word frequency, English concreteness ratings and language proficiency were all significant predictors of participants' translation performance. Any other interactions or the including predictors in the random slopes (Barr, Levy, Scheepers, & Tily, 2013) did not improve the model and did not account for more variance. This final model revealed that participants were more likely to provide the dominant translations for more frequent words and for more concrete words. Furthermore, more proficient bilingual participants tended to provide the dominant translations. Thus, English word frequency, concreteness and language proficiency all predicted participants' translation performance.

Table 8. Results of mixed-effects logistic regression analysis

Random effects	**Variance**		**SD**	
Subject	0.16		0.34	
Item	2.32		1.52	
Fixed effects	**Estimate**	**SE**	**z value**	***p***
Intercept	0.73	0.08	9.51	< .001
Word frequency	1.00	0.07	14.80	< .001
Concreteness	0.18	0.04	4.34	< .05
LexTALE scores	0.02	0.01	2.25	
Marginal R^2	0.08			
Conditional R^2	0.47			

Note. Marginal R^2 (the proportion of variance explained by the fixed factors alone) and conditional R^2 (the proportion of variance explained by both the fixed and random factors) were calculated using the function provided on http://jonlefcheck.net/2013/03/13/r2-for-linear-mixed-effects-models/, which is based on Johnson (2014).

2.3.4 Dominant and Subdominant Translations

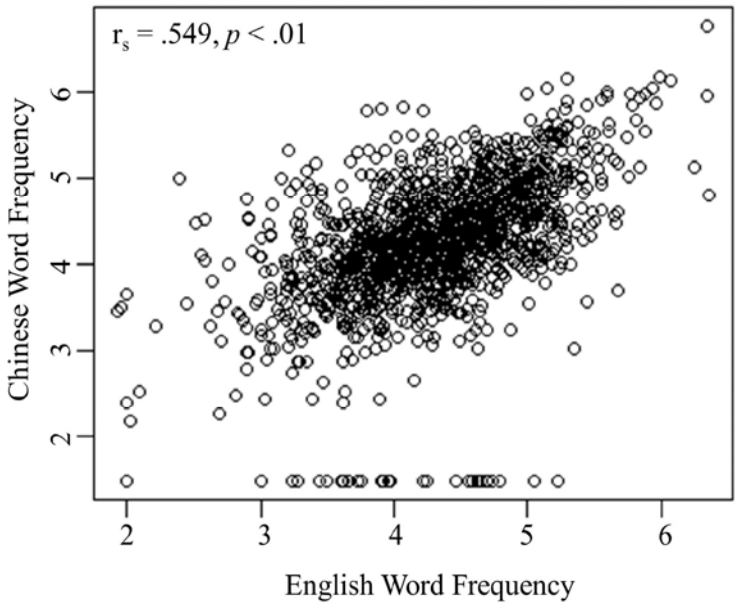
To investigate the effect of the word frequency on the translation choice, 730 English words were selected with a number-of-translations ranging from 2 to 5. The word

frequency (Zipf scale) of the dominant translations were compared with the subdominant translations in a Wilcoxon Signed Ranks Test. Results revealed that the word frequency difference between the dominant translations and the subdominant translations (4.40 vs. 3.92) was significant, $z = 11.29$, $p < .001$.[12] Thus, when participants were asked to translate English (L2) into Chinese (L1) they chose the more frequent Chinese translation equivalents.

2.3.5 Correlation between Source and Target Language Word Frequencies

The correlation between the English word frequencies and their dominant Chinese translation frequencies was significant, $r_s = .549$, $p < .01$. This positive correlation (see Figure 19) indicated that more frequent English words tend to have Chinese translations that are more frequent, even though their writing systems are very different.

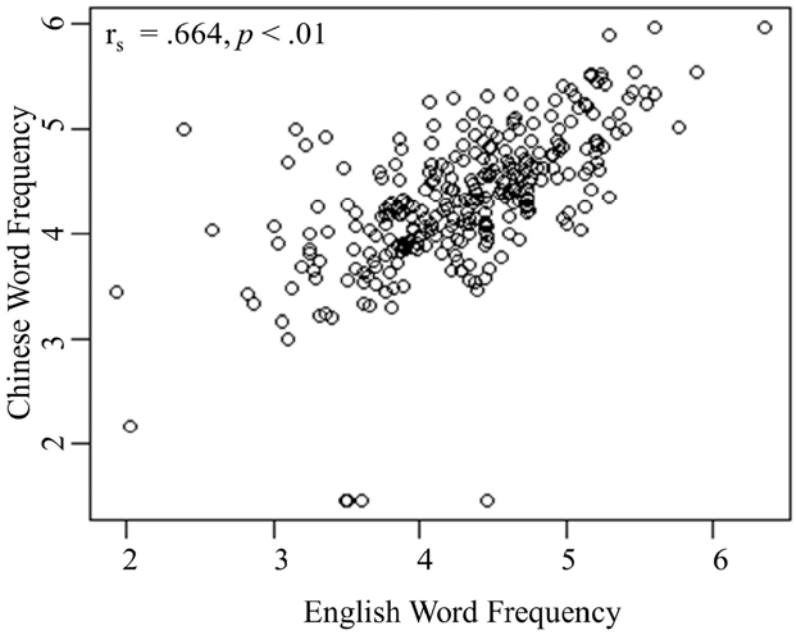
Figure 19. Relationship between word frequency between English words and their dominant Chinese translations



[12] We also conducted a further analysis using 289 English words with 2 alternative translations. Results also revealed that the word frequency difference between the dominant translations and the subdominant translations (4.53 vs. 3.89) was significant, $z = 9.89$, $p < .001$.

To investigate whether this positive correlation depends on the ambiguity of the translations the pairs were divided into translation unambiguous and translation ambiguous. In order to identify translation-unambiguous pairs the following criteria were used: 1) one unique correct translation was obtained for the English word, 2) at least 50% of participants provided that unique translation (the remaining participants either gave no response or an incorrect response), 3) the dominant part-of-speech for the English word was a noun or a name (e.g., bill, prince). This resulted in 307 translation-unambiguous pairs. In total, 560 translation-ambiguous pairs were selected using the following criteria: 1) more than one correct translation was obtained for the English word, 2) at least 50% of participants provided the dominant translation, 3) the dominant part-of-speech for the English word was a noun or a name. Spearman correlations revealed higher correlations between word frequencies for translation-unambiguous pairs, $r_s = .664$, $p < .01$ (Figure 20), than for translation-ambiguous pairs, $r_s = .518$, $p < .01$. Importantly, the correlation between the word frequencies of translation-unambiguous pairs were significantly higher than of translation-ambiguous pairs, $z = 3.174$, $p < .001$ (Diedenhofen & Musch, 2015). This suggests that translation ambiguity impacts the relationship between source and target language. However, some of the translation-unambiguous pairs in the present study were not classified as

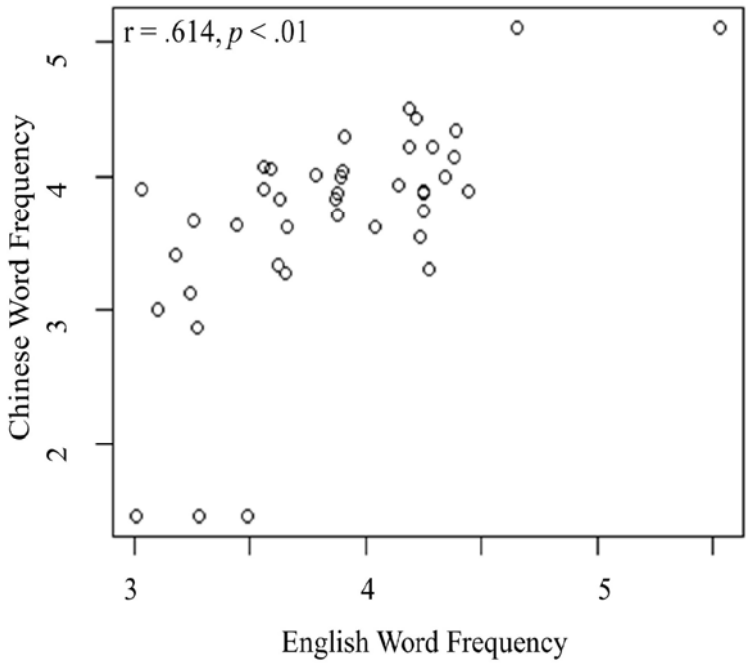
Figure 20. Relationship between English and Chinese word frequency for translation-unambiguous pairs (N = 307)



unambiguous in Tseng et al. (2014). For example, the word "*tail*" is a translation-unambiguous item in our database but ambiguous in Tseng et al. (2014). Furthermore, whereas the word "carrot" was translated into two correct translations in our study it was only translated into one correct translation in Tseng et al. (2014). Therefore, we further investigated the relationship between the translation pairs' frequencies using cognate pairs, which are considered to be semantically unambiguous.

Forty-one English-Chinese translation pairs were identified as cross-language cognates using the Contemporary Chinese Dictionary (2005). In this dictionary, transliterated loan words from English are specified together with their original English words. The correlation between the English frequencies and their Chinese cognate translation frequencies was significant, $r_s = .614$, $p < .01$ (Figure 21). As expected, this correlation was higher than the overall correlation between all the English words and the dominant translations (Figure 19).

Figure 21. Relationship of English and Chinese word frequency for cross-script cognates (N = 41)



2.4 Summary and Conclusions

The present translation norming study revealed that 71% of the English words were translated into more than one correct translation, and that more frequent words and less concrete words have more alternative translations as revealed by a small but significant correlation. A regression analysis further revealed that English word frequency and concreteness significantly predicted the number of Chinese translations of English words. Furthermore, an interaction between these predictors revealed that the number of translations was more affected by word frequency for more concrete words than for less concrete words. In addition, word frequency, concreteness and bilingual language proficiency were all significant predictors of whether participants provided a dominant translation for an English word or not. An important finding of the present study is that there are significant differences between the word frequency of the dominant translations and the subdominant translations, suggesting that when participants translated English (L2) into Chinese (L1) they selected the more frequent Chinese translation equivalent. Another important finding is the positive correlation between the word frequencies of the English words and their dominant Chinese translations. This positive correlation is higher for translation-unambiguous pairs than for translation-ambiguous pairs, and also high for cross-script cognate pairs.

The present database with backward translation (L2 to L1) norms is useful for future investigations of Chinese-English bilingual language processing, in particular L1 activation during L2 processing. Furthermore, the norms are also be useful for modelling bilingual word recognition and word translation (e.g., Dijkstra & Rekké, 2010) and for cross-language comparisons (e.g., Schepens, Dijkstra, Grootjen, & van Heuven, 2013). A limitation of the current study is that translation norms provided by this database were collected from unbalanced Chinese-English bilinguals conducting backward translation (L2 to L1) only. Therefore, researchers should be cautious when using this database for studies that investigate L2 (English) activation during L1 (Chinese) processing and when the research involves more balanced bilinguals. However, for studies with balanced bilinguals, the backward translation norms provide a baseline. To conclude, the Chinese translation norms for 1429 English words obtained in this study can be used by researchers investigating Chinese-English bilingual processing. The translation norms have been made available online for other

researchers together with lexical information about the words (see Appendix 1 for more details). All the stimuli used in the following chapters were based on English words and their dominant Chinese translation selected from this translation norming study. Therefore, this database ensures that a well-controlled set of stimuli could be selected in the investigation of Chinese translation activation during English word reading in Chinese-English bilinguals.

Chapter 3: Phonological Activation of Chinese Translations during English Word Reading: Behavioural Investigations

3.1 Introduction

As discussed in Chapter 1 (Section 1.4), several studies that used the hidden repetition paradigm have consistently reported that Chinese translations are activated when Chinese-English bilinguals read English words (e.g., Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a; Zhang et al., 2011). In their seminal study, Thierry and Wu (2007) presented proficient Chinese-English bilinguals with pairs of English words and asked them to judge whether these word pairs were related in meaning or not. Critically, they found a reduction of the N400 component for target words when they were preceded by prime words sharing an identical character in the Chinese translation (e.g., *train* [火车] - *ham* [火腿]). This hidden repetition effect suggests that bilinguals activate the Chinese translations of the English words. The hidden repetition effect is also observed in a behavioural experiment that combined the masked priming technique with the hidden repetition paradigm (Zhang et al., 2011). In this behavioural masked priming study, Chinese-English bilinguals were asked to perform an English lexical decision task (i.e., judging whether the presented letter string is a real English word or not), during which English target words were preceded by English primes presented for 59 ms. The results provided behavioral evidence for the activation of Chinese translations of the prime and target, which resulted in faster decisions to targets when the Chinese translation of the prime had the same the first character as the Chinese translation of the target (e.g., east [东] - thing [东西]). However, because the repeated characters in Thierry and Wu (2007) and Zhang et al. (2011) had the same pronunciation, it is unclear whether the observed hidden repetition effect originates from Chinese phonology or orthography. Wu and Thierry (2010) tried therefore to dissociate between repetition of Chinese phonological and orthographical information. Their results revealed a reduced N400 for the hidden phonological repetition (e.g., *experience* [经验, /**Jing1**Yan4/] - *surprise* [惊讶, /**Jing1**Ya4/] but not for the orthographic repetition (e.g., *accountant* [会计, /Kuai4Ji4/] - *conference* [会议, /Hui4Yi4/], suggesting that Chinese phonology

drives the hidden repetition effect. As noted in Chapter 1, Chinese phonology contains both segmental and tonal information, therefore it remains unclear whether segmental and/or tonal information drives the hidden repetition effect.

Two behavioural experiments reported in this chapter investigated therefore whether hidden repetition priming is driven by segmental and/or tonal information of Chinese translations of English words. In both experiments, the repetition of segmental and tonal information of the Chinese translations, i.e., the first character of Chinese translations for English) were systematically manipulated in semantically and orthographically unrelated English word pairs, e.g., *empire* [帝国, /Di4Guo2/] - *place* [地点, /Di4Dian3/], with repeated segment and tone; *enemy* [敌人, /Di2Ren2/] - *place* [地点, - /Di4Dian3/], with repeated segment; *author* [作者, /Zuo4Zhe3/] - *place* [地点, /Di4Dian3/], with repeated tone; *aspect* [方面, /Fang1Mian4/] – *place* [地点, /Di4Dian3/], with no repetition at all. As discussed in Chapter 1 (Section 1.7.2), Chinese phonological activation observed in Wu and Thierry (e.g., Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a) is most likely due to the feedback from semantics. Therefore, if the activation of the phonology of Chinese translations during English word reading is limited to phonology at the lexical level where segment and tone are combined, a priming effect would only be observed when both segment and tone are repeated and not when only either segment or tone is repeated. The second possibility is that the feedback from semantics also extends to the sublexical phonological level. In that case, any phonological repetition (segment and/or tone) would result in a priming effect. To explore the source of the hidden repetition effect is important because it helps to specify a central issue in bilingual language processing, namely, which L1 representations are activated during visual word recognition in L2.

In addition to manipulating the segmental and tonal repetition in the Chinese translations of the English words, the type of task was also manipulated to investigate whether Chinese phonological activation during English word reading is task-dependent. The first experiment (Experiment 2) in this chapter used an unmasked hidden priming paradigm with a semantic relatedness judgment task (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a), whereas the second experiment (Experiment 3) combined the masked priming paradigm with a lexical decision task (Zhang et al., 2011).

The reason for manipulating the task is because electrophysiological studies using the unmasked hidden priming paradigm with a semantic relatedness judgment task often reported an N400 reduction in hidden repetition priming and no significant effects in the behavioural data (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a), whereas a behavioural study using the masked hidden repetition priming paradigm with a lexical decision task (Zhang et al., 2011) found significant priming effects in response times. It is worth noting that in the literature behavioural cross-language priming effects are reported often stronger in semantic tasks (i.e., semantic relatedness/categorization judgment) than in lexical decision tasks (e.g., Chen et al., 2014b; Schoonbaert et al., 2009; Xia & Andrews, 2015). Therefore, it is surprising why an opposite pattern is found in the studies using the hidden repetition priming paradigm. In addition, as reviewed in Chapter 1 (Section 1.8.3), a considerable number of studies on phonological activation during Chinese character/word reading suggested that phonological priming was more salient in semantic tasks (i.e., semantic relatedness/categorization judgement task) than a lexical decision task. Therefore, it is necessary to understand the impact of task on the behavioural evidence of Chinese phonological activation during English word reading.

There are two possible explanations for the differences between the semantic relatedness judgement experiments (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a) and lexical decision experiments (Zhang et al., 2011). The difference in behavioural findings of these studies may lie in the differences between the statistical power in each study. The semantic relatedness judgement experiments only tested 15 bilinguals, whereas 32 bilinguals were tested in the lexical decision experiments. If statistical power is indeed an issue, a hidden repetition priming would be expected in both experiments presented in this Chapter because 32 participants were tested in each experiments. A second possible explanation is that phonological activation of Chinese translations might be short-lived, therefore it only impacts response times in lexical decision because the interval between primes and targets is short (i.e., 23 ms) in contrast to the long intervals used in semantic relatedness judgement experiments (e.g., 500 ms). If this explanation is correct, a hidden repetition priming would only be found in the lexical decision task (Experiment 3) and not in the semantic relatedness judgement task (Experiment 2).

To summarize, the present chapter investigates Chinese phonological activation when Chinese-English bilinguals read English words in two behavioural experiments. The results of a semantic relatedness judgement experiment (Experiment 2) and a lexical decision experiment (Experiment 3) are reported. In both experiments the segmental and tonal information of the Chinese translations were manipulated in English word pairs to explore the source of the hidden repetition priming.

3.2 Experiment 2: Semantic Relatedness Judgment Task

3.2.1 Method

3.2.1.1 Participants

35 Chinese-English bilinguals participated in this experiment and they received an inconvenience allowance. Three participants were excluded from analyses due to high error rates (>25%) in the experimental trials. The remaining 32 bilingual participants (Mean age: 22.6, SD = 2.46, male: 13) were first exposed to English at the age of 9.97 (range: 6-13). By the time of testing, they had learnt English for an average of 12.63 years (range: 9-19) and had lived in the United Kingdom for an average of 13.13 months (range 2-42). The unbalanced bilingual participants were all native Chinese speakers who learnt Chinese from birth and they were undergraduate or postgraduate university students studying in Nottingham. They met the minimum English language entry requirements to study at the University of Nottingham (IELTS 6.0 for undergraduates and 6.5 for postgraduates). Their English proficiency was objectively measured with LexTALE (Lemhöfer & Broersma, 2012), which is a quick and reliable English vocabulary test (See Section 3.2.5 in Chapter 2 for more information about LexTALE). The participants' LexTALE score was 57.18 on average (range 42.5-75). Their self-rating score of English reading proficiency on a 7-point scale (1 = very poor, 7 = native-like) was 4.75 (range 3-6). Participants received an inconvenience allowance of 3 pounds.

3.2.1.2 Materials and Design

Stimuli were taken from the English-Chinese translation norming study reported in

Chapter 2. In total, 72 English target words were selected that were preceded by four types of English primes: 1) translated same-tone homophone prime (+Segment +Tone), 2) translated different-tone homophone prime (+Segment -Tone), 3) translated same tone but different syllables prime (-Segment +Tone), 4) control prime (-Segment -Tone) which has no phonological overlap with the target. Importantly, the English prime and target words were all orthographically, phonologically, and semantically unrelated in English. An overview of the stimulus characteristics and examples of the stimuli are presented in Table 9 (see Appendix 3 for the list of all stimuli). Similar to Zhang et al. (2011) only the first character of the Chinese translations were repeated. All critical word pairs were unrelated in meaning. For the semantic judgement task a set of 72 semantically related filler pairs were included. The semantic relationship of the word pairs was checked against the University of South Florida free association database (Nelson, McEvoy, & Schreiber, 2004). This ensured that target words in the critical word pairs were not associated with the four prime words and that the target words of the fillers were always free associates of the prime words (e.g., *funeral* - *death*, mean association strength: 0.22). Four counterbalanced lists were created and participants were randomly assigned to one of the lists.

Table 9. Experimental design and stimulus characteristics of Experiments 2 (Means with SDs in brackets)

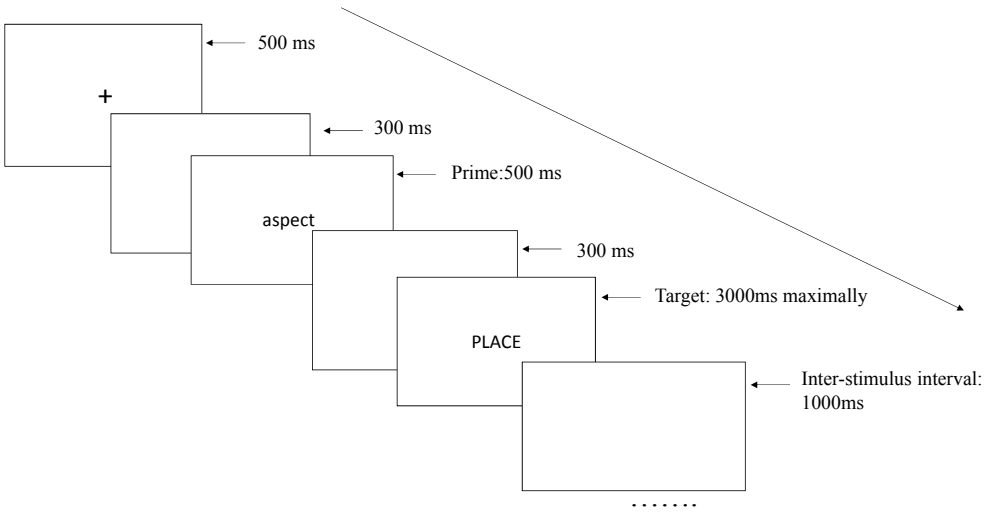
Experimental Trials (Semantically Unrelated)					
	Prime				Target
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*	
English	empire	enemy	author	aspect	place
Chinese Translation	帝国	敌人	作者	方面	地点
Pinyin	***Di4***Guo2	***Di2***Ren2	***Zuo4***Zhe3	***Fang1***Mian4	***Di4***Dian3
Word Length	5.88(1.61)	6.17(1.96)	5.56(1.21)	5.85(1.24)	5.97(1.88)
English Frequency	4.52(0.64)	4.38(0.60)	4.65(0.46)	4.40(0.62)	4.62(0.66)
Chinese Frequency	4.54(0.61)	4.39(0.69)	4.57(0.46)	4.36(0.62)	4.49(0.56)
Concreteness	3.46(1.07)	3.82(1.07)	3.98(1.01)	3.89(1.08)	3.90(0.98)
Filler Trials (Semantically Related)					
English	minute - hour				
Chinese Translation	分钟 小时				
Pinyin	Fen1Zhong1 Xiao3Shi2				
Word Length	5.81(2.00) 5.61(1.95)				
English Frequency	4.64(0.56) 4.78(0.61)				
Chinese Frequency	4.56(0.65) 4.66(0.71)				
Concreteness	4.20(0.89) 4.02(1.02)				

Note. English frequencies are Zipf values from SUBTLEX-UK (van Heuven et al., 2014). Chinese frequencies are Zipf values calculated using the word frequencies from SUBTLEX-CH(Cai & Brysbaert, 2010). Concreteness scores are based on Brysbaert et al. (2014). Pinyin is a Romanisation of the Chinese pronunciation.

3.2.1.3 Procedure

Participants were tested individually in a sound-attenuated experimental room. Each participant received a unique random presentation order of the stimuli. The stimuli were presented on a CRT monitor (19" Lacie Electron Blue IV, 1024×768 pixels, 85 Hz) using DMDX Display Software (Forster & Forster, 2003). The sequence of visual events for each trial is presented in Figure 22. Each trial began with a fixation cross, presented for 500 ms followed by a 300 ms black screen. Next, the lowercase prime word was presented for 500 ms, followed by an inter stimulus interval of 500 ms after which the uppercase target word was presented. Participants were asked to decide whether the presented word pairs (prime-target) were related in meaning or not by pressing buttons on a Cedrus response box attached to the computer. They were instructed to perform the task as quickly and as accurately as possible. The target word remained on the screen up to a maximum of 3000 ms or until the participants responded. The inter-trial interval was set at 1000 ms. The experiment was approved by the Ethics Committee of the School of Psychology at the University of Nottingham. Participants all signed a consent form before the experiment started.

Figure 22. Illustration of visual events in Experiment 2



3.2.2 Results

Nine word pairs had high error rates (> 50%). To keep the counterbalanced design 12 targets were excluded from analysis. This resulted in 60 critical word pairs in each list (15 per condition).

The mean reaction times and error rates are presented in Table 10. Incorrect responses (9.9%) and reaction times less than 300 ms or more than 2,500 ms (0.8%) were discarded. A 2 (Segment: +Segment vs. -Segment) × 2 (Tone: +Tone vs. -Tone) repeated measures ANOVA was conducted for both error rates and reaction times across subjects (F_1) and items (F_2). Analyses of the error rates revealed no significant main effects of Segment F_1 (1, 31) = 2.773, p = .106; F_2 (1, 59) = 2.099, p = .153, and Tone, and no significant interaction, all Fs < 1, ps > .40. The analysis of the reaction times also revealed no significant main effects of Tone, F_1 (1, 30) = 3.280, p = .080; F_2 (1, 59) = 3.404, p =.070, and Segment, and no significant interaction, all Fs < 1, ps > .30.

Table 10. Mean error rates and reaction times (with SD in brackets) of Experiment 2

	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*
Error Rates (%)	10.83	11.35	7.92	9.85
	(9.83)	(9.60)	(8.71)	(9.12)
RT (ms)	1130	1097	1105	1098
	(232)	(228)	(220)	(231)

3.2.3 Discussion

In Experiment 2, Chinese-English bilinguals were presented pairs of English words and made decisions about whether or not the presented word pairs were related in meaning or not. Unknown to the participants, the repetition of segmental and tonal information of the Chinese translations (i.e., the first character of Chinese translations for English) were systematically manipulated in semantically and orthographically unrelated English word pairs. This manipulation aimed to explore whether segmental and/or tonal information of the Chinese phonology is activated when Chinese-English bilinguals read English words. Unfortunately, the analysis of the reaction times only revealed a

trend towards an inhibitory effect of repeated tone (F_1 and F_2: $0.05 < ps < 0.10$).

An important question is whether this marginal result could become significant if more participants are tested.[13] As noted earlier, the sample size of participants was predetermined by using the same number of participants in Zhang et al. (2011). Importantly, it is dangerous to increase the predetermined sample size because the practice of optional stopping (testing more participants until having significant results) is doomed to increase Type I errors (Dienes, 2014; Rouder, 2014; Simmons, Nelson, & Simonsohn, 2011; Vasishth & Nicenboim, 2016; Wagenmakers, 2007; Yu, Sprenger, Thomas, & Dougherty, 2014). The marginal findings will be further discussed in the General Discussion section. A critical question is whether the current manipulation of segment and tone would result in significant behavioural results by using the same paradigm as Zhang et al. (2011). Therefore, Experiment 3 used the masked priming paradigm with a lexical decision task to explore whether segmental and/or tonal information of the Chinese phonology is activated when Chinese-English bilinguals read English words.

3.3 Experiment 3: Lexical Decision Task

3.3.1 Method

3.3.1.1 Participants

38 Chinese-English bilinguals participated in Experiment 3 and they received an

[13] There are no exact guidelines for what p-values can be considered as marginally significant (Pritschet, Powell, & Horne, 2016; Rousselet & Pernet, 2011). In the present thesis, a *p*-value smaller than 0.11 is considered as marginal. However, being aware of the criticism on interpreting "marginal significant" results as the evidence of an effect (Pritschet et al., 2016; Rousselet & Pernet, 2011), and the recommendation of interpreting marginal results as insignificant (Rousselet & Pernet, 2011), I use the phrase *marginal results* in the thesis.

inconvenience allowance (none of these participants had participated in Experiment 2). Six participants were excluded from the analysis due to high error rates (>25%) in the experimental trials. The remaining 32 bilingual participants (Mean Age: 21.6, SD = 2.35, Male: 7) were first exposed to English at the age of 9.3 (range: 6-13). At the time of testing, they had learnt English for an average of 12.3 years (range: 9-16) and had lived in the United Kingdom for an average of 7.25 months (range 1-37). The bilingual participants were all native Chinese speakers who learnt Chinese from birth and they were undergraduate or postgraduate university students studying in Nottingham. They met the minimum English language entry requirements to study at the University of Nottingham (IELTS 6.0 for undergraduates and 6.5 for postgraduates). Their English proficiency was objectively measured with LexTALE (Lemhöfer & Broersma, 2012). Their LexTALE score was 62.2 on average (range 42.5- 78.75). Their self-rating score of English reading proficiency on a 7-point scale (1 = very poor, 7 = native-like) was 4.3 (range 2-6).

3.3.1.2 Materials and Design

The critical word pairs were the same as those included in the analysis of Experiment 2 (see Table 11 and see Appendix 3 for the list of all stimuli). Thus, 60 English target words were preceded by four types of English primes: 1) translated same-tone homophone prime (+Segment +Tone), 2) translated different-tone homophone prime (+Segment -Tone), 3) translated same tone but different syllables prime (-Segment +Tone), 4) control prime (-Segment -Tone). A set of 60 pseudowords were selected from the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012). These pseudowords were matched with the target words in length and first letter. Pseudowords were preceded by 60 unrelated English words. Four counterbalanced lists were created and participants were randomly assigned to one of the lists.

Table 11. Experimental design and stimulus characteristics of Experiments 3 (Means with SDs in brackets)

Experimental Trials					
	Prime				Target
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*	
English	empire	enemy	author	aspect	place
Chinese Translation	帝国	敌人	作者	方面	地点
Pinyin	***Di4***Guo2	***Di2***Ren2	***Zuo4***Zhe3	***Fang1***Mian4	***Di4***Dian3
Word Length	5.85(1.57)	6.25(1.95)	5.57(1.18)	5.83(1.24)	5.72(1.68)
English Frequency	4.55(0.66)	4.31(0.56)	4.66(0.45)	4.40(0.63)	4.72(0.57)
Chinese Frequency	4.54(0.62)	4.34(0.67)	4.57(0.46)	4.36(0.63)	4.54(0.56)
Concreteness	3.50(1.11)	3.77(1.07)	3.98(1.00)	3.86(1.09)	3.02(0.88)

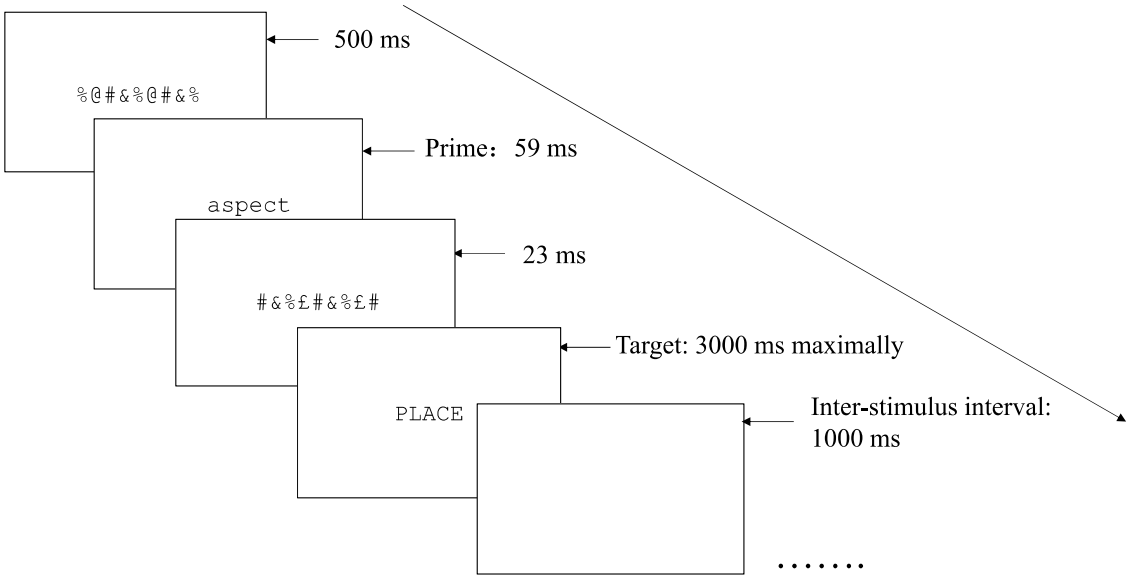
Note. English frequencies are Zipf values from SUBTLEX-UK (van Heuven et al., 2014). Chinese frequencies are Zipf values calculated using the word frequencies from SUBTLEX-CH(Cai & Brysbaert, 2010). Concreteness scores are based on Brysbaert et al. (2014). Pinyin is a Romanisation of the Chinese pronunciation.

3.3.1.3 Procedure

Participants were tested individually in a sound-attenuated experimental room. The stimuli were presented on a CRT monitor (19" Lacie Electron Blue IV, 1024×768 pixels, 85 Hz) using DMDX Display Software (Forster & Forster, 2003). The sequence of visual events for each trial is presented in Figure 23. Each participant received a unique random presentation order. The procedure was identical to Zhang et al. (2011). Each trial began with a forward mask (%@#$%@#$%@#%) presented for 500 ms. Following this mask was a lowercase English word presented for 59 ms, followed by a backward mask (#$%£#$%£#$%#) presented for 24 ms. Immediately following the backward mask the target, an uppercase letter string, was presented. Participants were not informed about the presence of the primes and they were asked to decide whether the uppercase letter strings were English words or not by pressing buttons on a Cedrus

response box attached to the computer. They were instructed to perform the task as quickly and as accurately as possible. The uppercase string remained on the screen up to a maximum of 3000 ms or until the participants responded. The inter-trial interval was set at 1000 ms. To explore whether the prime was visible for the participants, the participants conducted the same experiment again but now their task was to name aloud the lower case (prime). Before this task they were informed about the presence of the prime. Participants were hardly able to name the briefly-presented prime word and therefore this data is not reported here. The experiment was approved by the Ethics Committee of the School of Psychology at the University of Nottingham. Participants all signed a consent form before the experiment started.

Figure 23. Illustration of visual events in Experiment 3



3.3.2 Results

The mean reaction times and error rates are presented in Table 12. Incorrect responses were discarded (6.9%). Reaction times less than 300 ms or more than 2,500 ms were considered outliers (0.6%) were excluded before the analysis. A 2 (Segment: +Segment vs. -Segment) × 2 (Tone: +Tone vs. -Tone) repeated measures ANOVA was conducted with error rates and reaction times across subjects (F_1) and items (F_2). The error analysis revealed no significant main effects of Tone F_1 (1, 31) = 1.289, p = .265, F_2 (1, 59) = 2.331, p = .132, and Segment, and no significant interaction, all Fs < 1, ps > .40. Also the reaction time analysis revealed no significant main effects of segment and tone, and no significant interaction, all Fs < 1, ps > .40.

Table 12. Mean error rates and reaction times (with SD in brackets) of Experiment 3

	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*
Error Rates (%)	5.64	7.40	6.73	7.96
	(6.36)	(10.10)	(8.82)	(9.20)
RT (ms)	941	952	948	960
	(214)	(220)	(214)	(228)

3.3.3 Discussion

In Experiment 3, Chinese-English bilinguals were instructed to make decisions about the lexicality of the presented letter strings. Unknown to the participants, English targets words were preceded by masked and brief-presented (59 ms) English primes. Critically, the segmental and tonal information of the Chinese translations of primes and targets were systematically manipulated to explore Chinese phonological activation during English word reading. Similar to Experiment 2, no significant hidden repetition priming were found in either the error rates or reaction times. In contrast, Zhang et al. (2011) using the same paradigm found a significant facilitation effect in the reaction times. Therefore, the results of Experiment 3 are unexpected, and are not consistent with the predictions. The implications of these unexpected findings are further discussed in the next section.

3.4 General Discussion

In this chapter, the results of two experiments were reported that investigated the activation of segmental and tonal information of Chinese translations of English word pairs in an attempt to clarify the source of hidden repetition priming observed in Chinese-English bilinguals (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a). In addition to manipulating the segmental and tonal information of the Chinese translations in English word pairs, the type of task was also manipulated to investigate whether Chinese phonological activation during English word reading is task-dependent. Experiment 2 used an unmasked hidden priming paradigm with a semantic relatedness judgment task (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a), whereas Experiment 3 combined the masked priming paradigm with a lexical decision task (Zhang et al., 2011). As discussed in the introduction, if the failure to find significant behavioural effects in semantic relatedness judgement tasks (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a) was due to a lack of statistical power, testing the same number of participants as Zhang et al. (2011) would have more chance of finding significant results in both tasks. In contrast, if statistical power is not the issue it could be that phonological activation of Chinese translations is short-lived, therefore only in the masked lexical decision task significant hidden repetition priming effect would be expected. The results of the experiments reported in this chapter showed unexpectedly no significant hidden repetition priming effects in both experiments. It is difficult to interpret results that are not significant because the failure to reject the null hypothesis cannot be taken as evidence of the null hypothesis. In other words, "absence of evidence is not evidence of absence" (Altman & Bland, 1995, p. 485).

An important question with null results is whether enough participants were tested to detect the effect. One way to investigate this issue is to calculate the effect sizes and run a power analysis. The effect sizes for ANOVAs for within-subject designs can be calculated using the algorithm implemented in G*power (Faul, Erdfelder, Buchner, & Lang, 2009). However, no instructions in the manual have been clearly documented of how this should be done to get meaningful results (Dien, 2017). Judd, Westfall, and Kenny (2017) proposed a different approach to calculating the effect sizes for ANOVAs with within-subject designs and counterbalanced lists. Analogous to a linear mixed model analysis, Judd's et al. (2017) approach takes account of subjects and items as

two random factors in calculations in contrast to the traditional approach which mainly focuses on subjects as the random factor. Alternative to the above approaches, the Bayes factor could be calculated. In a Bayesian analysis, the Bayes factor is a ratio reflecting the relative support of the alternative hypothesis and the null hypothesis by weighing the probabilities of two hypotheses, e.g., $BF_{10} = \frac{p(D|H1)}{p(D|H0)}$, weighing the probability of the alternative hypothesis against the probability of the null hypothesis.[14] Therefore, the Bayes factor is a reflection of a direct comparison of two hypotheses. In this case, a BF (Bayes Factor) smaller than 1 would provide more support for the null hypothesis than the alternative hypothesis (Nicenboim & Vasishth, 2016). However, conventionally, a BF smaller than 0.3 provides support for the null hypothesis, a BF larger than 3 provides support for the alternative hypothesis, whereas a BF between 0.3 and 3 supports neither hypothesis, which would indicate insensitive data (Dienes, 2014). The BF of the two experiments reported in this thesis were obtained with the BayesFactor package in R (Morey & Rouder, 2016). For the reaction times analyses of Experiment 2, the BF of the effect of Tone was 0.395 in the by-subject analysis and 0.558 in the by-item analysis, and the BFs of Segment and the interaction between Segment and Tone were all smaller than 0.3. In the error rates analyses of Experiment 2, the BF of Segment was 0.750 in the by-subject analysis and 0.434 in the by-item analysis, the BFs of Tone and the interaction between Segment and Tone were all smaller than 0.3. In Experiment 3, all BFs were smaller than 0.3 except for BF of the effect of Tone in the error analysis, which was 0.404 in the by-subject analysis 0.366 in the by-item analysis. Thus, based on the Bayes Factors, the marginal results in our analyses do not support the null hypothesis or the alternative hypothesis, but the direction was more towards the null hypothesis (BFs < 1). Therefore, it is unlikely that significant results can be obtained by testing more participants.

Another important question is whether the null findings reported in this chapter suggest that the hidden repetition priming cannot be detected in behavioural data at all. It is crucial to note that the behavioural findings for the hidden repetition effects are not consistent in the literature. Three studies using semantic relatedness judgement task are

[14] Sometimes, researchers use BF_{01}, the reciprocal form of BF_{10}, which weights the probability of the null hypothesis against the probability of the alternative hypothesis.

consistent with this finding of this chapter because these studies did not observe any significant priming effects in the behavioural data (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a). However, one study by the same authors did find a behavioural effect (Thierry & Wu, 2004), which revealed significantly slower response times in semantically unrelated English word pairs with a hidden repeated character (e.g., *train* [火车] - *ham* [火腿]) relative to control pairs. It is important to point out that both Thierry and Wu (2004) and Thierry and Wu (2007) used the same hidden repetition paradigm. However, the character repetition effect in Thierry and Wu (2004) elicited an *increase* rather than a reduction (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a) of the N400 component. Although Thierry and Wu (2007) explained that the enhanced N400 effect observed in Thierry and Wu (2004) is caused by the failure to control for word concreteness in the stimuli, it is important to note that such potential confounds did not modulate ERP responses of native English controls. In addition, no detailed information of word concreteness was provided by Thierry and Wu (2004) or Thierry and Wu (2007), so it is unclear how to explain the differences in behavioural data. In contrast to the experiments using a semantic relatedness judgement task, Zhang's et al. (2011) lexical decision study reported a behavioural evidence of hidden repetition priming effect (e.g., *east* [东] - *thing* [东西]). Critically, unlike Experiment 3, the behavioural evidence of hidden repetition priming in Zhang et al. (2011) was a combination of orthography and phonology because the repeated characters had identical pronunciations (e.g., east [东, /Dong1/] - thing [东西, /Dong1Xi1/]). Therefore, the null findings of Experiment 3, which involved hidden phonological repetition seems to suggest that hidden repetition priming in the behavioural data of Zhang et al. (2011) is not driven by Chinese phonology. This hypothesis needs to be carefully investigated because it contradicts a series of electrophysiological studies in the literature (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a) that concluded that hidden repetition priming is driven by Chinese phonology rather than orthography. In addition, in an unpublished experiment by Zhang (2013), which used the same task and procedure as Zhang et al. (2011), the hidden repetition in Chinese translations revealed a significant *inhibition* effect in phonological repetition (e.g., *throw* [投, /Tou2/] - *head* [头, /Tou2/]), whereas no effect was found in the character repetition (e.g., *face* [面, /Mian4] - *noodle* [面, /Mian4/]). Taken together, the findings of studies using the hidden repetition paradigm are mixed.

To summarize, it seems that the hidden repetition effect is more stable in electrophysiological data than in behavioural data. Thus, electrophysiological measures might be a better tool than behavioural measures to investigate the hidden repetition effect. Therefore, the question whether the hidden phonological repetition priming is driven by segment and/or tone is investigated in the next chapter (Chapter 4) using the electrophysiological measures. Because of the contrast between the results of Experiment 3 and Zhang et al. (2011), further behavioural investigations will be reported in Chapter 5.

Chapter 4 Phonological Activation of Chinese Translation during English Word Reading: Electrophysiological Investigations

4.1 Introduction

The behavioural studies in the previous chapter investigated the activation of segmental and tonal information of Chinese translations during English word reading in an attempt to clarify the source of hidden repetition priming observed in Chinese-English bilinguals (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a). These behavioural investigations also manipulated the type of task to examine whether Chinese activation during English word reading is task-dependent. Experiment 2 used an unmasked hidden priming paradigm with a semantic relatedness judgment task (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a), whereas Experiment 3 combined the masked priming paradigm with a lexical decision task (Zhang et al., 2011). Unexpectedly, both experiments failed to observe any reliable behavioural effects of the hidden repetition priming.

Electrophysiological measures might be a better tool than behavioural measures to investigate the hidden repetition priming because the hidden repetition priming seems to be more stable in electrophysiological data than in behavioural data. As discussed in Chapter 1 (Section 1.4), several electrophysiological studies have reported a reduced N400 as the electrophysiological index of hidden repetition priming. For example, in their seminal study, Thierry and Wu (2007) presented Chinese-English bilinguals with English word pairs (e.g., *train* - *ham*) and asked them to judge whether these word pairs were related in meaning or not. Unknown to the participants, the Chinese translation of the critical English pairs had a repeated character (e.g., the Chinese translation of train and ham is 火车 and 火腿 respectively, with a repeated character 火). Although no behavioural effect was found for the hidden character repetition, the ERP data revealed evidence for the activation of Chinese translations in terms of a reduced N400 for the repeated character condition compared to the non-repeated control condition (e.g., *apple-table*, whose Chinese translations are 苹果 and 桌子 respectively). Critically, no

such difference was observed in the ERP data of native English speakers. Taken together, Thierry and Wu's (2007) study provided a strong evidence for non-selective access in a purely L2 context. In their follow-up study, Wu and Thierry (2010) further disassociated between phonological and orthographic activation of Chinese translations. Using the same hidden translation repetition priming paradigm, Wu and Thierry (2010) observed a reduced N400 for the hidden phonological repetition (e.g., *experience* [经验, /**Jing1**Yan4/] - *surprise* [惊讶, /**Jing1**Ya4/] but not for the orthographic repetition (e.g., *accountant* [会计, /Kuai4Ji4/] - *conference* [会议, /Hui4Yi4/]. These results suggest that Chinese-English bilinguals activate the Chinese phonology rather than the orthography of the Chinese translations. A N400 reduction was also found in another study by the same authors which manipulated the emotional valence of prime words (Wu & Thierry, 2012a). In addition, Chinese phonological activation during English word reading was also observed in a rhyming judgment experiment (Wu & Thierry, 2011) and other experiments using a non-linguistic task involving visual decisions (Wu et al., 2013; Wu & Thierry, 2012b), for example, judging whether each presented stimulus is a circle or a square.

As mentioned in previous chapters, Chinese is a tone language in which the phonological component of a Chinese character, which corresponds to a syllable, consists of segments (consonants and vowels) and lexical tone (for more details about Chinese phonology, see Section 1.8.1). Therefore, it is unclear yet whether the segmental and/or tonal information of the Chinese phonology is activated when Chinese-English bilinguals read English words. This question is related to the central question in bilingual language processing, namely, what representations of L1 are activated during visual word recognition in L2.

To investigate exactly what kind of Chinese phonological information is activated during English word reading, the first experiment of this chapter (Experiment 4) used a semantic relatedness judgment task with English word pairs (prime-target) in which the repetition of segmental and tonal information of the hidden Chinese translations (i.e., the first character of Chinese translations for the English words) was systematically manipulated in semantically and orthographically unrelated English word pairs. If Chinese phonological activation during English word reading is driven

by the combined activation of both segment and tone, the N400 reduction would be observed only when both segment and tone are repeated. The second possibility is that Chinese phonological repetition is only driven by the activation of segmental information. In that case, the N400 reduction should be obtained as long as the segment is repeated and independent of tonal repetition. Similarly, if tonal information is activated of the Chinese translations, the tonal repetition, regardless of the status of the segments, will attenuate the N400.

In addition to investigating ERPs, the current chapter also focussed on the time-frequency aspects of the electroencephalographic (EEG) data. The time-frequency data is usually referred to as oscillatory EEG or event-related oscillations (EROs). An advantage of EROs is that it might provide additional information about phonological activation because EROs can reveal non-phase-locked signals which conventional event-related potentials (ERPs) cannot capture (Cohen, 2014; Herrmann, Grigutsch, & Busch, 2005; Herrmann, Rach, Vosskuhl, & Strüber, 2014). For example, if the activation of Chinese tone is captured by non-phase-locked signals but the activation of Chinese segment is captured by phase-locked signals, only the segment activation will impact the ERP amplitude. According to Cohen (2014, p. 98), frequencies above 15 Hz are usually non-phase-locked. In addition, language studies often apply low-pass fillers on the EEG data (e.g., 25 Hz in Wu and Thierry, meaning that signals above 25 Hz were filtered out). Therefore, ERP analyses do not capture the complete picture of the electrophysiological signals.

Oscillations observed on the scalp EEG are assumed to reflect underlying neuronal synchronization (Bastiaansen & Hagoort, 2006; Bastiaansen, Mazaheri, & Jensen, 2011). Changes in EROs in response to linguistic manipulations have been reported in almost all major frequency bands (theta: 4-7 Hz, alpha: 8-12 Hz, lower beta: 15-20 Hz, higher beta: 20-30 Hz, gamma: above 30 Hz). Evidence has accumulated that distinct aspects of language processing lead to changes at different frequency bands (for a review, see Bastiaansen & Hagoort, 2006; Bastiaansen et al., 2011). For example, in a recent study, Bastiaansen and Hagoort (2015) reported a decrease in the beta-band power for syntactically anomalous sentences compared to syntactically correct sentences, whereas a decrease in the gamma-band power was elicited by semantically anomalous sentences relative to semantically correct sentences. These results reveal a

frequency-based segregation of syntactic and semantic processing. Following this frequency-based segregation approach, it is expected that if there is a functional dissociation between Chinese segmental and tonal processing during English word reading, segmental and tonal repetition may elicit different changes in EROs.

From the description above, it is obvious that the interpretations of time-frequency data and ERP data are very different. However, the preprocessing steps for time-frequency analyses and the conventional analysis of ERPs are very similar (e.g., re-reference, segmenting into epochs, remove artifacts, but not applying low-pass filter at 25 Hz if the focus is on oscillations above 25 Hz). After these preprocessing steps, the power spectra at different frequencies are extracted from all single trials by using wavelet transformations (for other methods used in time-frequency analyses, see Cohen, 2014). The power spectra are then averaged across trials for each condition of each participant. It is important to note that the value of power spectra at a given time point (e.g., 500 ms) can be seen as a weighted average of values within a certain time window centred at that point (e.g., 200 to 800 ms). For example, in wavelet transformations, the length of the time window is determined by the number of cycles, (e.g., the length of the time window at each frequency is 7/(2×frequency) second when using 7 cycles). Therefore, longer epochs are often desirable in time-frequency analyses because it is not possible to obtain values at 1000 ms post-stimuli if the epoch ended at 1000 ms post-stimulus. In addition, because of this, so called, temporal leakage issue, the baseline intervals used in time-frequency analyses often end before the stimulus onset (e.g., -400 to -200 ms) to avoid capturing signals at the stimulus onset (for further discussion of this issue, see Cohen, 2014, p. 64). This is different from ERP analyses which often use baseline intervals ending at 0 (e.g., -100 to 0 ms).

Although conducting time-frequency analyses to investigate EROs are a potential insightful approach that compliments the conventional analysis on ERPs, it is important to be aware of its limitations. First of all, the temporal resolution of EROs is generally poor relative to ERP data (for more discussion, see Cohen, 2014). Therefore, the onset of changes in EROs cannot be taken as the onset of the mechanisms underlying different experimental conditions (Hagoort et al., 2004). Secondly, there is no one-to-one mapping between a given frequency band and a single cognitive activity. For example, in addition to semantic violations, the gamma-band power is also modulated by world

knowledge conflicts (Hagoort et al., 2004), repetition priming and homophone priming (Matsumoto & Iidaka, 2008), and semantic priming (Mellem, Friedman, & Medvedev, 2013). Thus, the specific cognitive functions underlying oscillations at different frequency bands is still unclear. Therefore, more investigations of EROs are needed to understand the underlying neural mechanism of language processing.

Time-frequency analyses have been applied in only a few bilingual studies, such as sentence reading (Kielar, Meltzer, Moreno, Alain, & Bialystok, 2014), word production (Grabner, Brunner, Leeb, Neuper, & Pfurtscheller, 2007), and language learning (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Küssner, de Groot, Hofman, & Hillen, 2016; Kaan, Wayland, & Keil, 2013). However, as far as I know, there are no papers in the literature on bilingual word recognition that have reported time-frequency data. Therefore, I did not have prior knowledge about which frequency bands and which time window and electrodes would be sensitive to the effect of Chinese phonological activation during English word reading. Therefore, the time-frequency analysis should be considered as "exploratory" rather than "confirmatory" (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

A critical question for this exploratory approach is how to make sure that the statistical analyses are sensitive enough to reveal reliable findings and also control the chance of "fooling" oneself (Feynman, 1974; Nuzzo, 2015). A commonly used approach in neuroscience is to run statistical analyses on where (e.g., at which electrodes and/or which frequency) and when (e.g., at which time window) the contrast between two conditions is most pronounced. This kind of circular analysis approach has been condemned as "double dipping" (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). In addition, a recent simulation study has demonstrated that this approach is inappropriate because it dramatically increases Type I errors (Kilner, 2013). Luck and Gaspelin (2017) also pointed out that a visual inspection of the observed outcome should be avoided. Therefore, the circular analysis approach is clearly not in line with current publication guidelines for psychophysiological research, which explicitly states that "[m]easurement windows and electrode sites must be well justified and avoid inflation of Type I error rate" (Keil et al., 2014, p. 7).

An approach that has been recommended when there is no prior knowledge of a potential effect, is the mass univariate approach (Groppe, Urbach, & Kutas, 2011a, 2011b; Maris & Oostenveld, 2007). In this chapter, I used therefore the cluster-based random permutation test implemented in FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2010), which effectively controls for Type I errors across multiple comparisons while maintaining enough sensitivity to detect differences between conditions (see the statistical analysis section for more details). However, it is possible that Chinese segmental and tonal repetition during English word reading induces only a very small effect in the time-frequency data, which may therefore not be easy to detect using the cluster-based random permutation test with the three dimensions (i.e., frequencies, electrodes, times). Therefore, to increase the sensitivity of the analyses and to obtain prior knowledge of which frequency bands to focus the analyses, the second experiment reported in the current chapter (Experiment 5) was used. This experiment focused on Chinese phonological repetition in Chinese words with Chinese native speakers and used a Chinese semantic relatedness judgment task in which a different group of Chinese-English bilinguals read the Chinese translations of English word pairs used in Experiment 4. A cluster-based random permutation test was first run on the data of Chinese word reading by including all the frequencies. Although using all three dimensions (i.e., frequencies, electrodes, times) in the permutation test is a relatively conservative method, Chinese phonological repetition effects in Chinese word reading (Experiment 5) should be large enough to be observed in the time-frequency analysis. The significant frequency bands in the Chinese data were then used to focus on in the English data (Experiment 4).

Thus, the ERO results of Experiment 5 were used to guide the ERO analysis of Experiment 4. Experiment 5 investigated Chinese phonological activation during Chinese word reading. As reviewed in Section 1.8.3, previous electrophysiological studies on phonological activation in visual Chinese word/character recognition have revealed mixed findings. Some studies observed a N400 effect (e.g., Wu & Thierry, 2010, 2012a), whereas others reported a P200 effect (e.g., Kong et al., 2010; Zhang et al., 2009). For example, in a Chinese semantic relatedness judgment experiment, Wu and Thierry (2010) found a reduced N400 when two-character Chinese word shared the same pronunciation (e.g., 经验 [experience] - 惊讶 [surprise], /Jing1Yan4/ - /Jing1Ya4/)

relative to the control condition. Using the same task, Kong et al. (2010) found a larger P200 amplitude when a Chinese character was preceded by its homophone (e.g., 歉 - 嵌, /Qian4/ - /Qain4/). Furthermore, there are, as far as I am aware, no Chinese word recognition studies that investigated the activation of both segment and tone using EEG, and only a handful of behavioural investigations have been conducted (Li et al., 2013; Spinks et al., 2000). Although the empirical studies on segment and tone in Chinese word recognition are scarce, computational models of Chinese character recognition have generally coded segment and tone separately (Perfetti et al., 2005; Zhao & Li, 2009, 2013), suggesting that there is a dissociation between segment and tone.

To summarize, the current chapter reports two electrophysiological experiments that were conducted with Chinese-English bilinguals reading English (Experiment 4) or Chinese (Experiment 5) word pairs. The experiments involved a semantic relatedness judgment task with word pairs in which the repetition of segmental and tonal information of the hidden Chinese translations (Experiment 4) or of the Chinese words (Experiment 5) were systematically manipulated in semantically and orthographically unrelated word pairs. To advance the knowledge of the neural basis underlying Chinese phonological activation during English and Chinese word reading, time-frequency data of both experiments are also reported (Section 4.4).

4.2 Experiment 4: English Semantic Relatedness Judgment Task

4.2.1 Method

4.2.1.1 Participants

32 (Mean Age: 23.25, Male: 17) Chinese-English bilinguals participated in the experiment and they all received an inconvenience allowance for their participation. One participant was excluded due to high error rates (>25%), and one participants was excluded due to excessive eye blinks. A further eleven participants were excluded from the analysis due to slow drifts because the data were recorded during a hot summer in a lab without air-conditioning. A minimum of 40 epochs were required for each condition and the analyses were based on 19 participants. The bilingual participants

were first exposed to English at the age of 10 (range: 7-14). By the time of testing, they had learnt English for an average of 13 years (range: 9-20) and had lived in the United Kingdom for an average of 17.5 months (range 4-46). The bilingual participants were all native Chinese speakers who learnt Chinese from birth and they were undergraduate or postgraduate university students studying in Nottingham. They met the minimum English language entry requirements to study at the University of Nottingham (IELTS 6.0 for undergraduates and 6.5 for postgraduates). Their English proficiency was objectively measured by LexTALE (Lemhöfer & Broersma, 2012), which is an English vocabulary test. The average LexTALE score of the participants was 58.2 (range 43.75-77.5). Their self-rating score of English reading proficiency on a 7-point scale (1 = very poor, 7 = native-like) was 5 (range 4-7). All the participants had normal or corrected-to-normal vision and had no history of neurological or psychiatric impairment. According to an adapted version of Coren's (1992) handedness questionnaire, all the participants were right-handed.[15]

4.2.1.2 Materials and Design

Using the data from the English-Chinese translation norming study reported in Chapter 2, 240 experimental words pairs with similar word frequency, length and concreteness were selected to create four experimental conditions: +*Segment* +*Tone*, +*Segment* -*Tone*, -*Segment* +*Tone* and -*Segment* -*Tone*. Examples of stimuli are presented in Table 13 (see Appendix 4 for the list of all stimuli). All critical 240 word pairs were unrelated in meaning, whereas 60 filler pairs were semantically related.

[15] The questionnaire is available at http://faculty.washington.edu/chudler/rltablen.html

Table 13. Experimental design and stimulus examples in Experiment 4

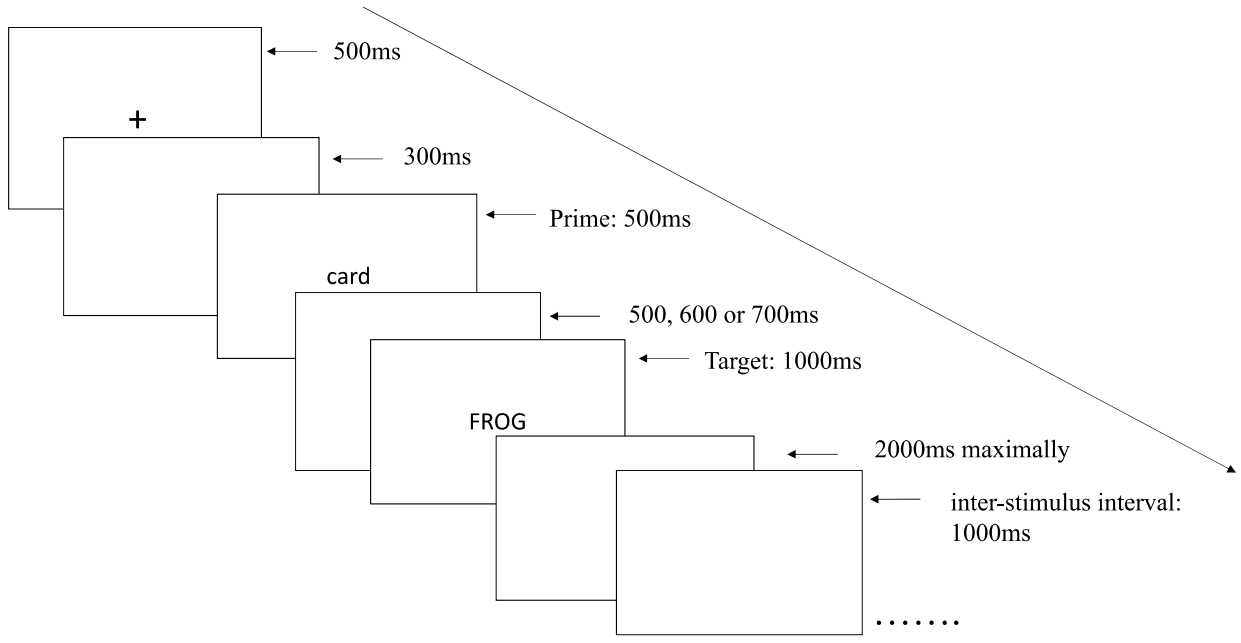
	Experimental Trials (Semantically Unrelated)								**Filler Trials**	
	+Segment +Tone		*+Segment -Tone*		*- Segment +Tone*		*-Segment -Tone*		**(Semantically Related)**	
English	media - rose		mirror - police		milk - file		card - frog		hobby	interest
Chinese Translation	媒体	玫瑰	镜子	警察	牛奶	文件	卡片	青蛙	爱好	兴趣
Pinyin	***Mei2***Ti3	***Mei2***Gui4	***Jing***4JZi3	***Jing***3Cha2	Niu***2***Nai3	Wen***2***Jian4	Ka3Pian4	Qing1Wa1	Ai4Hao4	Xing4Qu4
Semantic Scores (English)	1.24(0.21)		1.21(0.24)		1.23(0.25)		1.32(0.32)		4.34(0.24)	
Semantic Scores (Chinese)	1.19(0.19)		1.16(0.20)		1.15(0.19)		1.21(0.24)		4.36(0.26)	
Word Length	5.38(1.57)	5.77(1.58)	6.13(1.63)	5.73(1.51)	6.08(1.85)	6.10(1.60)	6.37(1.92)	5.75(1.71)	6.15(2.15)	5.72(1.90)
English Frequency	4.62(0.66)	4.67(0.57)	4.49(0.55)	4.52(0.60)	4.47(0.57)	4.45(0.60)	4.41(0.65)	4.56(0.48)	4.54(0.54)	4.80(0.56)
Chinese Frequency	4.49(0.64)	4.58(0.61)	4.40(0.63)	4.47(0.70)	4.47(0.50)	4.48(0.66)	4.48(0.45)	4.52(0.53)	4.31(0.67)	4.15(0.68)
Concreteness	3.94(1.00)	3.59(1.07)	3.86(1.06)	3.87(1.00)	3.99(0.98)	3.75(1.08)	3.70(1.05)	3.88(1.06)	3.90(1.08)	4.07(0.98)

Note. Pinyin is the alphabetic transposition of Chinese phonological form. Numbers for the Pinyin translations indicate tone. There are four different tones in Mandarin Chinese, including the high level tone (1), the rising tone (2), the falling-rising tone (3) and the falling tone (4). The English frequency is based on SUBTLEX-UK Zipf values (van Heuven et al., 2014) and Chinese frequency is based on SUBTLEX-CH Zipf values (Cai & Brysbaert, 2010). Concreteness score is based on Brysbaert et al. (2014). Semantic Scores of English word pairs and their Chinese equivalent pairs were obtained using from a 5-point semantic rating study with different group of 20 Chinese-English bilinguals to make sure that all the critical word pairs are not related in meaning.

4.2.1.3 Procedure

Participants were tested individually in a dimly-lit shielded experimental room. Each participant received a unique random presentation order of the stimuli. Stimuli were presented on a CRT monitor (17" Liyama Vision Master 400, 1024×768 pixels, 85 Hz) controlled by PsychoPy (Peirce, 2008). The sequence of visual events for each trial is presented in Figure 24. Each trial began with a fixation cross, presented for 500 ms followed by a 300 ms black screen. Next, the lowercase prime word was presented for 500 ms, followed by a variable interstimulus interval of 500, 600 or 700 ms (Thierry & Wu, 2007; Wu & Thierry, 2010) after which the uppercase target word with a fixed presentation time of 1000 ms was presented. Participants had to decide whether the presented word pairs (prime-target) were related in meaning or not by pressing buttons on a Cedrus response box attached to the computer. They were instructed to perform the task as quickly and as accurately as possible. If participants did not respond during the presentation of the target, a blank screen would appear up to a maximum of 2000 ms or until the participants responded. Response sides were counterbalanced between participants. The inter-trial interval was set at 1000 ms. The experiment was approved by the Ethics Committee in the School of Psychology at the University of Nottingham. Participants all signed a consent form before the experiment started.

Figure 24. Illustration of visual events in Experiment 4



All participants made semantic relatedness judgments to 300 English word pairs. Prior to the experiment, a practice block of 15 trials was used to familiarize the participants with the procedure. Participants were instructed to minimize blinks, eye-movements and body movements when the stimuli were presenting. Participants were prompted to have a rest after 60 trials. After the semantic relatedness judgment task, participants conducted an English vocabulary test (LexTALE, Lemhöfer & Broersma, 2012) and completed a language history questionnaire. The entire session lasted about 2 hours.[16]

4.2.1.4 EEG Recordings

The electroencephalogram (EEG) was recorded continuously at a sample rate of 1024 Hz using a Biosemi ActiveTwo system with 64 Ag/AgCl active electrodes placed according to the 10/20 system. Two additional electrodes placed close to Pz, namely, the Common Mode Sense [CMS] active electrode and the Driven Right Leg [DRL] passive electrode, were used for online referencing (for detailed explanations, see Metting van Rijn, Peper, & Grimbergen, 1990; Schutter, Leitner, Kenemans, & van Honk, 2006). In addition to the 66 electrodes on the scalp, 7 external electrodes were applied: two over left and right mastoids for off-line re-referencing, two at the outer canthus of each eye, two below each eye and one above the left eye for monitoring eye blinks and movements. For each measurement site, the voltage offset from the CMS was below 25 uV.

4.2.2 Data Analysis

4.2.2.1 Preprocessing

Data were preprocessed and analysed using FieldTrip (Oostenveld et al., 2010). The continuous EEG was filtered offline with a high-pass filter set at 0.1 Hz. The EEG data was then re-referenced off-line to the averaged mastoids and segmented in epochs of 3000 ms beginning with 1000 ms prior to the onset of the second word. Trials with

[16] After completing the English semantic relatedness judgement task, the participants also conducted a Chinese semantic relatedness judgement task in which the Chinese translation of English words were presented. However, because the N400 component is also sensitive to repetition and the poor quality of the EEG data in this Chinese section due to relatively prolonged testing, the Chinese data is not reported in this thesis.

incorrect responses or RTs less than 300 ms or more than 2,500 ms were discarded (2.8% of the data). Epochs contaminated by drifts and muscle activities were first manually dismissed (8.4% of the data), and eye-blinks and eye movements were corrected using an independent component analysis (ICA) algorithm (Jung et al., 2000). The epochs with potentials greater than 100 μV were automatically rejected (16.0% of the data). For the data of each participant to be included in the analysis, a minimum of 40 epochs were required for each condition. The preprocessed data were then used for the ERP analysis and the time-frequency analysis reported in Section 4.4.

4.2.2.2 ERP Analysis

The clean EEG data was re-segmented in epochs of 1100 ms beginning with 100 ms prior to the onset of the second word. Baseline correction was applied, using the interval from -100 to 0 msec. The epochs were low-pass filtered at 30 Hz. Finally, the average of the ERPs was computed for each participant in each of the experimental condition.

4.2.2.3 Statistical Analysis

The ERP data analysis focused on the N400 component (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a). Nine electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz) which showed the maximal effect in Thierry and Wu (2007) were selected for the analysis. The mean global field power (Lehmann, 1987; Lehmann & Skrandies, 1980; Picton et al., 2000) was calculated across all the electrodes and all semantically unrelated conditions (*+Segment +Tone, +Segment -Tone, -Segment +Tone*, *-Segment -Tone*) to detect the latency of the peak amplitude in the N400 window from 350 ms to 500 ms (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a). The peak latency was 392 ms. Next, the mean amplitude of the time window which extended ±25 ms surrounding the peak (367 ms to 417 ms) was calculated. The mean amplitude of the next 50-ms time window (417 ms to 467 ms) was also calculated. A 2 (Segment: +Segment vs. -Segment) × 2 (Tone: +Tone vs. -Tone) × 9 (Electrodes: FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz) repeated measures ANOVA was conducted for the mean amplitudes in both time windows. This approach aims to capture the fine-grained changes in the N400 component, which is analogous to using pairwise t-tests for each sampling point in previous studies (e.g., Thierry & Wu, 2007; Wu & Thierry, 2010). Greenhouse-Geisser was applied whenever applicable.

4.2.3 Results

4.2.3.1 Behavioural Data

Table 14 presents a summary of the mean error rates and reaction times for the 4 experimental conditions. A 2 (+Segment vs. -Segment) × 2 (+Tone vs. -Tone) repeated measures ANOVA was run for both error rates and reaction times. The reaction time analyses revealed a significant main effect of Segment, F (1,18) = 8.968, p < .01, BF = 1.216, no effect of Tone, F <1, p > .30, BF = 0.308, and no interaction between Segment and Tone, F = 2.176, p = .157, BF = 0.375. No effects were found in the error analyses, all Fs < 1, ps > .40, BFs < 0.33.

Table 14. Mean error rates and reaction times (with SD in brackets) of Experiment 4

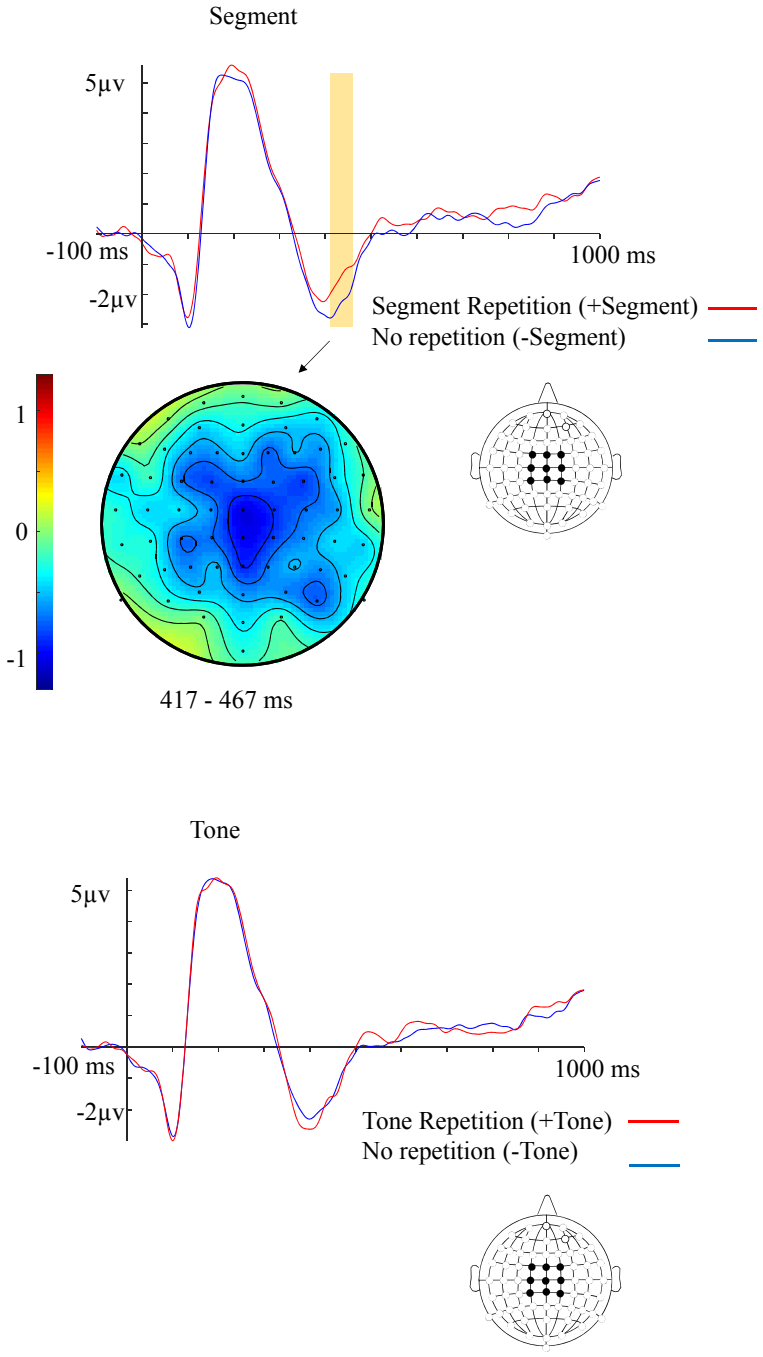
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*
Error Rates (%)	2.11	2.28	1.58	2.11
	(2.14)	(1.59)	(1.52)	(2.60)
Reaction Time (ms)	966	947	967	974
	(247)	(245)	(248)	(250)

4.2.3.2 ERP Data (N400 time window)

Between 367 - 417 ms, analyses of the mean amplitudes revealed no significant main effects of Segment, F (1,18) = 2.409, p = .138, and Tone, F (1,18) = 2.234, p = .152, and no significant interaction between Segment and Tone F < 1, p > .50. However, between 417 - 467 ms, the analysis of the mean amplitudes revealed a significant effect of Segment, F (1,18) = 6.099, p = .024, indicating a reduced N400 for segmental repetition (see Figure 25). The main effect of Tone and interaction between Segment and Tone were not significant, Fs < 1, ps > .70.[17]

[17] Similar results were also found when using the cluster-based random permutation test implemented in FieldTrip (for details Section 4.3.2.4) to the window from 350 ms to 500 ms. For the effect of Segment, a significant cluster (p = .046) in the time window from 400 to 460 ms was found. There were no significant clusters found for Tone or for the interaction between Segment and Tone.

Figure 25. Results of ERP analysis of critical trials in Experiment 4. ERPs time-locked to the onset of target words averaged across 9 electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz, all shaded in the schematic head) on the scalp with topography of ERP effect (-*Segment* minus +*Segment*)



4.2.4 Discussion

In Experiment 4, behavioural and electrophysiological data was collected from Chinese-English bilinguals in an English semantic relatedness judgement task. Critically, participants were not aware that the phonology of the Chinese translations of the semantically unrelated English pairs were repeated at the segmental and/or tonal level. The behavioural results revealed that when the segment of Chinese translations was repeated, participants were significantly faster in responding. Tone repetition did not impact the behavioural responses. The behavioural segmental repetition effect found in Experiment 4 was unexpected because Experiment 2 reported in Chapter 3 that used the same paradigm and involved the same manipulations did not find any behavioural effects. A possible explanation is that the conditions involved 60 trials each in Experiment 4 but only 15 trials in Experiment 2. As discussed in Section 1.3.4, using more trials may increase priming effects by reducing the noise in the data (Van den Bussche et al., 2009), therefore it is likely that Experiment 4 was able to detect a small significant priming effect (14 ms) in the behavioural data, whereas Experiment 2 could not. However, the Bayes factor of this small behavioural segment repetition effect was larger than 1 but smaller than 3, suggesting that the alternative hypothesis was not strongly supported. Therefore, a further replication study would be needed to investigate whether the behavioural segmental repetition effect observed in Experiment 4 is robust or fragile.

In line with the behavioural data, the ERP data also revealed a significant main effect of Segment. When segmental information of Chinese translation was repeated (+Segment, i.e., *+Segment +Tone, +Segment -Tone*), the ERP amplitude was reduced in the time window of 417 - 467 ms. Consistent with the behavioural data, no effect of Tonal repetition was found in the ERP data. The hidden segmental repetition effect appeared however later and lasted for a shorter period than the phonological repetition effect in Wu and Thierry (2010), i.e., 310 - 420 ms. A possible explanation is that the Chinese phonological repetition effect in Wu and Thierry (2010) only involved both segmental and tonal repetition, so the effect might therefore be stronger. To test this possible explanation, a repeated measures ANOVA was conducted for the contrast between *+Segment +Tone,* and *-Segment –Tone* using the same pre-selected 9 electrodes for the mean amplitudes between 310 and 420 ms. The results revealed

however no significant effect of Chinese phonological repetition within that time window. Therefore, the phonological repetition effect in Experiment 4 was indeed weaker than the effect reported in Wu and Thierry (2010).

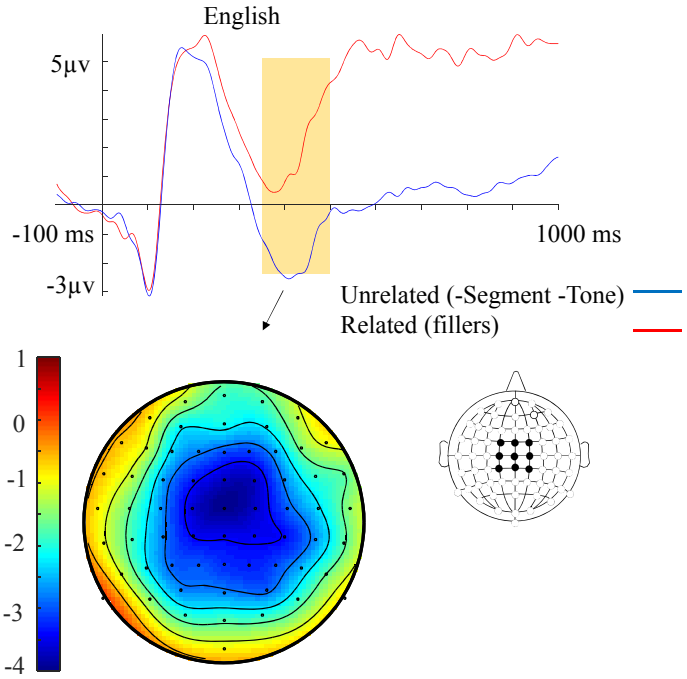
To understand why the ERP response was weaker, I carefully looked at the differences between Experiment 4 and the experiment reported in Wu and Thierry (2010). Experiment 4 in this thesis differs from Wu and Thierry (2010) in how EEG was recorded and how the data was analysed (e.g., different EEG recording systems, different filters applied, the on-line reference channel and off-line re-reference measures). However, it is not likely that these differences are the key issue. Take for example the difference in re-referencing of the EEG data. The EEG was re-referenced to the average reference (the average of all channels on the scalp) in Wu and Thierry (2010), whereas the average mastoids reference was used in Experiment 4. Although different re-referencing methods indeed impact ERP data (Luck, 2014), it is more common in the literature to use the average mastoids reference in studies focusing on the N400 component (Ito, Martin, & Nieuwland, 2016) because the average mastoids reference can maximize its effect (Dien, 2017). Therefore, it is unlikely that this difference would cause a weaker effect in Experiment 4.

The second possible explanation is that the ratio of unrelated trials (240 critical trials) to related trails (60 fillers) in Experiment 4 (4:1) lead to a weaker effect. This explanation is also unlikely. Although the ratio of unrelated to related trails was lower in Wu and Thierry (2010) (3:1) than in Experiment 4 (4:1), Wu and Thierry (2012a) also found the Chinese phonological repetition effect between the 300-500 time-window using the same ratio of 4:1. Furthermore, the ERP data of the filler trials were also analysed, and these revealed a classic N400 reduction in related trials compared to unrelated trials of -*Segment* -*Tone* (Figure 26).[18] These results indicate that participants processed the English words up to the semantic level, thus ruling out the potential concern over the unequal number of unrelated and related trails. Thus, it is unlikely that

[18] A 2 (related fillers vs. unrelated -*Segment* -*tone*) × 64 electrodes repeated measures ANOVA was conducted for the mean amplitudes between 350 to 500 ms. The results revealed a main effect of relatedness, $F(1,18) = 31.153$, $p < 0.001$.

differences in the ratio of unrelated trails to related trials could explain a weaker effect in Experiment 4.

Figure 26. Results of ERP analysis for semantic related (fillers) vs. unrelated trials (-*Segment* -*Tone*)) in Experiment 4. ERPs time-locked to the onset of target words averaged across 9 electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz, all shaded in the schematic head) on the scalp with topography of ERP effect (-*Segment* -*Tone* minus related fillers)



It is also unlikely that the weaker effect is due to insufficient number of participants. The analysis of Experiment 4 included 19 participants, whereas only 15 participants were tested in Wu and Thierry (2010). As Cohen (2014, p. 502) pointed out, testing around 20 participants is powerful enough to detect both small and large effects in ERP data.

However, the differences in reaction times between two studies might provide a clue to reconcile the dissimilarities in the ERP effects. Participants in Wu and Thierry's (2010) study responded faster (around 900 ms) than participants in Experiment 4 (around 950

ms), although participants in both studies had a similar linguistic background.[19] The behavioural differences may be due to the difference in the way the target words were presented. The target words in Experiment 4 were presented in upper case, whereas the target words in Wu and Thierry (2010) were presented in lower case.[20] Chinese-English bilinguals may need more processing time to recognize English words presented in upper case than in lower case, resulting in slower response times in Experiment 4. Therefore, it is likely that the relatively demanding efforts in processing English words in upper case lead to a weaker activation of Chinese phonology.

Taken together, segmental repetition in Chinese translations of unrelated English pairs was observed in both behavioural and ERP data. This suggests that hidden segmental repetition rather than hidden tonal repetition drives the ERP amplitude reduction observed in previous studies (Wu & Thierry, 2010, 2012a). The next experiment (Experiment 5) used a Chinese semantic relatedness judgment task in which another group of Chinese-English bilinguals read the Chinese translations of English word pairs used in Experiment 4. As previously mentioned, the goal of Experiment 5 is two-fold. First and foremost, conducting the time-frequency analyses on Chinese data (Experiment 5) can provide the knowledge about which frequency bands to focus on when analysing the English data (Experiment 4). Second, it contributes to the investigations of Chinese phonological activation during Chinese word reading.

[19] Chinese-Ensglih bilingual participants in both studies were native Chinese speakers who learnt Chinese from birth and they were undergraduate or postgraduate university students studying in UK. They met the minimum English language entry requirements studying in both univsesities.

[20] Wu and Thierry (2010) did not report that they used lowercase target words, but this was confirmed in a personal communication with Professor Thierry at a conference in Birmingham (June 2014).

4.3 Experiment 5: Chinese Semantic Relatedness Judgment Task

4.3.1 Methods

4.3.1.1 Participants

32 (Mean Age: 24.5, Male: 17) Chinese-English bilinguals participated in the experiment and they received an inconvenience allowance for their participation. Nine participants were excluded from the analysis due to slow drifts because the data were recorded during the summer in a lab without air-conditioning. A minimum of 40 epochs were required for each condition and the analyses are based on 23 participants. Although Chinese-English bilinguals were presented with pairs of Chinese words, their language profiles were also accessed. The bilingual participants were first exposed to English at the age of 10 (range: 7-14). By the time of testing, they had learnt English for an average of 14 years (range: 9-17) and had lived in the United Kingdom for an average of 21 months (range 1-70). The bilingual participants were all native Chinese speakers who learnt Chinese from birth and they were undergraduate or postgraduate university students studying in Nottingham. Their English proficiency was objectively measured using LexTALE (Lemhöfer & Broersma, 2012). The average LexTALE score of the participants was 57.2 (range 46.25-73.75). Their self-rating score of English reading proficiency on a 7-point scale (1 = very poor, 7 = native-like) was 5 (range: 4-6). All the participants had normal or corrected-to-normal vision and had no history of neurological or psychiatric impairment. According to an adapted version of Coren's (1992) handedness questionnaire, all the participants were right-handed.

4.3.1.2 Materials and Design

The materials and design were identical to Experiment 4 except that participants read the Chinese translations of English word pairs. Examples of stimuli are presented in Table 15 (see Appendix 5 for the list of all stimuli). All critical 240 word pairs were unrelated in meaning, whereas 60 Chinese filler pairs were semantically related.

Table 15. Experimental design and stimulus examples of Experiment 5

	Experimental Trials (Semantically Unrelated)								**Filler Trials**	
	+Segment +Tone		*+Segment -Tone*		*- Segment +Tone*		*-Segment -Tone*		**(Semantically Related)**	
Chinese	媒体	玫瑰	镜子	警察	牛奶	文件	卡片	青蛙	爱好	兴趣
Pinyin	***Mei2***Ti3	***Mei2***Gui4	***Jing***4JZi3	***Jing***3Cha2	Niu***2***Nai3	Wen***2***Jian4	Ka3Pian4	Qing1Wa1	Ai4Hao4	Xing4Qu4
Semantic Scores (Chinese)	1.19(0.19)		1.16(0.20)		1.15(0.19)		1.21(0.24)		4.36(0.26)	
Chinese Frequency	4.49(0.64)	4.58(0.61)	4.40(0.63)	4.47(0.70)	4.47(0.50)	4.48(0.66)	4.48(0.45)	4.52(0.53)	4.31(0.67)	4.15(0.68)

Note. Pinyin is the alphabetic transposition of Chinese phonological form. Numbers added to the Pinyin translations indicate tone. There are four tones in Mandarin Chinese: high level tone (1), the rising tone (2), the falling-rising tone (3) and the falling tone (4). Chinese frequency is based on SUBTLEX-CH Zipf values (Cai & Brysbaert, 2010). Semantic Scores of Chinese word pairs were obtained using a 5-point semantic rating study with different group of 20 Chinese-English bilinguals to make sure that all the critical word pairs were not related in meaning.

4.3.1.3 Procedure

The procedure was the same as in Experiment 4 apart from two differences. First, pairs of Chinese words were presented. Second, participants were prompted to have a rest after 30 trials instead of 60 trials because a fan was used more often to cool the lab.[21]

4.3.1.4 EEG Recordings

Identical to Experiment 4.

4.3.2 Data Analysis

4.3.2.1 Preprocessing

The preprocessing procedures were the same as Experiment 4. The preprocessed data were the used for the ERP analysis and the time-frequency analysis will be reported in Section 4.4.

4.3.2.2 ERP Analysis

The ERP analysis is the same as Experiment 4.

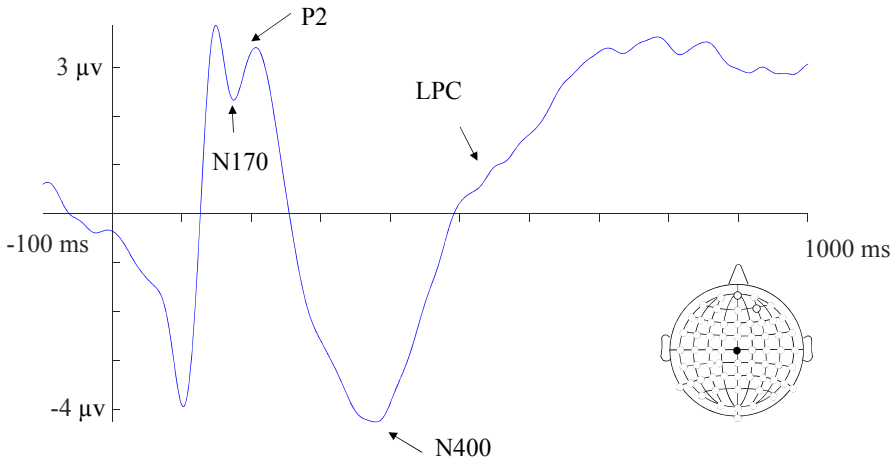
4.3.2.4 Statistical Analysis

The original plan for the ERP data analysis was the same as for Experiment 4. The focus is the P200 component (Kong et al., 2010; Zhang et al., 2009) and the N400 component (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a). Using the mean global field power (Lehmann, 1987; Lehmann & Skrandies, 1980; Picton et al., 2000) measured across all the electrodes and all the unrelated conditions (*+Segment +Tone, +Segment -Tone, -Segment +Tone*, *-Segment -Tone*), the peaks for 100-250 ms and 350-500 were detected at 148 ms and 459 ms respectively. The latencies of the peaks were unexpected because the 148-ms peak was relatively early for the P200 component and the 459-ms peak was relatively late for the N400 component. Based on the latencies of peaks, it is very likely that the P200 component temporally overlaps with an early component and the N400 component temporally overlaps with a late component. It is common that an observed ERP waveform is the combination of more than one underlying components

[21] The data were recorded during a hot summer in a lab without air-conditioning.

(Kappenman & Luck, 2012; Luck, 2014). However, the mean global field power, which is an unbiased measure of electric signal amplitude across the scalp, may fail to capture distinguished peaks for temporally overlapping components (Dien, 2012). To investigate the concern of overlapping components, the ERP waveforms were averaged across all the unrelated conditions (+*Segment* +*Tone,* +*Segment* -*Tone,* -*Segment* +*Tone*, -*Segment* -*Tone*). As can be seen in Figure 27, the P200 component is proceeded by the N170 component and the N400 component is followed by the LPC component.

Figure 27. ERPs time-locked to the onset of target words average across all the unrelated conditions at the Cz electrode with the schematic head depicting the location of Cz.



Because the patterns in the ERP data were more complex than expected and different from previous studies (e.g., Kong et al., 2010; Wu & Thierry, 2010), the findings of previous studies (e.g., Kong et al., 2010; Wu & Thierry, 2010) were not very helpful to guide the analysis. Therefore, the ERP data in Experiment 5 was analysed using the cluster-based random permutation test as implemented in FieldTrip (for details see Maris & Oostenveld, 2007). As mentioned earlier, the cluster-based random permutation test effectively controls for Type I errors across multiple comparisons

while maintaining enough sensitivity to detect differences between conditions. The method involves conducting dependent-sample t-test for every data point (electrode by time). The adjacent data points ($p < .05$) were combined to create clusters. For each cluster, the sum of the t-statistics was calculated at cluster level statistics. A null-distribution was then created by first randomly assigning one of the two conditions within subjects for several times (here 1000 times) and then calculating the largest cluster level statistics for each randomization. Finally, the observed cluster level statistics were compared against the null-distribution, with clusters falling in the highest or lowest 2.5% considered to be significant. For the main effect of Segment and Tone and their interaction, a cluster-based random permutation test was applied from 200 ms to 800 ms post-stimulus.

4.3.3 Results

4.3.3.1 Behavioural Data

Table 16 presents a summary of the mean reaction times and error rates for the 4 experimental conditions. A 2 (+Segment vs. -Segment) × 2 (+Tone vs. -Tone) repeated measures ANOVA was conducted with reaction times and errors. The analysis of the reaction times revealed no significant main effects of Segment and Tone, and no significant interaction, all Fs < 1, ps > .40, BFs < 0.30. The analysis of the error rates also revealed no significant main effect of Segment, F (1, 22) = 2.529, p = .126, BF = 0.523, Tone, F (1, 22) = 1.762, p = .198, BF = 0.515, and no significant interaction, F (1, 22) = 2.007, p = .171, BF < 0.30.

Table 16. Mean error rates and reaction times (with SD in brackets) of Experiment 5

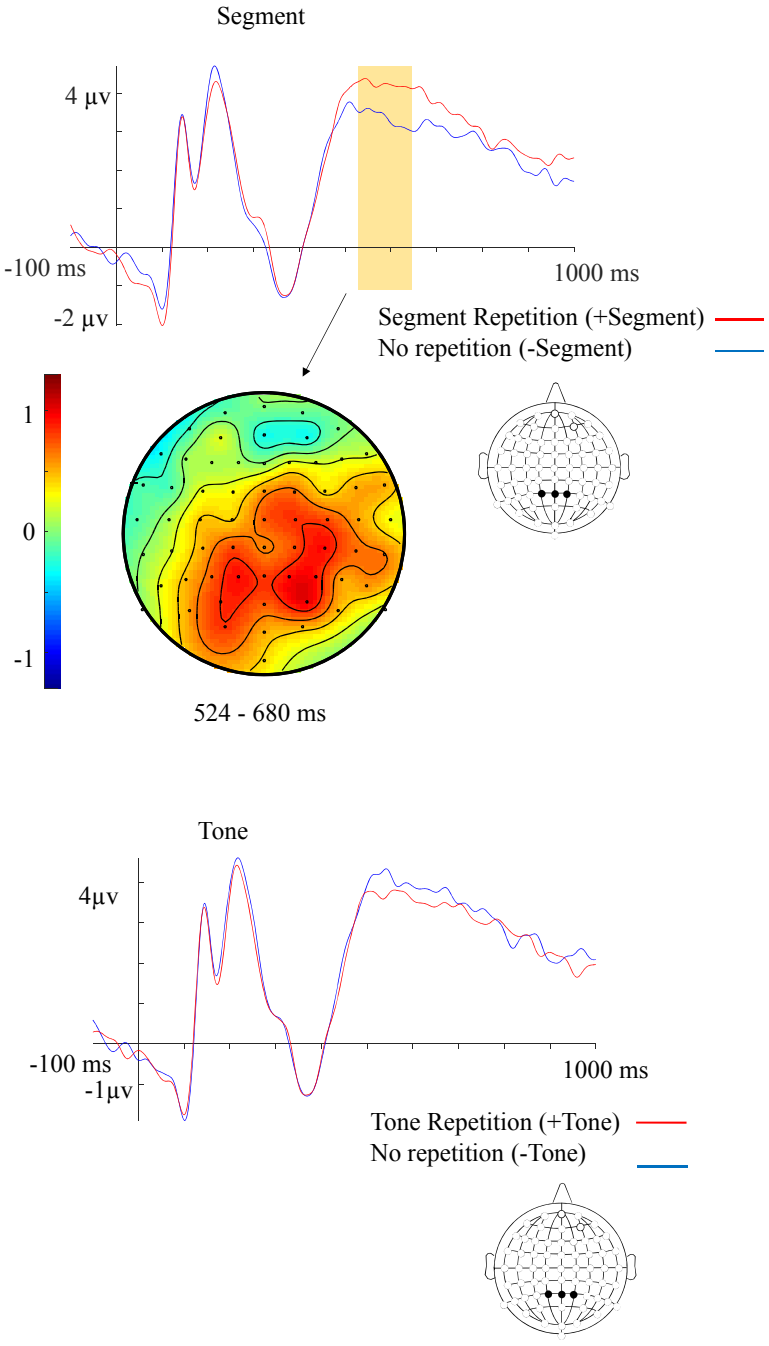
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*
Error Rates (%)	1.00	1.01	1.01	1.81
	(1.93)	(2.23)	(2.23)	(3.55)
Reaction Time (ms)	643	642	643	637
	(130)	(135)	(142)	(137)

4.3.3.2 ERP Data

The results of ERP analysis are presented in Figure 28. For the main effect of Segment, a significant cluster (p = .015) in the time window from 524 to 680 ms was found, which indicates a significant difference between the segmental repetition condition (+*Segment*) and the no repetition condition (-*Segment*). This suggests that the segmental repetition condition elicited a larger LPC compared to the non-repeated condition, and the difference was more pronounced over posterior electrodes (Figure 28). There were no significant clusters for the main effect of Tone or the interaction between Segment and Tone.[22]

[22] In addition to using a large window (200 - 800 ms), the cluster-based permutation test was also conducted for two smaller time windows, i.e., 200 – 500 ms and 500 – 800 ms. The results are very similar. For the main effect of Segment, a significant cluster (p = .006) in the time window from 524 to 680 ms was found, revealing the segmental repetition condition elicited a larger LPC compared to the non-repeated condition. There were no significant clusters for the main effect of Tone or the interaction between Segment and Tone.

Figure 28. Results of ERP analysis of critical trials in Experiment 5. ERPs time-locked to the onset of target words averaged across 3 electrodes (Pz, P1, P2, all shaded in the schematic head) on the scalp with topography of ERP effect (+*Segment* minus -*Segment*)



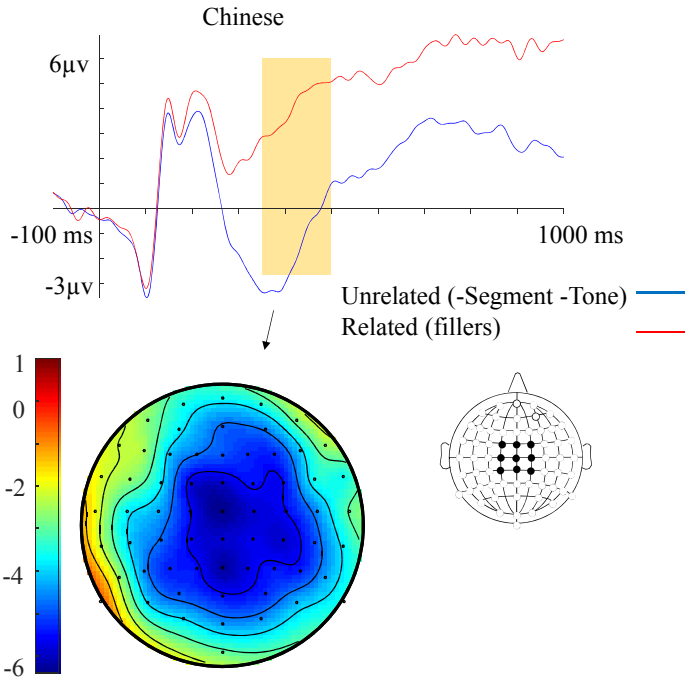
4.3.4 Discussion

In Experiment 5, behavioural and electrophysiological data was collected from Chinese-English bilinguals in a Chinese semantic relatedness judgement task. All the Chinese word pairs presented were Chinese translations of English word pairs used in Experiment 4. Therefore, the phonology of the semantically unrelated Chinese pairs was repeated at the segmental and/or tonal level. In the behavioural data, no significant hidden repetition priming effect was found in either the error rates or reaction times. In contrast, the ERP data revealed that segmental repetition condition elicited a larger LPC compared to the (non-repeated) control condition. An LPC effect has been associated with conflict resolution at late decision stages (Kolk & Chwilla, 2007; Misra & Holcomb, 2003), which suggests that word pairs with segmental repetition might have caused interference for the participants.

Although Experiment 5 indicates that Chinese segmental repetition created an LPC effect, an obvious puzzle is that why the effect is later in Chinese (Experiment 5) than in English (Experiment 4). To confirm that participants in Experiment 5 processed the presented words up to the semantic level, the ERP data of semantically related fillers were also analysed. Similar to Experiment 4, a classic N400 effect was observed in related trials compared to unrelated trials of -*Segment* -*Tone* (Figure 29).[23] Therefore, it is certain that participants in Experiment 5 did process the presented words up to the semantic level.

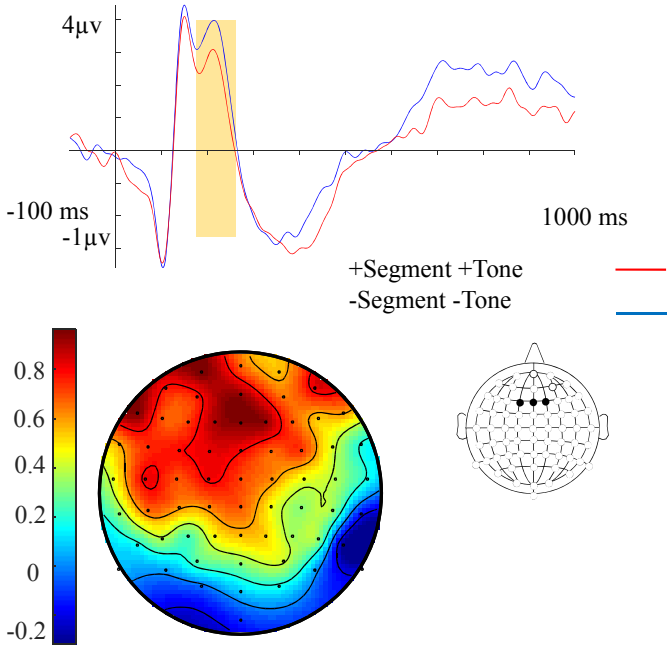
[23] Because Thierry and Wu (2007) used the window between 350 and 500 ms and found the N400 reduction in the semantic related pairs compared to unrelated pairs, the same window was used here to capture the N400 effect. A 2 (related fillers vs. unrelated -*Segment* -*tone*) × 64 electrodes repeated measures ANOVA conducted for the mean amplitudes between 350 to 500 ms revealed a main effect of relatedness, $F(1,22) = 68.698$, $p < 0.001$.

Figure 29. Results of ERP analysis semantic related (fillers) vs. unrelated trials (-*Segment* -*Tone*) in Experiment 5. ERPs time-locked to the onset of target words averaged across 9 electrodes (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz, all shaded in the schematic head) on the scalp with topography of ERP effect (-*Segment* -*Tone* minus related fillers)



To understand the surprising difference between the English and Chinese data, a new analysis was conducted to compare +*Segment* +*Tone,* and -*Segment* -*Tone* because this was the contrast used previous studies which observed either the N400 effect (Wu & Thierry, 2010, 2012a) or the P200 effect (Kong et al., 2010; Zhang et al., 2009). Interestingly, the cluster-based permutation test revealed a significant difference between the +*Segment* +*Tone* and -*Segment* -*Tone* (p = .012, 180 to 250 ms). This finding reveals a significant P200 reduction in +*Segment* +*Tone*, which was most pronounced over frontal and central electrodes (Figure 30). However, it is unclear why the phonological repetition elicited an enhanced P200 in previous studies (Kong et al.,

Figure 30. Results of ERP analysis of *-Segment -Tone* vs. *+Segment +Tone* in Experiment 5. ERPs time-locked to the onset of target words averaged across 3 electrodes (FC1, FC2, FCz), all shaded in the schematic head) on the scalp with topography of ERP effect (*-Segment -Tone* minus *+Segment +Tone*)



2010; Zhang et al., 2009), whereas a reduced P200 was found in Experiment 5. Although the P200 component has been associated with lexical form overlap (Liu et al., 2003; Misra & Holcomb, 2003), the exact mechanism of decrease and increase in the P200 is still unclear (for a discussion, see Guo et al., 2012). A possible explanation is that the stimuli in Experiment 5 involved two-character Chinese words in contrast to single Chinese characters used in previous studies (Kong et al., 2010; Zhang et al., 2009). As discussed in Section 1.8, it is difficult to tell whether or not all single characters can be considered as words, and there are much fewer one-character than two-character Chinese words (e.g., 5,321 vs. 45,871 in SUBTLEX-CH). In fact, the N170 component, which has been considered as an index of visual familiarity (Qin, Maurits, & Maassen, 2016) and is usually more pronounced in parietal electrodes (e.g., P7, P8, PO7, PO8), was actually so strong in Experiment 5 that it impacted the central electrodes. This pattern was not observed in previous studies using single characters

(Kong et al., 2010; Zhang et al., 2009). In addition, it is obvious that the phonological repetition just occurred in one of the characters in two-character Chinese words, whereas the phonological repetition in single Chinese characters is a complete overlap. Therefore, it is likely that the inherent differences in the presented Chinese stimuli impact the P200 effect. Taken together, the contrast of phonological repetition effects between two-character Chinese words and single Chinese characters in terms of P200 effect should be further investigated in future studies.

Another puzzling result is that Experiment 5 did not reveal a N400 effect in contrast to the N400 effect observed in previous studies using two-character Chinese words (Wu & Thierry, 2010, 2012a). A potential issue is that the present study and previous studies (Wu & Thierry, 2010, 2012a) used Chinese translations of English words as stimuli, so it is impossible to control for the potential impact of the phonological neighbour density (i.e., the number of words containing a character with the same pronunciation). This potential drawback needs to be addressed with a new set of well-controlled Chinese stimuli in the future.

To summarize, segmental repetition was observed in the ERP data (i.e., a larger LPC) during a Chinese semantic relatedness judgement task, indicating that word pairs with segmental repetition might have caused interference for the participants. Experiment 5 contributes to the literature on the involvement of phonology in reading Chinese words by illustrating the predominant role of segmental activation. In the next section, a time-frequency analysis was conducted with the data of Experiments 4 and 5, in order to further explore Chinese phonological activation during English and Chinese word reading. As previously discussed, time-frequency analysis is a potential insightful approach that compliments the conventional analysis on ERPs because it can reveal non-phase-locked signals. More importantly, the time-frequency analysis can be used as a frequency-based segregation approach. Therefore, if there is a functional dissociation between Chinese segmental and tonal processing during English and Chinese word reading, segmental and tonal repetition may elicit different changes in the time-frequency data.

4.4 Time-frequency analysis of Experiment 4 and Experiment 5

4.4.1 Time-frequency Analysis

The preprocessed data of Experiment 4 (see Section 4.2.1) and Experiment 5 (see Section 4.3.1) were used for the time-frequency analyses. The power spectra were computed for all single trials using complex Morlet wavelets (Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1996). A wavelet of seven-cycle width was used, with frequencies increasing from 5 to 60 Hz, in 1-Hz steps. The power spectra were then averaged across trials for each condition in each participant. The power obtained were expressed as a percentage change relative to the power in a baseline interval from 300 to 100 ms before the target words.

4.4.2 Statistical Analysis

The time-frequency data were analysed using the cluster-based random permutation test implemented in FieldTrip (for details see Maris & Oostenveld, 2007). The cluster-based random permutation test effectively controls for Type I errors across multiple comparisons while maintaining enough sensitivity to detect differences between conditions. The method involves conducting dependent-sample t-test for every data point. The adjacent data points ($p < .05$) were combined to create clusters. For each cluster, the sum of the t-statistics was calculated at cluster-level statistics. A null-distribution was then created by first randomly assigning one of the two conditions within subjects for several times (here 1000 times) and then calculating the largest cluster level statistic for each randomization.[24] Finally, the observed cluster level statistics were compared against the null distribution, with clusters falling in the highest or lowest 2.5% considered significant. For Experiment 5 (Chinese experiment), the main effects of Segment and Tone and their interaction were analysed using a cluster-based random permutation test applied from 200 ms to 800 ms post-stimuli for all frequencies. Because significant changes were found in the gamma-band power (30-60 Hz) approximately 400 ms post-stimulus (see Section 4.4.3.1), the gamma-band power (30-60 Hz) was first averaged for each electrode and each condition in Experiment 4 (English experiment) to increase the sensitivity of the permutation test (Maris &

[24] This analysis was also conducted 10,000 times. The results remained the same.

Oostenveld, 2007). For the main effects of Segment and Tone and their interaction, a cluster-based random permutation test was then conducted for the averaged gamma-band power from 400 ms to 800 ms post-stimulus.

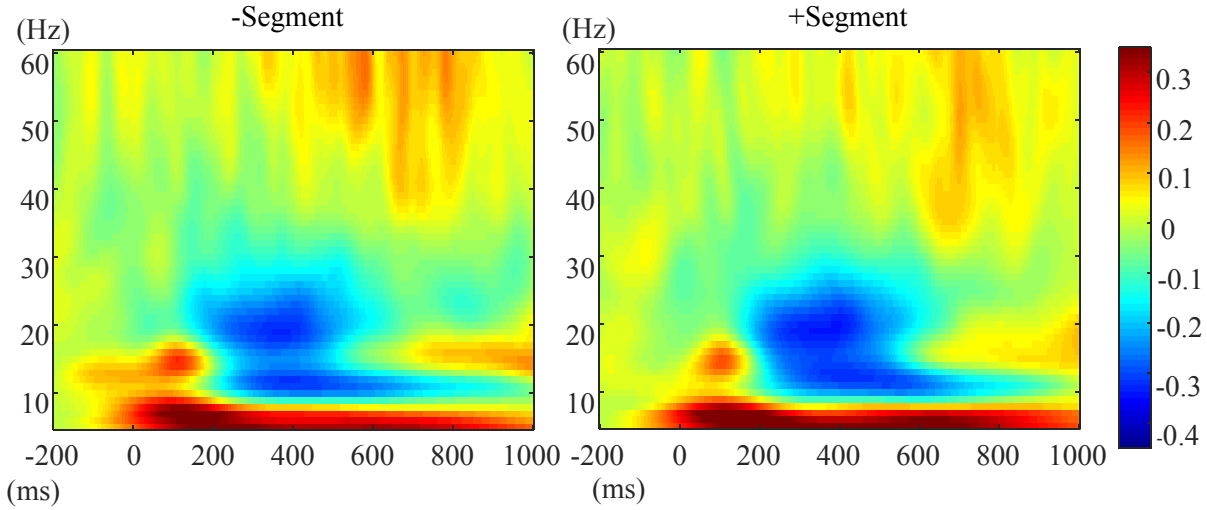
4.4.3 Results

4.4.3.1 Chinese (Experiment 5)

For the effect of Segment, a significant cluster (p = .039, 531 to 676 ms) was found in the gamma-band power (40-60 Hz), revealing that gamma-power was significantly weaker in the segmental repetition conditions (+*Segment*) than in the non-repeated segmental conditions (-*Segment*). As can be seen in Figure 31, the decreased gamma-band power in segmental repetition was widely distributed. For the effect of Tone, a significant cluster (p = .017, 465 to 800 ms) was also found in gamma band power (31-60 Hz), revealing that gamma-power was significantly stronger in the tone repetition conditions (+*Tone*) than in the non-repeated tonal conditions (-*Tone*). The increased gamma-band power in Tone repetition was widely distributed (see Figure 32). There were no significant clusters in the interaction between Segment and Tone.

Figure 31. Results of time-frequency analysis for the effect of Segment in Experiment 5. (A) Time-frequency representations at the Cz electrode for -Segment (left) and +Segment (right). (B) Time-frequency representations of the contrast -Segment minus +Segment at the Cz electrode with a mask showing the gamma frequency range (left); the corresponding topography of gamma frequency range from 531 to 676 ms with the white star marking the location of the Cz electrode (right). The colour scale indicates the relative power changes (0 = no difference).

A



B

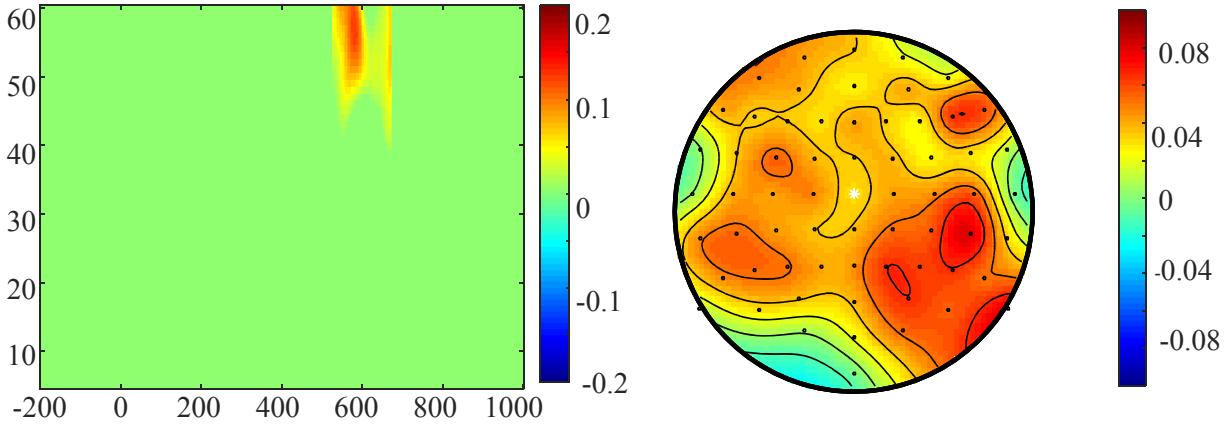
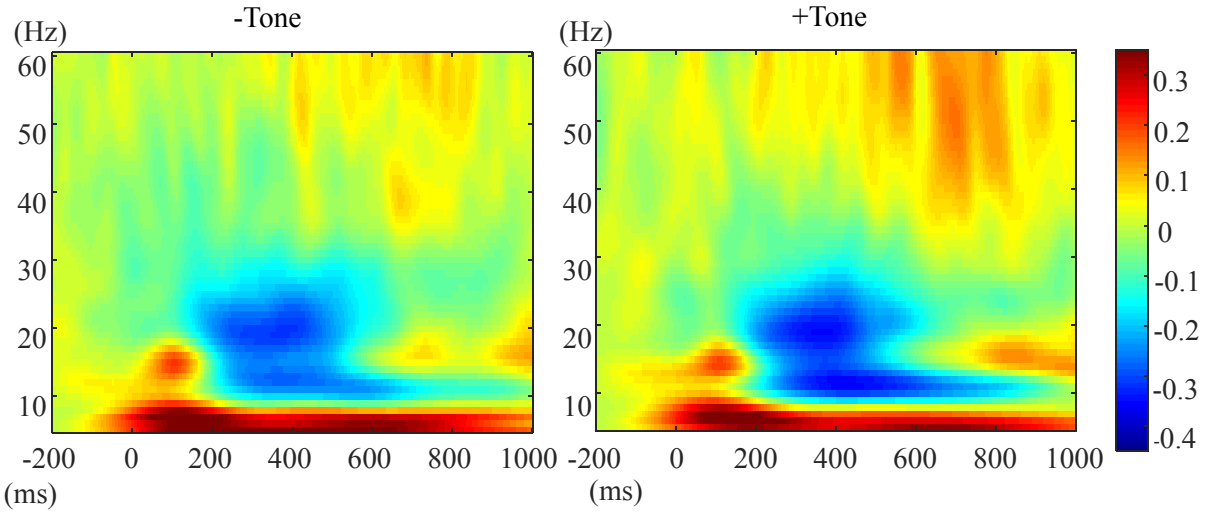
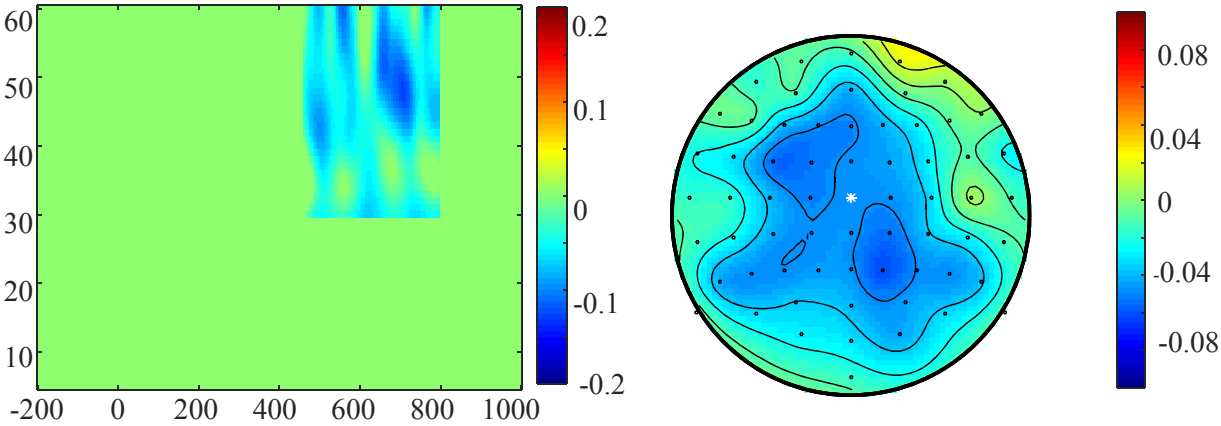


Figure 32. Results of time-frequency analysis for the effect of Tone in Experiment 5. (A) Time-frequency representations at the Cz electrode for -Tone (left) and +Tone (right) conditions. (B) Time-frequency representations of the contrast -Tone minus +Tone at the Cz electrode with a mask showing the gamma frequency range (left); the corresponding topography of gamma frequency range from 465 to 800 ms with the white star marking the location of the Cz electrode (right). The colour scale indicates the relative power changes (0 = no difference).

A



B

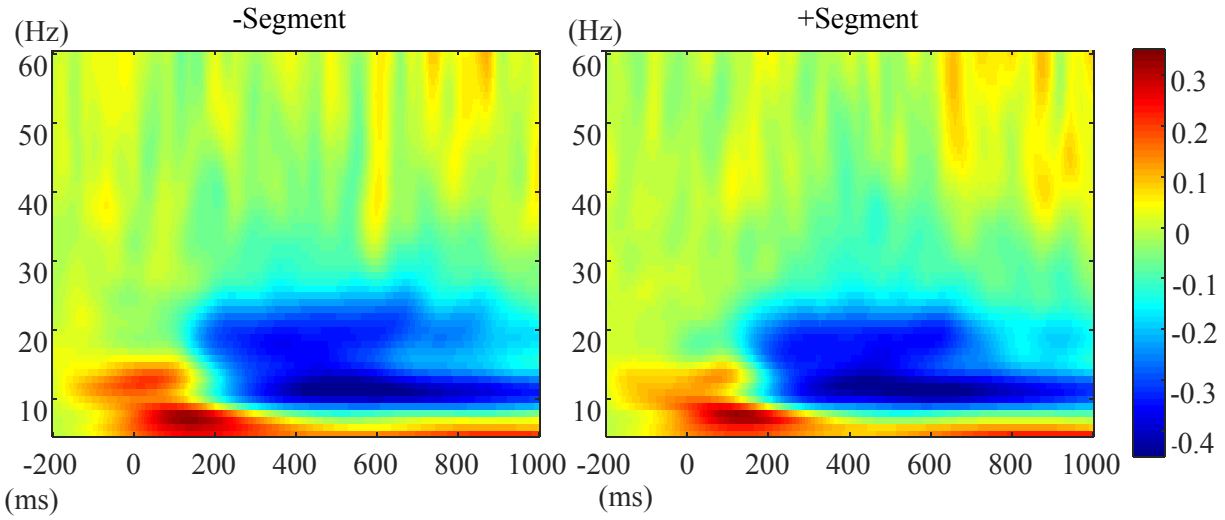


4.4.3.2 English (Experiment 4)

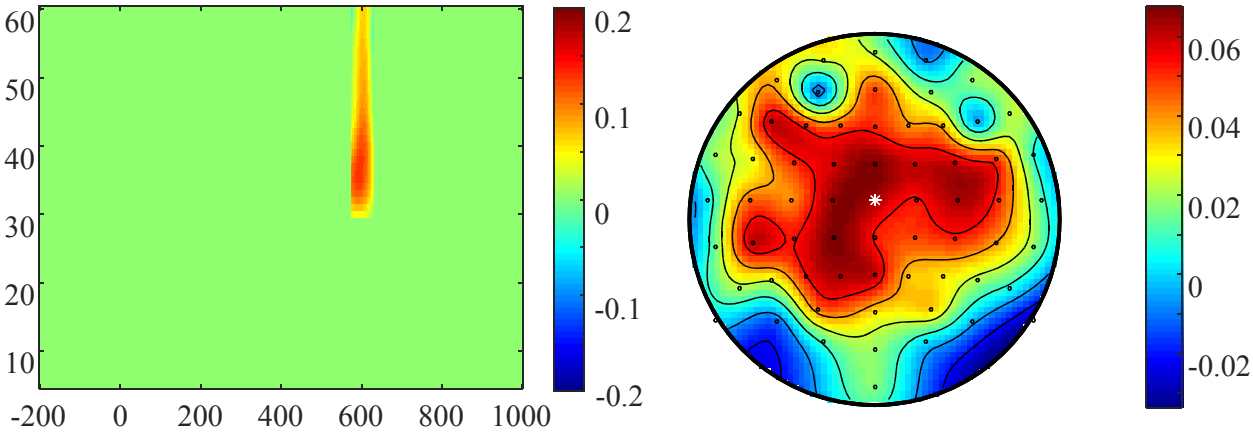
The results of time-frequency analysis are presented in Figure 33. A significant cluster (p = .0120, 578 to 632 ms) was found in gamma-band power (30-60 Hz) for the contrast between +*Segment* and –*Segment*, indicating that gamma power was significantly weaker in the segmental repetition conditions (+*Segment*) than in the non-repeated segmental conditions (-*Segment*). Within the significant time range, the decrease was most pronounced over central electrodes (see Figure 33). There were no significant clusters within the gamma-band power for the effect of Tone or the interaction between Segment and Tone.

Figure 33. Results of time-frequency analysis for the effect of Segment in Experiment 4. (A) Time-frequency representations of Cz electrode for –*Segment* (left) and +*Segment* (right) conditions. (B) Time-frequency representations of the contrast *Segment* minus +*Segment* in FCz electrode with a mask showing the gamma frequency range (30 - 60 Hz) from 578 to 632 ms (left); the corresponding topography of gamma frequency range from 578 to 632 ms with the white star marking the location of the Cz electrode (right). The colour scale indicates the relative power changes (0 = no difference).

A



B



4.4.4 Discussion

In this section, the time-frequency data of Experiment 4 and Experiment 5 were reported to further explore segmental and tone activation during English and Chinese word reading in Chinese-English bilinguals. As indicated in Section 4.2 and 4.3, both experiments involved a semantic relatedness judgment task with word pairs in which the repetition of segmental and tonal information of the hidden Chinese translations (Experiment 4) or of the Chinese words (Experiment 5) was systematically manipulated in semantically and orthographically unrelated word pairs. During English and Chinese word reading, a decrease in the gamma-band power was elicited by segmental repetition. In contrast, an increase in the gamma-band power was observed in tonal repetition when participants were reading Chinese word pairs and not when they were reading the English translations. These results indicate that there is a functional dissociation between Chinese segment and tone in the time-frequency data. Importantly, the finding that only segmental repetition in Chinese translations impacts the gamma-band power is compatible with behavioural and ERP data of Experiment 4. As reported in Section 4.2, segmental repetition in Chinese translations of unrelated English pairs resulted in a slightly quicker reaction times and elicited a reduced N400.

As far as I am aware, the current study is the first to explore cross-language activation by analysing the oscillatory EEG data. The key finding is that gamma-band power is differently affected by segmental and tonal repetition during Chinese word reading and that during English word reading the segmental repetition of the Chinese translations only affects the gamma-band power. Although it is unclear why segmental and tonal repetition lead to either an increase or decrease in the gamma-band power, similar findings have been found in a masked priming experiment with a lexical decision task (Matsumoto & Iidaka, 2008). In particular, the gamma-band power showed a decreased in repetition priming (e.g., *desk* - *desk*) and homophone priming (e.g., *pair* - *pear*) compared to control condition (e.g., *pen* - *milk*). Gamma-band power changes have also been associated with other linguistic processing, for example, semantic unification in sentence processing (e.g., Bastiaansen & Hagoort, 2015; Hagoort et al., 2004) and lexical-semantic retrieval (e.g., Mainy et al., 2008; Mellem et al., 2013), so it is difficult to conclude at the moment what its role is in Chinese phonological activation during English and Chinese word reading.

4.5 General Discussion

The focus of the current chapter was to investigate whether hidden phonological repetition priming is driven by segmental and/or tonal information of Chinese translations of English words. In the critical experiment (Experiment 4), behavioural and electrophysiological data was collected from Chinese-English bilinguals when they were conducting an English semantic relatedness judgement task. Critically, unknown to the participants, the phonology of the Chinese translations in the semantically unrelated English pairs was repeated at the segmental and/or tonal level. The behavioural results revealed that when the segment of Chinese translations was repeated, participants were significantly faster in responding. Importantly, the ERP data also revealed evidence for segmental repetition in Chinese translations of unrelated English pairs in terms of a reduced N400. It is important to emphasise that Experiment 4 not only successfully replicated the findings of Wu and Thierry (2010), it also suggested that the N400 reduction observed in previous studies (Wu & Thierry, 2010, 2012a) is driven by hidden segmental repetition.

Another unique contribution of Experiment 4 is the exploration in the time-frequency data, which revealed a decrease in the gamma-band power elicited by segmental repetition of Chinese translations in English word pairs. In contrast, tonal repetition of Chinese translations did not modulate the gamma-band power. Therefore, the time-frequency data also revealed the dominant role of Chinese segmental information activation during English word reading, which is compatible with the behavioural and ERP data. Taken together, the converging evidence of the behavioural, ERP and time-frequency data suggest the dominant role of segmental activation of Chinese translations during English word reading. Importantly, these findings of cross-language phonological activation in a purely monolingual context are consistent with the idea that bilingual visual word recognition involves non-selective lexical access in bilinguals (Dijkstra & van Heuven, 2002; van Heuven, Dijkstra, & Grainger, 1998) and they are in line with previous studies which also found that words in a given language activate their translation in another language (e.g., Guo et al., 2012; Marian & Spivey, 2003; Morford, Wilkinson, Villwock, Piñar, & Kroll, 2011; Thierry & Wu, 2007; Wu et al., 2013; Wu & Thierry, 2010, 2012a, 2012b).

In addition to the English semantic relatedness experiment, a Chinese semantic relatedness judgement experiment (Experiment 5) was also conducted. In this experiment, the Chinese translations of the English word pairs used in Experiment 4 were presented. Therefore, the phonology of the semantically unrelated Chinese pairs was repeated at the segmental and/or tonal level. Although no significant phonological repetition priming effects were found in the behavioural data, the ERP data revealed a larger LPC for segmental repetition. Because an LPC effect has been associated with conflict resolution at late decision stages (Kolk & Chwilla, 2007; Misra & Holcomb, 2003), a larger LPC in segmental repetition suggests that word pairs with segmental repetition might have caused interference for participants. More importantly, the time-frequency data provided evidence for effects of both segmental and tonal repetition. A decrease in the gamma-band power was elicited by segmental repetition, whereas an increase in the gamma-band power was elicited by tonal repetition. The opposite patterns of the gamma-band power in response to segmental and tonal repetition seem to suggest that there seems to be a functional dissociation between segmental and tonal activation during Chinese word reading. Thus, the time-frequency analysis was crucial in revealed segmental and tonal repetition effects and therefore future studies should use this approach when investigating Chinese phonological activation.

It is clear that the present study has several important implications for Chinese phonological activation in English and Chinese word reading. However, there might be some potential limitations of the experiments. There was, for example, an unbalanced ratio of unrelated trials (240 critical trials) to related trials (60 fillers) in both experiments. Because related trials were also selected from the translation database reported in Chapter 2 to ensure that Chinese translations of English words were known to participants, it is impossible to create 240 related trials. It is important to note that this set of 60 related pairs was carefully validated by an English rating task and a Chinese rating task collected from another group of 20 Chinese-bilinguals. In fact, a classic N400 effect observed in both experiments further confirmed the validity of the related trials and suggested that participants did processed the words up to the semantic level. However, the reaction times of related trials relative to unrelated controls (-*Segment* -*Tone*) showed a 79-ms facilitation in the English experiment (t (18) = 2.596, p = .018 in Experiment 4) in contrast to a 48-ms inhibition in the Chinese experiment (t (22) = 3.909, p = .001 in Experiment 5). This puzzling contrast of the behavioural

data between two experiments could be caused by the unequal number of unrelated trials to related trials. Another potential issue of the present study is the variable interstimulus interval (i.e., 500, 600 or 700 ms) between the prime and the target. Although this variable interstimulus interval is also used in previous studies (Thierry & Wu, 2007; Wu & Thierry, 2010), it is unclear how this procedure would impact the time-frequency data which have a temporal leakage issue.

To summarize, the current chapter revealed converging behavioural and electrophysiological evidence of the role of segment repetition in Chinese translations during English word reading. This provides evidence for Chinese phonological activation during English word reading, which supports the idea of non-selective lexical access (Dijkstra & van Heuven, 2002; van Heuven et al., 1998). As mentioned in Chapter 3, the next chapter will further investigate the source of Chinese activation (phonology or/and orthography) during English word reading in a series of behavioural experiments using the masked priming paradigm. These behavioural investigations are motivated by the contradictory results of Experiment 3 and Zhang et al. (2011). As discussed in Chapter 3, Zhang's et al. (2011) lexical decision study reported behavioural evidence of hidden repetition priming (e.g., east [东, /Dong1/] - thing [东西, /Dong1Xi1/]), which is a combination of orthographic and phonological repetition in Chinese translations. However, Experiment 3 found no behavioural evidence of the hidden phonological repetition. The contrast between the results of Experiment 3 and Zhang et al. (2011) seems to suggest that hidden repetition priming in the behavioural data of Zhang et al. (2011) is not driven by Chinese phonology. Therefore, further behavioural investigations are needed to understand the source of Chinese translation activation in the masked priming paradigm.

Chapter 5 Orthographical and Phonological Activation of Chinese Translation during English Word Reading: Behavioural Investigations

5.1 Introduction

The previous chapter provided converging behavioural and electrophysiological evidence for the activation of the segments of Chinese translations during English word reading. This finding is consistent with several electrophysiological studies by Wu and Thierry (e.g., Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a, 2012b). It is important to point out that the electrophysiological study presented in Chapter 4 as well as other electrophysiological studies by Wu and Thierry (e.g., Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a, 2012b) always involved the processing of *visible* English words to investigate Chinese translation activation. To investigate whether Chinese translations are activated when English words are *subliminally* presented, Zhang et al. (2011) combined the masked priming technique with the hidden repetition paradigm developed by Thierry and Wu (Thierry & Wu, 2007; Wu & Thierry, 2010). In Zhang's et al (2011) behavioural masked priming study, Chinese-English bilinguals were asked to perform a lexical decision task with English target words, which were preceded by English primes presented for 59 ms. Faster responses were found to targets when the Chinese translation of the prime was the first character in the Chinese translation of the target (e.g., east [东] - thing [东西]), but not when the Chinese translation of the prime was the second character in the Chinese translation of the target (e.g., *west* [西] - *thing* [东西]). These results provided behavioural evidence for the activation of Chinese translations of the English prime and target. Although the findings of Zhang et al. (2011) support the idea of fast and automatic translation of primes and targets, it is unclear whether this behavioural masked hidden repetition priming is also driven by Chinese phonology.

Understanding the source of fast and automatic translation in masked priming (Zhang et al., 2011) is important because it contributes to the ongoing investigation into bilingual lexical representation and lexical access, which is fundamental for

psycholinguistic research of bilingual language processing. Obviously, the hidden repetition priming originates from L1 (Chinese) and can only occur when both the L2 (English) prime and L2 target are translated into L1. This automatic translation is consistent with the assumption of non-selective lexical access (e.g., Dijkstra, 2005; Dijkstra & van Heuven, 2002; van Heuven et al., 1998). However, given that the L2 prime was subliminally and briefly presented in Zhang et al., (2011), it is surprising that the L1 translation of the L2 prime is activated so fast, because masked priming studies using non-cognate translation equivalents found strong significant L1-L2 (L1 primes and L2 targets) translation priming effects, whereas L2-L1 (L2 primes and L1 targets) translation priming effects were significantly smaller (but still significant, see meta-analysis presented in the next chapter). Although this translation asymmetry has been discussed in terms of different strengths between lexical and semantic links for L1 and L2 (Kroll & Tokowicz, 2001) or slower activation of L2 representations (Allen et al., 2015), this translation asymmetry is generally attributed to a representational asymmetry between L1 and L2 words, with L1 words connected to more semantic features than L2 words (Finkbeiner et al., 2004; Schoonbaert et al., 2009). However, if L2 primes quickly activate the L1 translations, it is unlikely that translation asymmetry would occur. Furthermore, L2-L1 masked translation priming ERP studies have not reported a N250 effect, which is associated with translation activation (e.g., Hoshino et al., 2010; Midgley et al., 2009). Therefore, this would suggest that L2 primes did not activate L1 translations. In addition, it is difficult to explain why the 75 ms hidden repetition priming effect in Experiment 1 of Zhang's et al. (2011), which requires the translation of the L2 prime into L1, is comparable to the L1-L2 translation priming effect reported in other masked priming studies with Chinese-English bilinguals (i.e., a mean priming effect of 69 ms from 4 experiments: Chen et al., 2014b, Experiment 2; Jiang, 1999, Experiment 1; Xia & Andrews, 2015, Experiment 1B and 2B). Thus, it is crucial to further investigate fast and automatic translation in masked priming at the behavioural level and to reconcile these findings with other relevant empirical evidence.

The present chapter focuses on the behavioural effects of fast and automatic translation. The first aim is to determine the source of hidden repetition priming in masked priming. To investigate whether Chinese phonology drives hidden repetition priming effects, Experiment 6 compared English word pairs with identical characters in Chinese translation and English word pairs with identical pronunciation in Chinese translation.

If the hidden repetition priming is driven by Chinese phonology, a hidden repetition priming effects would be observed in English word pairs with identical pronunciation in Chinese translation. In contrast, if the hidden repetition priming is not driven by Chinese phonology, a hidden repetition priming effects would *only* be observed in English word pairs with identical Chinese characters in their Chinese translations. To test whether the results favoured the null hypothesis (no priming effect) or the alternative hypothesis (a priming effect), the Bayes factor (BF) was also calculated. A BF smaller than 0.3 provides support for the null hypothesis, a BF larger than 3 provides support for the alternative hypothesis, whereas a BF between 0.3 and 3 supports neither hypothesis, which would indicate insensitive data (Dienes, 2014).

5.2 Experiment 6: Lexical Decision Task

5.2.1 Method

5.2.1.1 Participants

33 Chinese-English bilinguals participated in Experiment 6. One participant was excluded because of high error rates in the experimental trials (> 25%). In all experiments reported here, the bilingual participants were all native Mandarin Chinese speakers who learnt Chinese from birth. Participants were undergraduate or postgraduate university students studying in Nottingham. They met the minimum English language entry requirements to study at the University of Nottingham (IELTS 6.0 for undergraduates and 6.5 for postgraduates). The details of participants' language background are summarized in Table 17.

Table 17. Summary of participants' language background data from Experiments 6-8

	Mean (SD)		
	Experiment 6	Experiment 7	Experiment 8
Age (years)	22.5 (2.7)	21.2 (1.3)	22.8 (2.2)
Age exposed to formal English education	9.3 (2.2)	8.2 (1.8)	8.4 (2.8)
Time studies English (years)	13.2 (2.4)	13.0 (2.1)	14.1 (3.1)
English immersion experience (months)	6.2 (10.9)	4.0 (3.7)	9.0 (12.3)
Self-rated English Reading ability	4.9 (1.0)	5.3 (0.8)	5.1 (0.8)
LexTALE test score	62.7 (8.0)	62.3 (9.4)	63.5 (8.3)

Note. LexTALE (Lemhöfer & Broersma, 2012) is a quick and valid English vocabulary test; subjective reading ability were rated on a 7-point scale (1 = very poor, 7 = native-like). Results of Mann-Whitney Test showed that there was no significant difference in the self-ratings between participants from across 3 experiments ($p > 0.25$) and independent t-tests on their LexTALE scores also revealed no significant difference between the participant groups ($p > 0.25$).

5.2.1.2 Materials and Design

Stimuli were selected using the English-Chinese translation norming study reported in Chapter 2. The stimulus material consisted of two sets of 40 English target words. The first set of 40 words were preceded by either an unrelated English word whose two-character Chinese translation shared the first character with the Chinese translation of the target word (hidden phonological and orthographical repetition prime), or by an English control word matched in length and word frequency. The second set of 40 words were preceded by either an English word in which first characters in their Chinese translations was the homophone of the first characters in Chinese translations of the target words (hidden phonological repetition prime), or by an English control word matched in length and word frequency. Examples of the stimuli are presented Table 18 (see Appendix 6 for the list of all stimuli). Importantly, the English prime and

target words were all orthographically, phonologically, and semantically unrelated in English. For the English lexical decision task, eighty English pseudowords were selected from the British Lexicon Project (Keuleers et al., 2012). These pseudowords were matched with the target words in length and the first letter. Pseudowords were preceded by 80 unrelated English words. Two counterbalanced lists were created and participants were randomly assigned to one of the lists.

Table 18. Experimental design and stimulus examples of Experiment 6

	Phonological & Orthographical Repetition			***Phonological Repetition***		
	PO control	*PO*	*Target*	*P*	*P prime*	*Target*
English	cancer	career	fact	wealth	winter	thing
Chinese	癌症	事业	事实	财富	冬天	东西
Pin Yin	Ai2Zheng4	***Shi4***Ye4	***Shi4***shi2	Cai2Fu4	***Dong1***Tian1	***Dong1***Xi1
Word Length	6.18	6.23	5.78	6.10	6.13	5.35
English Frequency	4.55	4.44	4.65	4.52	4.52	4.76
Concreteness	3.51	3.99	3.79	3.98	3.48	3.97

Note. English frequencies are Zipf values from SUBTLEX-UK (van Heuven et al., 2014). Concreteness scores are based on Brysbaert et al. (2014). Pinyin is a Romanisation of the Chinese pronunciation

5.2.1.3 Procedure

Participants were tested individually in a sound-attenuated experimental room. The stimuli were presented on a CRT monitor (19" Lacie Electron Blue IV, 1024×768 pixels, 85 Hz) using DMDX Display Software (Forster & Forster, 2003). Each participant received a unique random presentation order. The procedure was identical to Zhang et al. (2011). Each trial began with a forward mask ($@#$£@£%) presented for 500 ms. Following this mask was a lowercase English word presented for 59 ms, followed by a backward mask (%$%£@£$#) presented for 24 ms. Immediately following the backward mask was the target, an uppercase letter string. Participants were not

informed of the presence of the primes and were asked to decide whether the uppercase letter string was an English word or not by pressing buttons on a Cedrus response box attached to the computer.[25] They were instructed to perform the task as quickly and as accurately as possible. The uppercase string remained on the screen up to a maximum of 3000 ms or until the participants responded. The inter-trial interval was set at 1000 ms. The experiment was approved by the Ethics Committee of the School of Psychology at the University of Nottingham. Participants all signed a consent form before the experiment started.

5.2.2 Results and discussion

The mean reaction times and error rates for Experiment 6 are presented in Table 19. Incorrect responses were discarded (7.7%). Reaction times less than 300 ms or more than 2,500 ms were considered outliers (0.3%). Because two types of prime conditions (i.e., hidden phonological and orthographical repetition prime and hidden phonological repetition prime) used different target words that were not matched on lexical properties, and each of the prime conditions was matched with different controls, priming effects were analysed separately (see Guo et al., 2012 for a similar approach). Thus, paired by-subject and by-time t-tests were conducted separately for both priming conditions using reaction times and error rates. For the hidden orthographic and phonological repetition prime, no significant differences were found in the reaction times, all ts < 1, ps > .25,

[25]After completing the English lexical decision task, all the participants also performed a Chinese lexical decision task to Chinese targets with masked Chinese primes. The Chinese words were translation equivalents of the English words. However, because the majority of participants realized that the Chinese words were translation equivalents of the English words, the Chinese data should be interpreted with caution. For the phonological repetition condition, all error rates and reaction times analyses revealed no significant differences, all ts < 1, ps > .25, all BFs < 0.3.For the orthographic and phonological repetition prime, significant differences were found in the reaction times in the by-subject analysis, t_1 (31) = 3.000, p = .005, BF > 3; t_2 (39) = 1.558, p = .127, BF = 0.517. Similarly, error rates also showed significant patterns in the by-subject analysis, t_1 (31) = 2.048, p = .049, BF = 1.184; t_2 (39) = 1.621 p = .113, BF = 0.566.

all BFs < 0.3, and error rates, t_1 (31) = 1.735, p = .093, BF = 0.71 and t_2 (39) = 1.397, p = .170, BF = 0.42. For the hidden phonological repetition prime, the analyses of error rates and reaction times revealed no significant differences, all ts < 1, ps > .25, all BFs < 0.3.

Table 19. Mean error rates and reaction times (with SD in brackets) of Experiment 6

	Phonological & Orthographical Repetition		***Phonological Repetition***	
	PO control	*PO prime*	*P control*	*P prime*
RT (ms)	963 (231)	946 (211)	914 (204)	913 (195)
Error Rates (%)	8.63 (8.54)	10.87 (9.41)	6.10 (6.31)	5.47 (5.73)

In contrast to the findings of Zhang et al. (2011), no behavioural evidence of a hidden repetition priming effect was found. It is worth to note that electrophysiological studies using the hidden repetition paradigm also reported no significant effects in the behavioural data (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a; cf. Thierry & Wu, 2004). At first glance, the clear contrast at behavioural levels between Zhang et al. (2001) using a masked priming paradigm and all those electrophysiological studies using an unmasked priming paradigm is rather puzzling, because the behavioural priming effects are generally stronger in the unmasked paradigm than in the masked paradigm (e.g., Brown & Hagoort, 1993; de Groot & Nas, 1991). However, one possible explanation for a significant behavioural effects found in Zhang et al. (2011) lies in a crucial difference of morphological complexity of the Chinese translations. Zhang et al. (2011) used the primes which can be translated into one-character Chinese words (e.g., east - 东), whereas the Chinese translations of English pairs are both two-character words in Experiment 6 and in the electrophysiological studies (e.g., train - 火车). Therefore, Experiment 7 examined the impact of morphological complexity on the Chinese translations of English primes (i.e., either the Chinese translation is a two-character or one-character word) on masked hidden repetition effects. If morphological

complexity of Chinese translation does impact the hidden repetition priming effect, hidden repetition priming effects would only be observed when English primes could be translated into one-character Chinese words. Furthermore, it is important to examine whether hidden repetition priming would get stronger with a longer presentation time because priming effects are often more salient when primes are presented longer (Chen et al., 2014b; Grainger & Frenck-Mestre, 1998). Therefore, the prime duration was also manipulated in the next experiment (59 ms vs. 118 ms). Following the common practice of creating the long prime duration by doubling the short prime duration (e.g., Ferrand & Grainger, 1992, 1993, 1994), the 118 ms prime duration was chosen. Presenting the primes for 118 ms followed by the 24 ms backward mask, the SOA was still shorter than 167 ms, which is the threshold required for measuring automatic priming in the absence of strategic processing (Neely & Kahan, 2001).

5.3 Experiment 7: Lexical Decision Task

5.3.1 Method

5.3.1.1 Participants

34 Chinese-English bilinguals participated in Experiment 7. Two participants were excluded, one because of high error rates in the experimental trials (>25%) and the other because of a failure to follow the task instructions. The details of participants' language background data are summarized in Table 17.

5.3.1.2 Materials and Design

The stimulus material consisted of two sets of forty English words used as targets. The first set of 40 words were taken from Experiment 1 of Zhang et al. (2011). These words were preceded by either English words whose one-character Chinese translations were the first characters in Chinese translations of the target words, or by English control words matched in length and word frequency. Using the data from an English-Chinese translation norming study reported in Chapter 2, a second set of 40 English target words was selected. These words were preceded by either an unrelated English word whose two-character Chinese translations shared the first characters with Chinese translations

of the target words, or by an unrelated English control word matched in length and word frequency. Examples of stimuli are presented in Table 20 (see Appendix 7 for the list of all stimuli). The duration of the primes (59 ms vs. 118 ms) was further manipulated in two counterbalanced blocks. Eighty English pseudowords were selected from the British Lexicon Project (Keuleers et al., 2012). These pseudowords were matched with the target words in length and the first letter. Similar to Experiment 6, pseudowords were preceded by 80 English words. Four counterbalanced lists were created for both sets of target words to ensure that in each list each target was viewed once and primed with all 4 experimental conditions (2 prime types × 2 prime durations) across lists. Participants were randomly assigned to one of the lists.

Table 20. Experimental design and stimulus examples of Experiment 7

	Hidden One-character Prime			***Hidden Two-character Prime***		
	Control	*Repetition*	*Target*	*Control*	*Repetition*	*Target*
Word	night	water	level	cancer	career	fact
Chinese Translation	夜	**水**	**水**平	癌症	**事**业	**事**实
Word Length	5	5	6.3	6.18	6.23	5.78
English Frequency	5.12	5.16	4.72	4.55	4.44	4.65
Concreteness	3.71	3.65	3.28	4.06	3.60	3.85

Note. English frequencies are Zipf values from SUBTLEX-UK (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Concreteness scores are based on Brysbaert, Warriner, and Kuperman (2014).

5.3.1.3 Procedure

The procedure was similar to Experiment 6 except for two aspects. First, the prime duration was either 59 ms or 118 ms, this was manipulated between blocks and the order of the block was counterbalanced across participants. Second, the uppercase string remained on the screen up to a maximum of 4000 ms instead of 3000 ms.

5.3.2 Results and discussion

The mean error rates and reaction times for Experiment 7 are presented in Table 21. Incorrect responses were discarded (9.5%). Reaction times less than 300 ms or more

than 3,000 ms were discarded as outliers (0.7%).[26] Because the two types of prime conditions (i.e., Hidden One-character Prime and Hidden Two-character Prime) had different target words that were not matched on lexical properties, and each of the hidden repetition prime was matched with different controls, priming effects were as in Experiment 7 analysed separately. For both Hidden One-character Prime and Hidden Two-character Prime condition, a 2 (control vs. repetition) × 2 (59 ms vs. 118 ms) repeated measures ANOVA was separately run using reaction times and error rates across subjects (F_1) and across items (F_2). For Hidden One-character Prime, reaction times revealed no significant main effects of prime duration and condition, , and no significant interaction, all Fs < 1, ps > .25, BFs < 0.3 expect for main effect of prime duration in by-subject analysis, BF = 0.31. The error analysis revealed also no significant main effects of prime duration, F_1 (1, 31) = 1.4, p = .246, BF = 0.47, F_2 (1, 39) = 1.414, p = .242, BF = 0.4, and priming, and no significant interaction, all Fs < 1, ps > .25, BFs < 0.3. For the Hidden Two-character conditions, reaction times also revealed no significant main effects of prime duration, all Fs < 1, ps > .25, BFs < 0.3, and priming, F_1 (1, 31) < 1, p > .25, BF <0.3, F_2 (1, 39) = 1.05, p >.25, BF < 0.3, and no interaction, all Fs < 1, ps > .25, BFs < 0.3. Similarly, the error analysis revealed no

[26] To keep more data, 3000 ms instead of 2500 was used ms as the cut-off threshold. The pattern remained the same if 2500 ms was used as the threshold.

significant main effects of prime duration and priming, Fs < 1, ps > .25, BFs < 0.3, and no significant interaction, F_1 (1, 31) < 1, p > .25, BF < 0.3, F_2 (1, 39) = 1.372, p =.249, BF < 0.3.

Again, no evidence was found for a hidden Chinese character repetition effect, not even for the Hidden One-character Prime conditions, which involved the same critical stimuli and 59 ms prime duration as Zhang et al. (2011). In other words, we failed to replicate Zhang et al. (2011). To understand the unexpected findings, it is crucial to test whether the masked hidden repetition priming findings are replicable at all. Therefore, an exact replication of Zhang et al. (2011) is needed because the design of the current experiment was not the same as the original study. To ensure that this replication has enough statistical power to detect an effect, the number of participants needed for this replication experiment was determined by a power analysis using the effect size of Zhang et al. (2011) (d = 0.349, power = 0.8, one-tailed paired t-test). The results of this power analysis indicated that 52 participants were needed.

Table 21. Mean error rates and reaction times (with SD in brackets) in Experiment 7

	Hidden One-character Prime			
Prime Duration	*59 ms*		*118 ms*	
Priming Types	*Control*	*Repetition*	*Control*	*Repetition*
RT (ms)	1068 (281)	1093 (311)	1055 (228)	1047 (270)
Error Rate (%)	6.6 (7.5)	6.0 (6.2)	8.4 (11.1)	8.2 (6.9)
	Hidden Two-character Prime			
Prime Duration	*59 ms*		*118 ms*	
Priming Types	*Control*	*Repetition*	*Control*	*Repetition*
RT (ms)	1096 (251)	1099 (249)	1091 (275)	1084 (285)
Error Rate (%)	12.7 (10.9)	12.0 (11.5)	10.0 (8.4)	12.3 (11.6)

To investigate whether a possible null effect is not due to technical or procedure issues and to confirm that the same participants are processing the primes, two additional and critical tasks were included. Thus, the experiment consisted of three tasks, which were conducted in the same fixed order so that the first task was an exact replication of Zhang et al. (2011). The second task involved masked cross-language priming with Chinese primes and English targets using the same prime duration as in the first task (59 ms). The goal of this second task was twofold. First, it served as a control task by using Chinese words as primes and their English translation equivalents as targets. Given that robust L1-L2 translation priming effects have been reported in numerous studies (e.g., Duñabeitia et al., 2010a; Gollan et al., 1997; Hoshino et al., 2010; Wang & Forster, 2010), observing a masked translation priming effect would suggest that the participants processed the primes up to the semantic level, ruling out the possibility that the absence of hidden repetition priming effect is due to participants' failure to process the primes or any experimental/technical issues. The second aim of this experiment is to investigate whether the Chinese translations of English target words are activated, which is critical to explain the absence of hidden repetition priming effect. This aim is achieved by using Chinese prime words unrelated in meaning with the English target words but sharing the first character of the Chinese translation for the English target words (e.g., prime 事业, meaning "career", and target FACT, which Chinese translation is 事实, repeated character 事). Only if the Chinese translations of English target words are activated, the character repetition primes would impact participants' response times. It is important to note that a hidden repetition priming effect in task 1 (English primes and targets) can occur only if English primes and targets are both translated into Chinese, whereas a character repetition priming effect in task 2 (Chinese primes) only requires that English target words are translated into Chinese. Therefore, if a character repetition priming effect is found in task 2 but no hidden repetition priming effect found in task 1, it would suggest that automatic Chinese translation is limited to English targets. In addition, the character repetition priming effect, if found, should be smaller than the L1-L2 translation priming effect. This is because the character repetition priming effect originates from a partial form overlap, whereas the L1-L2 translation priming effect is driven by the semantic overlap as well as a potential complete form overlap via semantic feedback (Hoshino et al., 2010; Midgley et al., 2009; Schoonbaert et al., 2011).

It is also possible that we do not find any character repetition priming or hidden repetition priming effects. To investigate whether an absence of any priming effects from partial form overlap in task 1 and 2 is due to the unconscious processing of the masked prime, a translation recognition task was added as the final task of this experiment. A translation recognition task requires explicit processing of the prime (Chinese words) and target (English words) in order to decide whether the prime and target are translation equivalents. Therefore, the prime is presented for 500 ms in order for participants to consciously translate the prime. The design of this experiment and the stimuli included were the same as for the second task of this experiment. Previous studies using translation recognition tasks have reported significant through-translation interference effects when primes and targets were not translation equivalents but contained form similarities (e.g., Ferré et al., 2006; Guasch et al., 2008; Guo et al., 2012; Moldovan et al., 2016; Moldovan et al., 2012), which indicates that translation equivalents are activated. For example, in a recent study by Guo et al. (2012), Chinese-English bilinguals were asked to perform an English-Chinese (L2-L1) translation recognition task. Participants were less accurate and slower in responding when the Chinese targets (e.g., 塘) had similar pronunciation and forms with Chinese translations of the English primes (e.g., *sugar* whose Chinese transition is 糖), suggesting that L1 translations of L2 words were activated. Following this logic, we expect that character repetition results in an interference effect in task 3.

5.4 Experiment 8: Exact Replication of Zhang et al. (2011)

5.4.1 Method

5.4.1.1 Participants

57 Chinese-English bilinguals participated in Experiment 8. Five participant were excluded from the analysis due to high error rates in the experimental trials (> 25%). Details of the participants' language background data in this experiment are summarized in Table 17.

Table 22. Experimental design and stimulus examples of Experiment 8 task 1

	First Character			***Second Character***		
	Control	*Repetition*	*Target*	*Control*	*Repetition*	*Target*
Word	wall	east	thing	town	west	thing
Chinese Translation	墙	**东**	**东**西	镇	**西**	东**西**
Word Length	5.0	5.0	6.3	4.9	4.9	6.3
English Frequency	5.12	5.16	4.72	5.10	5.13	4.72
Concreteness	3.71	3.65	3.28	3.7	3.7	3.28

Note. English word frequencies are Zipf values taken from SUBTLEX-UK (van Heuven et al., 2014). Concreteness scores are based on Brysbaert et al. (2014).

5.4.1.2 Materials and Design

The experiment consisted of three tasks that were presented in the same fixed order so that the critical replication of the first experiment of Zhang et al. (2011) was conducted first. The first two tasks were masked priming lexical decision tasks, whereas the final task was a translation recognition task. The stimuli of the first task were identical to Experiment 1 of Zhang et al. (2011). Thus, 40 English target words were preceded by four types of primes: 1) an English word whose Chinese translation is the first character in the Chinese translation of the English target word (hidden first character repetition), 2) an English control word matched in length and English frequency with the hidden first character , 3) an English word whose Chinese translation is the second character of the Chinese translation of the English target word, 4) an English control word matched in length and English frequency with the hidden second character prime. Examples of stimuli are presented in Table 22 (see Appendix 8 for the list of all stimuli). Four counterbalanced lists were created and participants were randomly assigned to one of the lists.

A new set of stimuli for task 2 and 3 were selected from the English-Chinese translation norming database reported in Chapter 2. In total, 36 English target words were selected, and these were preceded by three types of two-character Chinese primes: 1) the Chinese

translation equivalent, 2) a Chinese word whose first character is the same with the

Table 23. Experimental design and stimulus examples of Experiment 8 task 2 & 3

	Prime			*Target*
	Translation Equivalent	***Character Repetition***	***Control***	
Word	事实	事业	蜡烛	fact
English Translation	fact	career	candle	
Chinese Frequency	4.58	4.49	4.50	
Strokes	15.39	15.31	15.14	

Note. Chinese frequencies are Zipf values calculated using the word frequencies from SUBTLEX-CH(Cai & Brysbaert, 2010).

Chinese translation of the English word (character repetition), 3) a Chinese control word matched in number of strokes and word frequency with the other primes. Examples of the stimuli are presented in Table 23 (see Appendix 9 for the list of all stimuli). For the lexical decision task (task 2) a set of 36 English pseudowords were selected from the British Lexicon Project (Keuleers et al., 2012). These pseudowords matched the target words in length and the first letter. For the final task, 12 additional translation equivalent pairs were included as fillers in order to have an equal number of Yes and No responses. Importantly, the critical stimuli presented in Task 2 and 3 were exactly the same. Three counterbalanced lists were created and participants were randomly assigned to one of the lists in task 2 and presented with the same list in task 3.[27] We did not counterbalance the order of tasks 2 and 3 because we did not want participants to conduct the explicit translation task first, which would potentially impact their performance in the masked priming lexical decision.

[27] For Task 2 and 3, list 1 had 18 participants, whereas lists 2 and 3 had 17 participants. The results did not change if one participant was excluded from list 1.

5.4.1.3 Procedure

The procedure of the first task was identical to Experiment 1 of Zhang et al. (2011). Furthermore, the exact same DMDX files from Zhang et al. (2011) were used, expect that the timings (in terms of number of frames) had to be modified because the monitor used was different from the original monitor used by Zhang et al. (2011). For task 2 a similar procedure as in Experiment 6 was used, except that primes were two-character Chinese words. For the translation recognition task (task 3), each trial began with a fixation (+) presented for 500 ms, followed by a blank screen for 200 ms. A two-character Chinese word was presented for 500 ms followed by a blank screen for 200 ms (Guo et al., 2012). Immediately following the blank screen the English target word was presented in uppercase. Participants were asked to decide whether the English word was the correct translation of the Chinese word by pressing buttons on a response box (Cedrus) attached to the computer. Furthermore, participants were instructed to perform the task as quickly and as accurately as possible. The uppercase string remained on the screen up to a maximum of 3000 ms or until the participants responded. The inter-trial interval was set to 1000 ms.

5.4.2 Results

Table 24 and 25 present the mean reaction times and error rates for each task in Experiment 8. Incorrect responses were discarded (task 1: 6.0%; task 2: 3.2%; task 3: 4.1%). Reaction times less than 300 ms or more than 2,500 ms were discarded (task 1: 0.51%; task 2: 0.54%; task 3: 0%). Data of task 1 was analysed using a 2 (control vs. repetition) × 2 (first character vs. second character) repeated measure ANOVA. For task 2, a one-way 3 (translation equivalent vs. character repetition vs. control) repeated measure ANOVA was conducted, and planned comparisons were conducted when there was a significant main effect of priming condition. The results of task 3 were analysed with paired t-tests that compared translation equivalent word pairs with translation non-equivalent word pairs and critically comparing the two translation non-equivalent conditions (character repetition word pair vs. control word pairs).

Table 25. Mean error rates and reaction times (with SD in brackets) in Experiment 8 task 1

	First Character		***Second Character***	
	Control	*Repetition*	*Control*	*Repetition*
RT (ms)	941 (184)	944 (197)	926 (180)	949 (202)
Error Rate (%)	5.19 (7.3)	6.73 (8.6)	7.00 (9.2)	5.24 (7.5)

Table 24. Mean error rates and reaction times (with SD in brackets) for tasks 2 and 3 of Experiment 8

		Translation Equivalent	***Character Repetition***	***Control***
Task 2	RT (ms)	828	1016	1060
(Masked priming)		(174)	(182)	(196)
	Error Rate (%)	2.7	12.5	17.9
		(4.9)	(9.7)	(14.7)
Task 3	RT (ms)	541	815	730
(Translation recognition)		(106)	(179)	(159)
	Error Rate (%)	1.9	8.0	2.4
		(3.5)	(9.6)	(5.0)

5.4.2.1 Task 1: Masked Hidden Repetition Piming with English Primes and Targets

The analysis of the reaction times revealed no significant effect of priming, F_1 (1, 51) = 1.365, p = .248, F_2 (1, 39) < 1, p >.25, or morpheme position, Fs < 1, ps > .25, and no significant interaction, Fs < 1, ps >.25. Also, the error analyses revealed no main effects, Fs < 1, ps > .25, and no significant interaction, F_1 (1, 51) = 2.72, p = .105, F_2 (1, 39) = 3.03, p = .090. For all analyses, BFs < 0.3.

5.4.2.2 Task 2: Masked Cross-language Priming with Chinese Primes and English Targets

The RT analysis revealed a significant effect of priming, F_1 (2, 102) = 76.23, p < .001, η_p^2 = .599, BF > 100, F_2 (2, 70) = 83.85, p < .001, η_p^2 = .706, BF > 100. Planned comparisons revealed a significant 232 ms translation priming effect, t_1 (51) = 9.74, p < .001, BF > 100, t_2 (35) = 13.44, p < .001, BF > 100, and a 44 ms character repetition priming effect that was significant in the by-subject analysis, t_1 (51) = 2.48, p = .017, BF = 2.40 and a marginal result in the by-item analysis, t_2 (35) = 1.68, p = .102, BF = 0.64.

A similar pattern was found in the error analysis. A reliable effect of priming was found, F_1 (2,102) = 35.63, p < .001, η_p^2 =.411, BF > 100, F_2 (2, 70) = 25.56, p < .001, η_p^2 = .422, BF > 100. Planned comparisons revealed that participants made less errors when first presented with translation equivalent primes, t_1 (51) = 7.20, p < .001, BF > 100, t_2 (35) = 6.13, p < .001, BF > 100. Participants made also less errors when first presented with character repetition primes compared to control primes, t_1 (51) = 2.78, p = .008, BF = 4.74, t_2 (35) = 2.86, p = .007, BF = 5.73.

5.4.2.3 Task 3: Chinese-English Translation Recognition

As expected, participants responded faster to translation equivalent word pairs, which required a Yes response, than to word pairs that were not translation equivalent and required a No response, t_1 (51) = 11.77, p < .001, BF > 100, t_2 (35) = 13.84, p < .001, BF > 100. Importantly, the crucial comparison between word pairs that required a No response revealed a significant interference effect for the repeated character word pairs compared to control word pairs, t_1 (51) = 6.22, p < .001, BF > 100, t_2 (35) = 3.44, p = .002, BF = 21.62 > 3.

Analyses of error rates revealed no significant differences between translation equivalent and non-translation equivalent word pairs, *p*s > .25, BFs < 0.3. In line with the response time analysis, participants made more rejection errors for word pairs with a character repetition than to control word pairs, t_1 (51) = 5.03, p < .001, BF > 100, t_2 (35) = 3.16, p = .003, BF = 11.13> 3.

5.4.2.4 Evaluate Replication Attempts

To evaluate the two replications using the same critical stimuli from Zhang et al. (2011), namely Experiment 7 (the hidden One-character Prime condition with the prime duration of 59 ms), and the first task of Experiment 8 (the First Character condition in Task 1), effect sizes (Cohen's d) and the fixed-effect meta-analysis Bayes factors were calculated. The rationale is that if the effect size of the original study is outside the 95% confidence interval of the replications, the replication studies fail to replicate the original study (Open Science Collaboration, 2015). As for the fixed-effect meta-analysis Bayes factor test, it aims to reveal whether an experimental effect is present (Meta BF > 3) or absent (Meta BF < 0.3) on the basis of the original study and the replication (Verhagen & Wagenmakers, 2014). As Table 26 shows, the effect size of Zhang et al. (2011) is clearly outside the 95% confidence intervals of two replications and Meta BFs provides evidence for an absence of the priming effect.

Table 26. Effect sizes and Meta BF of two replications against Zhang et al. (2011)

Experiment	Effect size d [95% CIs]	Meta BF
Original study Zhang et al. (2011)	0.349 [0.53 0.17]	/
Replication 1: Experiment 7	- 0.085 [- 0.29 0.12]	0.019
Replication 2: Experiment 8	- 0.012 [- 0.17 0.14]	0.015

5.5 General Discussion

The initial aim of the present chapter was to determine whether the hidden repetition priming effects reported by Zhang et al. (2011) were due to phonological repetition. Surprisingly, across three experiments there was no evidence of hidden repetition priming effects in the masked priming paradigm. Most importantly, even in the exact replication study (first task of Experiment 8) of the original study by Zhang et al., no evidence was found of a hidden repetition priming effect with a larger sample size (N = 52) and the effect was numerically in the opposite direction of the original effect. It is important to note that that the failure to replicate Zhang et al. (2011) findings was not due to the lack of statistical power, and furthermore, Bayesian analyses strongly supported the null hypothesis. The fixed-effect meta-analysis Bayes factor tests also revealed no evidence of hidden repetition effects. Overall, the experiments failed to replicate the hidden repetition priming effects with English primes and target words in the masked priming lexical decision experiment.[28]

Although it is not uncommon to fail to replicate an experimental effect (Open Science Collaboration, 2015), a critical question is why the present experiments did not observe the hidden repetition priming effects. One possible explanation is that participants in the experiments presented in this chapter did not process the masked primes up to the semantic level. This possibility is very unlikely because the results of task 2 of Experiment 8 revealed a significant masked translation priming effect, which is generally assumed to originate at the semantic level (Finkbeiner et al., 2004; Schoonbaert et al., 2009).

A second possible explanation of the failure to observe the hidden repetition effects with English primes and targets in the current study is that the English proficiency of the bilingual participants were different from that of the bilingual participants in Zhang et al. (2011). However, the average self-rated proficiency scores in Zhang et al. (2011) were not statistically different from the participants of all the three experiments in the

[28] Reaction times were also analysed using mixed-effects modelling, which revealed the same findings (see Appendix 10 for full details).

present study (all *p*s > .25) ruling out this possible explanation. It is also noteworthy that the bilinguals in Zhang et al. (2011) and the present study had very similar linguistic profiles and were recruited from the same university. However, the impact of English proficiency differences cannot be completely ruled out because self-assessments of language proficiency may be not as reliable as objective English proficiency measures (Khare et al., 2013; Lemhöfer & Broersma, 2012; Wen & van Heuven, 2016). Although the English language proficiency of the bilinguals in the present study were objectively measured using an English vocabulary test (LexTALE, Lemhöfer & Broersma, 2012), this test was not used in Zhang et al. (2011) and therefore, it is impossible to compare the objective English proficiency measures across studies. However, the participants' LexTALE scores did not correlate with the hidden repetition priming effects in the exact replication (Experiment 8) at all, r = 0.02, p =.90. Therefore, it is unlikely that the English proficiency differences can explain the failure to find hidden repetition effects.

A third possible explanation is that there were errors in the original analyses. To explore this possibility, the raw data of Zhang et al. (2011) were reanalysed. First, it is important to emphasise that no error were found in the original analyses. Additional analyses conducted using the mixed-effects modelling also confirmed the findings of the ANOVA analyses presented in Zhang et al. (2011). Crucially, the Bayesian factors also supported the presence of priming effects, BF > 3. However, the reaction times of the control condition for the hidden first character (e.g., *wall - thing*) were significantly slower than those of the control condition for the hidden second character (e.g., *town - thing*), 1,005 ms vs. 946 ms, t_1 (30) = 2.337, p = .026, t_2 (30) = 2.118, p = .041. Given that the same set of target words were used and both groups of control word were similar in word frequency and length, the differences in reaction times between the two controls might be false positives.

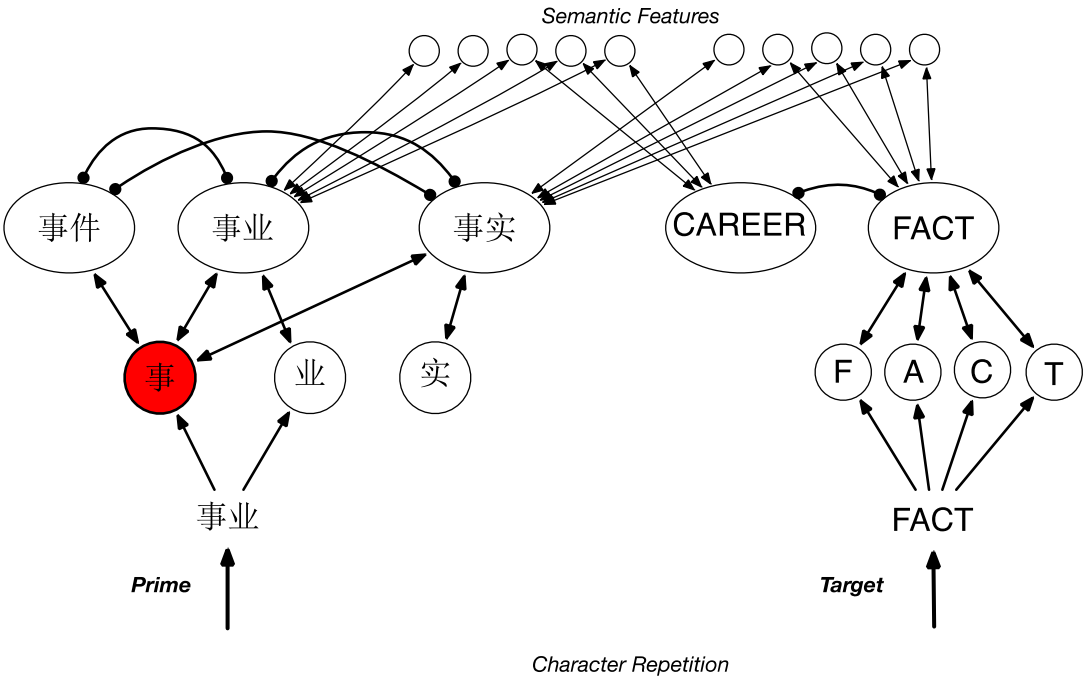
The final possible explanation is that the effect size of Zhang et al. (2011) was small and the power was not high, leading to a low chance of reproducing the effect. This seems to be the most plausible account. Despite a strong priming effect (i.e., 75 ms), the standardized effect size was small (i.e., d = 0.349). With such a small effect size, a power analysis indeed indicated that Zhang et al. (2011) was not a high-powered study because the power was only 0.61 (N = 32, d = 0.349, one-tailed paired t-test). Given that the power is the probability of successfully rejecting the null hypothesis (i.e., 1

minus type II error), the chance of finding significant results in three successive replication experiments in a row would be $0.61^3 = 0.227$ at best (Ioannidis & Trikalinos, 2007). However, it is important to note that this probability of successful replication might still be an over-estimation because the observed power is an over-estimation of the true power (Hoenig & Heisey, 2001; Vasishth & Nicenboim, 2016). This is because the true effect size of hidden repetition effects may be even smaller than what was observed in Zhang et al. (2011). One obvious support for this is that the effect size of character repetition effects observed with masked Chinese primes in Task 2 of Experiment 8 was even smaller (d = 0.23) than the effect size of hidden repetition effects in Zhang et al. (2011). It is difficult to understand why hidden repetition effects involving primes and targets translations would be larger than character repetition effects, which only require targets to be translated.

Another finding in the present study that should be carefully examined is the result of the character repetition priming in Experiment 8. This effect was significant in the by-subject analysis but marginally significant in the by-item analysis. Additional mixed-effects modelling revealed however a significant priming effect (t = 3.192, p = .002). In terms of a classical ANOVA analysis the effect is not robust in the by-item analysis. To explain this puzzle, it is necessary to consider how character repetition priming effects could occur. One possible account is that the Chinese primes activate other Chinese words sharing the same characters, e.g., Chinese prime **事**业 [career] activates Chinese word **事**实 [fact] via a shared character **事**. These pre-activated Chinese words (e.g., 事实 [fact]) could be processed up to the semantic level (and potentially form level, e.g., translation) so the recognition of English equivalents (e.g., fact) would be facilitated. This explanation seems unlikely for several reasons. First, Chinese words generally share the same characters with many words (on average 67 first-character neighbours for the primes in Experiment 8). For example, the Chinese character 事 occurs in 事情 [thing], 事件 [incident], 事故 [accident], 事迹 [deeds], 事物 [object], 事务 [routines], 故事 [story], 公事 [business], 私事 [personal issue]. Assuming competition at the lexical level, the prime (e.g., **事**业 [career]) would suppress these other words (see Figure 34). In addition, as Forster (2013) suggested, the maximal capacity for the lexical processor is two words at the same time. Because the prime is only briefly presented for 59 ms and replaced by the target after a 24-ms backward

mask, it is more likely that lexical processor is already busy with processing the prime and target so there is no capacity to process the other neighbours further. Thus, it is unlikely that all those same-character neighbours are strongly activated and that the Chinese translation of the target (e.g., 事实 [fact]) would activated its semantic features so that target word recognition is facilitated. Furthermore, the pre-activated same-character neighbours account also runs into difficulty when considering the interference effect in the translation recognition task of Experiment 8. The logic is that if the prime pre-activated same-character neighbours Chinese words (e.g., 事实 [fact]) and trigger lexical processing, a facilitation in character repetition condition should have been observed, which would be similar to translation equivalent condition.

Figure 34. The origins of the character repetition priming effect



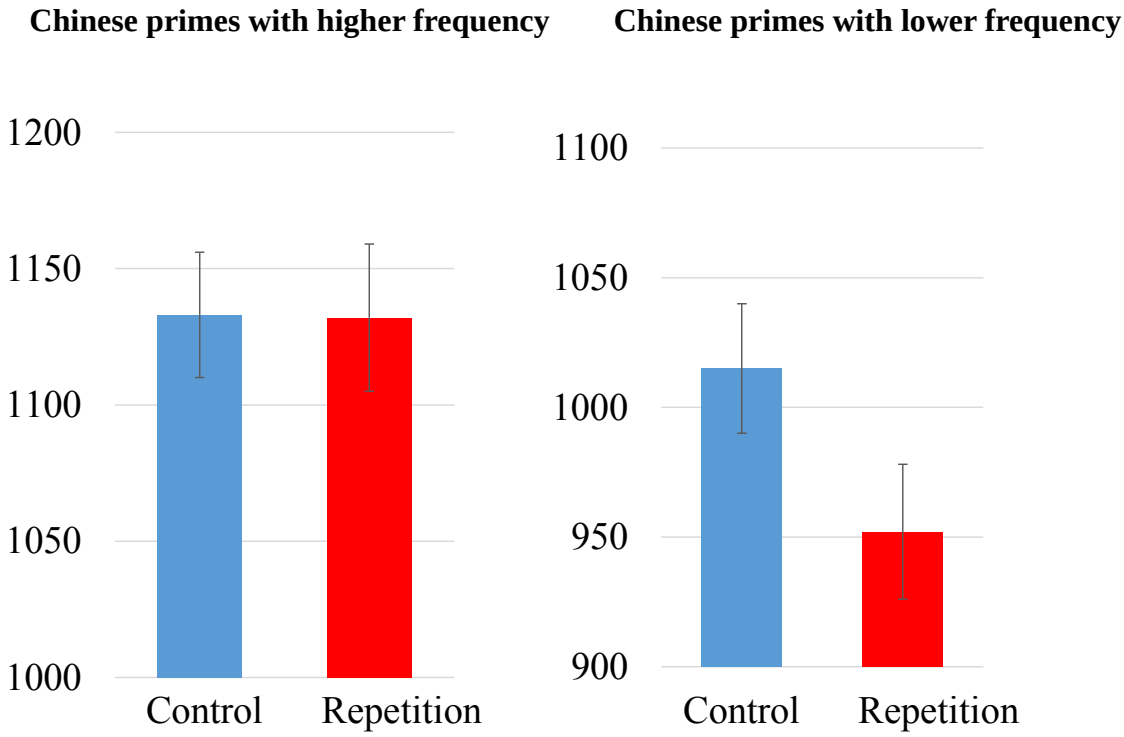
Note. 事件 [incident], 事业 [career], 事实 [fact])

Another explanation for the character repetition priming is that the English target is translated into Chinese. When the translation of the target (e.g., 事实 [fact]) is activated, its activation receives a boost because it is already activated by the prime (because it

shares the first character) and therefore it will increase the activation of the semantic features, which will in turn facilitate the target word decisions relative to the control prime (which does not activate the translation of the target). If this explanation is correct, the word frequency of the Chinese prime and Chinese translation of the target would be modulated the character repetition effect because of the possible competition between these representations at the early stages of processing. To investigate this possibility, the reaction times in the by-item analysis were reanalysed according to the relative frequency between Chinese primes and Chinese translations of the English targets. This analysis revealed that when the Chinese primes have a higher word frequency (mean = 4.78 in Zipf values) than the Chinese translations of the English targets (mean = 4.30), there is no character priming effect, t (16) = 0.025, p = .980, BF = 0.249 (priming effect: 1 ms, see Figure 35), whereas when the Chinese translations of the targets have a higher word frequency (mean = 4.82) than the Chinese primes (mean = 4.23), there is a significant priming effect, t (18) = 2.468, p = .024, BF = 2.549 (priming effect: 63 ms).[29]

[29] This impact of the relative frequencies could also influence the explanation in terms of pre-activation of the English target. However, as indicated above there are many neighbours activated by the prime, it is unlikely that all the neighbours are processed up to the semantic level.

Figure 35. Mean response times for Task 2 in Experiment 8 with error bars indicating the standard errors



Note. Left: Chinese primes have a higher word frequency than the Chinese translations of the English targets; Right: Chinese primes have a lower word frequency than the Chinese translations of the English targets

Importantly, the resonance effect between same-character neighbours through translation can also be applied to the interference effect in the translation recognition task. That is, when the translation of the target (e.g., 事实 [fact]) is activated, its activation receives a positive feedback from the prime (i.e., its same-character neighbour). This boost effect should be supressed in order to make a no response, resulting in a slower response times compared to control primes. In sum, from all the available evidence so far, it is more likely that character repetition priming originates from the Chinese translation activation of the English target words.

Taken together, although the systematic replication experiments did not find behavioural evidence for a hidden repetition effect in the masked lexical decision task, it does not mean that Chinese translation equivalents are not activated during English

word recognition. Evidence of character repetition priming effect was found in the second task of Experiment 8 with Chinese masked primes, indicating that automatic Chinese translation is limited to English target words. Obviously, the effect size of this character repetition effects was small (d = 0.23), so a critical question is whether the character repetition priming effect observed in the present study is replicable. Using a different set of items and an identical procedure, Zhang (2013) also reported a facilitation effect when masked Chinese primes shared a character with the Chinese translations of the English target words (d = 0.29). Therefore, this provides additional support that Chinese translations of English target words are activated in the masked lexical decision task. In addition, the finding that English (L2) targets are translated into Chinese was further supported in the third task of Experiment 8. In this translation recognition task with unmasked Chinese-English word pairs, the character repetition showed an interference effect on the reaction times (d = -0.49). Consequently, these findings can only be explained by assuming that the Chinese translations of English target words were activated and that the primes were processed. In contrast, as previously noted, a key aspect of the hidden repetition effect in the masked lexical decision task is that this effect can only occur when the prime and target are translated. Therefore, the failure to find a hidden repetition effect with English primes and targets in behavioural data can be only explained by the failure of the masked English (L2) prime to be translated into Chinese (L1).

The present study focused on behavioural effects of masked hidden repetition priming. The failure to find behavioural evidence of the translation of masked English primes could be a limitation of the behavioural method, which might not be sensitive enough to detect the hidden repetition effects. Could electrophysiological measures reveal the hidden repetition effects in the masked priming paradigm? As already demonstrated in seminal studies by Thierry and Wu (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a, 2012b), electrophysiological measures might be a better tool than behavioural measures because electrophysiological data can unravel the time course of bilingual language processing, which cannot be captured by behavioural data (Duñabeitia et al., 2016). However, as previously discussed, some ERP electrophysiological have suggested that masked L2 primes did not quickly activate L1 translations (e.g., Hoshino et al., 2010; Midgley et al., 2009). Therefore, it seems unlikely that electrophysiological measures would capture the hidden repetition effects in masked priming paradigm.

To summarize, three behavioural experiments in the present chapter with in total 118 Chinese-English bilinguals did not find any behavioural evidence of a hidden repetition priming effect in the masked priming paradigm. However, a classic masked translation priming was found, indicating that the Chinese primes were processing up to the semantic level. Furthermore, character repetition priming was obtained with masked and unmasked Chinese primes, indicating that automatic translation is restricted to English target words. The experiments in this chapter do, however, not provide an answer to whether or not masked English prime words are processed up to the semantic level. A possible answer might be found in the literature. However, as discussed in Chapter 1 (Section 1.3) studies using the masked translation priming paradigm have reported mixed findings for L2-L1 translation priming. To investigate whether L2-L1 translation priming in the masked translation priming paradigm reported the literature is reliable and to explore the possible factors modulating the translation priming effects, the next chapter will report a meta-analysis of existing studies using the masked translation priming paradigm.

Chapter 6: A Meta-analysis of Masked Translation Priming[30]

6.1 Introduction

As discussed in Chapter 1, although the non-cognate masked translation priming paradigm has been widely used to investigate word processing in bilinguals, studies with unbalanced bilinguals have reported mixed findings, in particular for L2-L1 translation priming. Several studies have provided narrative reviews of translation priming effects reported with unbalanced bilinguals in the literature (e.g., Altarriba & Basnight-Brown, 2007; Dimitropoulou et al., 2011b; Nakayama et al., 2016; Xia & Andrews, 2015), which discussed a range of possible factors modulating the translation priming effects. For example, factors related to statistical power (i.e., number of participants), the prime-target presentation procedure (i.e., prime duration, inter-stimulus interval, stimulus onset asymmetry), the stimuli (i.e., number of experimental items, the languages involved are the same or different scripts) and general processing speed (i.e., response speed of participants). Because existing empirical studies differ dramatically in terms of all these factors and no studies in the literature have so far considered some or all of these potential moderators systematically, the tentative conclusions of narrative reviews remain inconclusive. Importantly, these insightful reviews have so far focused mainly on the magnitude of the priming effects (in milliseconds), which are unstandardized estimates of the effect sizes and thus cannot be compared across studies. Surprisingly, there are no meta-analytic reviews conducted in the literature, so far as we are aware, that quantitatively assessed the standardized

[30] This data of this chapter has been published in *Psychonomic Bulletin & Review* as Wen, Y., and van Heuven, W. J. B. (2016). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analytic review. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-016-1151-1. This chapter is an adapted version of the published paper because slight modifications of the text have made to the introduction and the discussion in order to integrate it in the thesis.

effect sizes of L1-L2 and L2-L1 translation priming and the impact of potential experimental moderators on translation priming effects.

To fill this important gap in the literature, we present in this chapter a meta-analysis that investigated masked translation priming effects of non-cognates word pairs in lexical decision tasks. A meta-analysis uses standardized effect sizes and their variance observed in studies and tests statistically whether the overall effect size provides evidence of the experimental effect (Borenstein, Hedges, Higgins, & Rothstein, 2009). Another unique advantage of a meta-analysis is that potential moderators can be tested statistically. These moderators may explain inconsistent findings in experiments reported in the literature. Therefore, the aim of the meta-analysis in the present study was twofold. First, the primary goal was to determine the overall effect size of L1-L2 and L2-L1 translation priming and to statistically compare the effects sizes between the two translation directions. The second aim was to statistically test whether effect sizes of translation priming are influenced by moderators previously suggested in the literature. The following seven potential moderators were considered: the number of participants, the prime duration, the SOA (Stimuli Onset Asynchrony, i.e., the interval between the onset of prime and the onset of target), the ISI (Inter-Stimulus Interval, i.e., interval between the offset of prime and the onset of target), script type, number of items per cell and response speed.

6.2 Method

6.2.1 Literature Search and Study Selection

A literature search was conducted using “masked translation priming” as the search string in PsycINFO, Web of Science and PubMed (up to 31 March, 2016) to identify possible studies to be included in the meta-analysis. To find additional studies, recent studies and reviews of masked translation priming were consulted (Dimitropoulou et al., 2011b; Duñabeitia et al., 2010b; Nakayama et al., 2016; Schoonbaert et al., 2009; Xia & Andrews, 2015). The following criteria were used to select the final set of studies and experiments for the meta-analysis: (1) prime duration ≤ 100 ms, (2) primes were

masked, (3) a lexical decision task was used, (4) prime-target pairs were non-cognates, (5) L1/L2 of bilinguals were clearly specified, and (6) either the F or t value of the translation priming effect were reported. Using these selection criteria, we found 24 published articles. For the L1-L2 translation priming direction, 31 experimental observations were extracted from 20 studies. For the L2-L1 translation priming, 33 experimental observations were extracted from 18 studies. A detailed description of these studies has been provided in the Table 2 and 3 in Chapter 1.[31]

6.2.2 Meta-analysis

6.2.2.1 Effect Sizes

The effect sizes (d) were calculated using t values or F values, the number of participants (n) and the formula proposed by Rosenthal (1991): $\mathrm{d} = \frac{t}{\sqrt{n}}$ or $\mathrm{d} = \sqrt{\frac{F}{n}}$. Previous meta-analyses (e.g., Van den Bussche et al., 2009) have also used this formula to estimate effect sizes in within-subject experiments. In line with other meta-analyses for masked priming effects in monolingual studies (Lucas, 2000; Van den Bussche et al., 2009), t values or F values was taken from the subject analyses. To indicate the direction of the priming effects, the effect size was specified as positive or negative based on the means of the translation priming and unrelated (control) conditions. Thus, a positive effect size indicates a facilitatory translation priming effect. Sampling variance of the effect sizes was calculated using the formula provided by Morris and DeShon (2002):

$$Sampling\ variance = \left(\frac{1}{n}\right)\left(\frac{n-1}{n-3}\right)(1+nd^2) - \frac{d^2}{[c(n-1)]^2}$$

in which d is the effect size, n is the number of participants and $c(n-1)$ is defined as (Hedges, 1981, 1982):

$$c(n-1) = 1 - \frac{3}{4(n-1)-1}.$$

[31] A detailed description of these studies can also be downloaded from http://link.springer.com/article/10.3758%2Fs13423-016-1151-1

6.2.2.2 Moderator Coding

The seven factors mentioned in the introduction were included as moderators in the present meta-analysis. Six factors were included as continuous moderators: the number of participants, number of items per cell, the prime duration (in ms), the ISI (in ms), the SOA (in ms) and overall response speed as measured by the mean reaction time in the unrelated (control) condition (in ms). Script type was coded categorically, as either as same-script languages (e.g., Dutch and English) or different-script languages (e.g., Chinese and English).

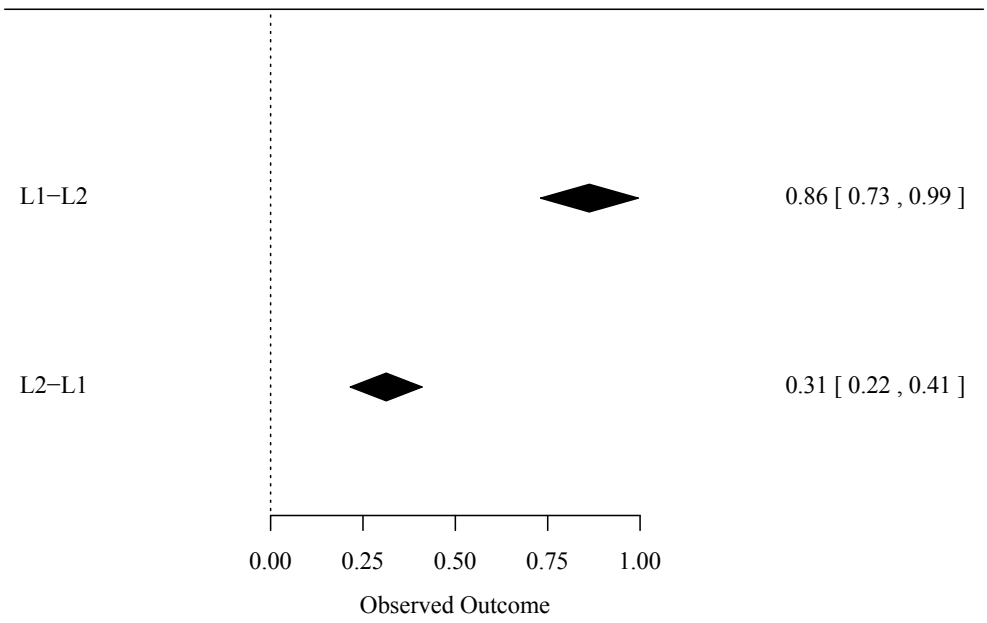
6.2.2.3 Data Analyses

The meta-analysis was conducted using the metafor package (Viechtbauer, 2010) in R v.3.2.4 (R Core Team, 2016). For both translation directions, a random-effects model without any moderators was first conducted to estimate the effect sizes of L1-L2 and L2-L1 translation priming. A Z test was conducted to compare the overall effect sizes of the two translation directions (Borenstein et al., 2009). A significant Z test would suggest that the effect sizes of the two translation directions are different and separate analyses are warranted. Next, for both translation directions Q tests of variance were conducted to investigate the heterogeneity of the observed effect sizes. A significant Q test would indicate that the observed effect sizes are heterogeneous, and that potential moderators are likely to exist. To investigate the influence of the potential moderators, we used a meta-regression approach similar to Van den Bussche et al. (2009). First, each of the seven moderators were separately entered into a random-effects model. Next, when more than one of the moderators was significant, we included these significant moderators in the initial model of the meta-regression. In order to address the issue of the collinearity between moderators, we orthogonalised moderators that significantly correlated by fitting a linear model to obtain the residuals (see, for example, Siyanova-Chanturia et al., 2011, for a similar approach). The residuals of the model were then included in the meta-regression. Finally, a backward model selection procedure was used in which non-significant moderators were step-by-step eliminated from the model.

6.3 Results

The overall effect size (d) for the L1-L2 translation priming direction was 0.86, z = 12.869, $p < 0.0001$, whereas the overall effect size for L2-L1 translation priming was 0.31, z = 6.3481, $p < 0.0001$. The difference of 0.55 between the overall effect sizes of the L1-L2 and L2-L1 translation priming directions was significant, z = 6.61, $p < 0.0001$. Figure 36 illustrates the effect sizes of translation priming for the two directions with 95% CIs. Because the effect sizes are significantly different for each translation direction the next analyses were conducted for each translation direction separately.

Figure 36. Overall effect sizes for L1-L2 and L2-L1 non-cognate masked translation priming (with their 95% confidence intervals)



6.3.1 L1-L2 Translation Priming

Figure 37 presents an overview of the observed effect sizes for L1-L2 translation priming. A Q test of variance revealed that the effect sizes across experiments were heterogeneous, Q = 82.00, df = 30, $p < 0.001$. The separate random-effects model analyses that each included a different moderator revealed that ISI and SOA were the significant moderators (Table 27). Because ISI and SOA are highly correlated, r = 0.98,

$p < 0.001$, the collinearity between ISI and SOA was reduced by using the residuals from the linear model in which ISI was predicted by SOA. When ISI and the residuals of SOA were both entered into a random-effects model, SOA was not significant anymore, $\beta = 0.0015$, SE = 0.0056, z = 0.262, $p = 0.793$, whereas ISI was still significant, $\beta = 0.0022$, SE = 0.0011, z = 2.072, $p = 0.0382$. The final model only included ISI and it explained 16.00% of the variance between studies. The AIC (Akaike's Information Criterion) of this model was 35.925, which is smaller than 38.177 for the model without any moderators, indicating that the model with ISI was a better model.

Figure 37. Observed effect sizes for L1-L2 non-cognate masked translation priming ordered by magnitude of the effect size and the overall effect sizes (with their 95% confidence intervals)

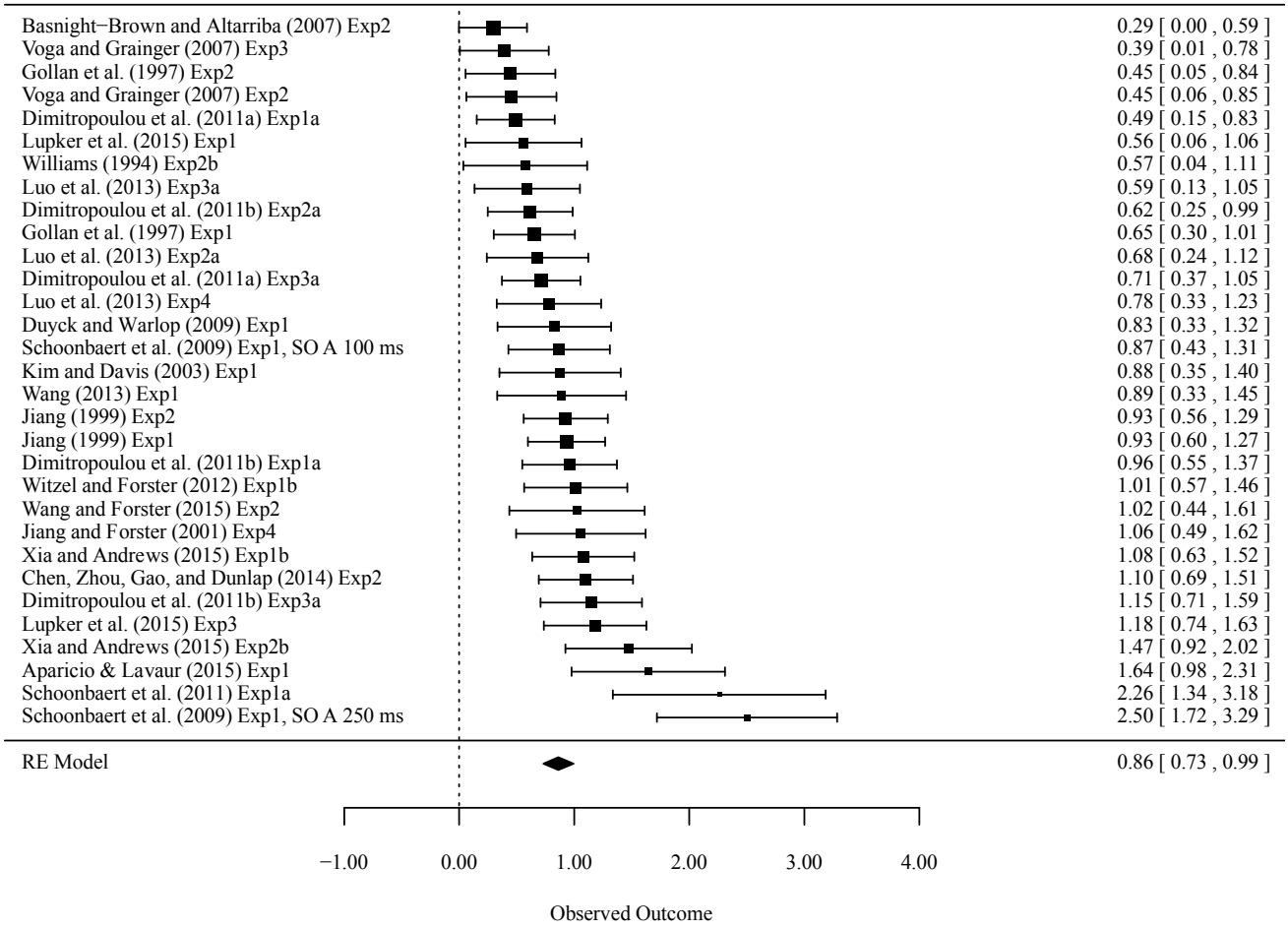


Table 27. Meta-regression analysis with one moderator for L1-L2 translation priming

	Intercept	Estimate	SE	Z value	*p*	R^2 (%)
number of participants	1.0951	-0.0071	0.0076	-0.942	0.346	0.00
prime duration	0.8467	0.0004	0.0058	0.061	0.951	0.00
ISI	0.7728	0.0022	0.0010	2.128	0.033	16.00
SOA	0.6558	0.0022	0.0011	2.104	0.035	10.25
number of items per cell	0.6905	0.0073	0.0041	1.791	0.073	4.77
script type	0.9929	-0.1624	0.1676	-0.969	0.332	0.00
response speed	1.7479	-0.0012	0.0008	-1.643	0.100	2.82

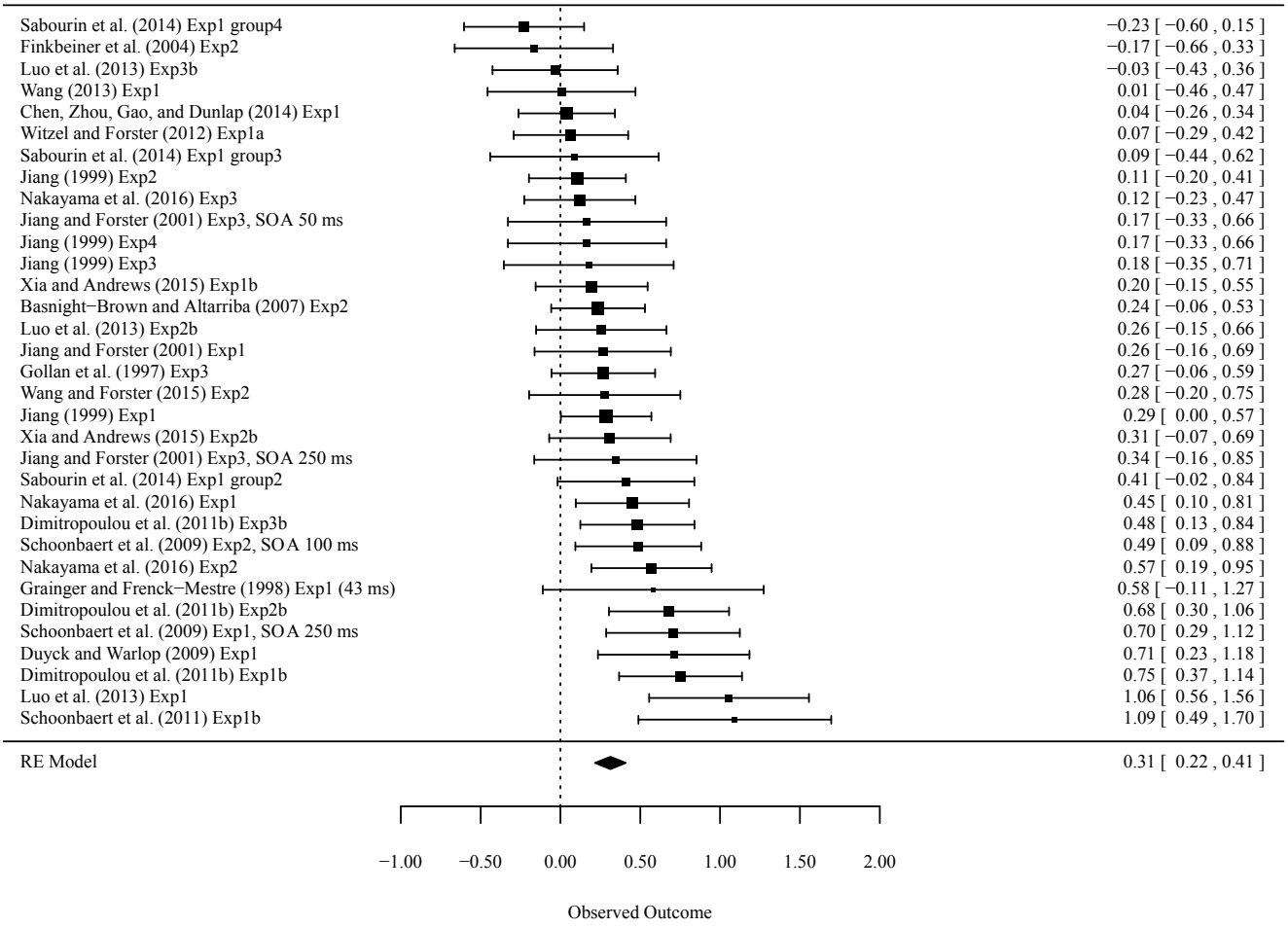
6.3.2 L2-L1 Translation Priming

Figure 38 presents an overview of the observed effect sizes for L2-L1 translation priming. A Q test of variance showed that the effect sizes across studies were again heterogeneous, Q = 60.93, df = 32, $p = 0.0015$. Separate random-effects models with each moderator revealed that number of items per cell was the only significant moderator (Table 28). The final model included number of items per cell as the moderator, which explained 74.60% of the heterogeneity between studies. The AIC of this model was 7.784, which is smaller than the 17.219 for the model without any moderators, which suggested that the model with the number of items per cell was a better model.

Table 28. Meta-regression analysis with one moderator for L2-L1 translation priming

	Intercept	Estimate	SE	Z value	*p*	R^2 (%)
number of participants	0.3297	-0.0005	0.0051	-0.104	0.917	0.00
prime duration	0.1115	0.0037	0.0040	0.935	0.350	0.00
ISI	0.3383	0.0004	0.0006	-0.619	0.536	0.00
SOA	0.3504	-0.0003	0.0006	-0.485	0.628	0.00
number of items per cell	0.0585	0.0102	0.0026	3.914	<0.001	74.60
script type	0.4046	-0.1214	0.1156	-1.050	0.293	0.00
response speed	-0.1039	0.0007	0.0007	0.961	0.336	0.00

Figure 38. Observed effect sizes for L2-L1 masked non-cognate translation priming ordered by magnitude of the effect size and the overall effect sizes (with their 95% confidence intervals)



6.4 General Discussion

A meta-analysis of 64 experimental observations across 24 studies was conducted to quantitatively assess the overall effect sizes of masked translation priming effects from L1 to L2 and vice versa. The results revealed *significant* translation priming effects for *both* directions with L1-L2 translation priming significantly larger than L2-L1 translation priming (i.e., overall effects sizes: 0.86 vs. 0.31). This finding supports the

view that the translation priming asymmetry between L1-L2 and L2-L1 is quantitative rather than qualitative (Schoonbaert et al., 2009).

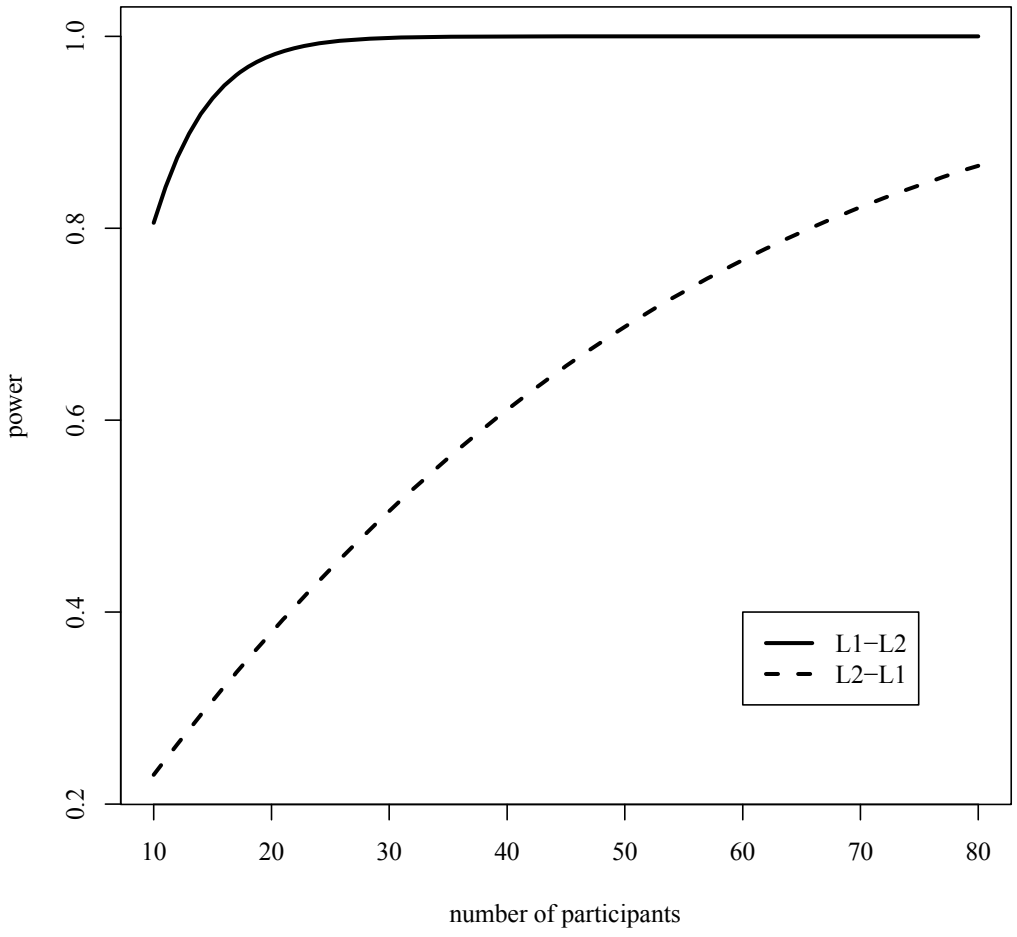
The meta-analysis further investigated the influence of seven potential moderators (number of participants, prime duration, ISI, SOA, number of items per cell, script type and general response speed). The results revealed that the effect sizes of L1-L2 translation priming were moderated by ISI, and the effect sizes of L2-L1 translation priming were moderated by the number of items per cell.[32] These findings have two important implications. For L1-L2 translation priming, it is very likely that a longer ISI increases the time to process the prime, resulting in a stronger priming effect. Therefore, it is crucial for future studies to systematically investigate how considerable variations in ISI influence L1-L2 translation priming. For L2-L1 translation priming, our results confirm earlier concerns in the literature about the large variation in the number of items per cell in masked translation priming experiments (Dimitropoulou et al., 2011b; Nakayama et al., 2016). A possible explanation for the impact of the number of items on L2-L1 priming is that using more items may increase priming effects because it reduces the noise in the data (Van den Bussche et al., 2009). Although the number of items per cell varied from 12 to 80 in the studies included here, researchers rarely provided a justification for the choice of the number of items per cell and no studies, as far as we know, have investigated this systematically. Therefore, future studies should use a sufficient number of items per cell to investigate the L2-L1 translation priming. It is beyond the scope of the meta-analysis to provide a recommendation about the number of items per cell, but it is important to note that it is possible to calculate the number of items needed a priori for a high powered experiment (see Stevens, Mandera, Keuleers, & Brysbaert, 2015). Selecting more translation pairs could easily be accomplished by using large-scale databases with translation norms (e.g., Prior et al., 2007; Tokowicz et al., 2002; Wen & van Heuven, 2016). Taken together, the impact of

[32] Following the suggestion of a reviewer of the paper, we also investigated the impact of script in masked experiments with only a prime duration of up to 60 ms and an ISI of up to 50 ms. The results again indicated that script type was not a significant moderator, *p*s > .80. See Appendix 11 or supplementary material of the published paper (http://link.springer.com/article/10.3758%2Fs13423-016-1151-1) for more details.

ISI and the number of experimental items should be considered when judging the mixed findings.

Our meta-analysis is also useful for researchers because the overall effect sizes estimated here can be used as a benchmark to calculate the number of participants needed for a study to detect an effect. For example, in a one-tailed paired t test with an effect size of 0.86, only 10 participants are required to obtain a power of 0.8 for L1-L2 translation priming studies. In contrast, for a L2-L1 translation priming effect size of 0.31, 66 participants are necessary to obtain a power of 0.8 in a one-tailed repeated t test. As can been seen in Figure 39, due to the differences between the effect sizes of

Figure 39. The relationship between power and number of participants estimated with the overall effect sizes for L1-L2 and L2-L1 masked non-cognate translation priming



L1-L2 and L2-L1 translation priming, there is a large difference in power between L1-L2 and L2-L1 translation priming when the same number of participants are tested.

The findings of the current meta-analysis are clear. However, there are some limitations. First of all, because the literature search did not include any unpublished data, it is not feasible to reliably estimate the publication bias in the field. Secondly, although studies have found that the L2 proficiency of bilinguals (Dimitropoulou et al., 2011b; Nakayama et al., 2016) and the age of L2 acquisition (Sabourin et al., 2014) modulated translation priming effects, we, unfortunately, could not consider the second language profiles of the bilingual participants (e.g., L2 proficiency, language dominance, age of L2 exposure/acquisition) in the present study because the majority of the studies included in the present meta-analysis failed to assess or provide detailed descriptions of the participants' second language profiles. Across the 24 studies, only 12 used self-assessed proficiency ratings as an estimate of the bilinguals' L2 proficiency, and only 10 studies provided information about the age of first L2 exposure/acquisition. Critically, studies have suggested that self-assessment is less reliable than objective language proficiency measures such as obtained with LexTALE (Khare et al., 2013; Lemhöfer & Broersma, 2012). Surprisingly, only two studies provided detailed (objective) information about the L2 proficiency (Dimitropoulou et al., 2011b; Nakayama et al., 2016). To move the field forward, future studies are strongly encouraged to include objective L2 proficiency information so that future meta-analyses can shed more light on whether or not L2 proficiency moderates translation priming.

In addition, it is crucial to notice that studies in the field seldom reported the standardized effect sizes, which is not in line with the guidelines of the American Psychological Association (2010). Future meta-analyses would benefit if standardized effect sizes of translation priming are reported (for more discussion about effect sizes, see Judd et al., 2017; Lakens, 2013).

To summarize, we conducted the first meta-analysis of L1-L2 and L2-L1 masked translation priming in the literature, which quantitatively assessed the effect sizes of translation priming. The results not only revealed significant translation priming effects for both directions with larger L1-L2 than L2-L1 translation priming, but also revealed

that the effect sizes of L1-L2 were moderated by ISI and those of L2-L1 translation priming were moderated by the number of items used per cell. These findings contribute to the discussion about the mixed findings for the existing translation priming studies and provide methodological recommendations for future research. In the next chapter, a general discussion is presented, which considers theoretical and methodological implications of the overall findings reported in the thesis.

Chapter 7 General Discussion and Conclusions

7.1 Introduction

This chapter discusses the findings of five experimental chapters presented in this thesis. In the following sections, I will first summarise the results of the behavioural and electrophysiological experiments reported in Chapter 2 to Chapter 6. Next, I will discuss the methodological implications in the bilingual studies involving translation activation, and how the current findings contribute to the theoretical models of bilingual language processing. Finally, this chapter ends with discussions of the limitations of the experiments reported in the thesis, further directions and conclusions.

7.2 Summary of Findings

This thesis investigated L1 (Chinese) translation activation during L2 (English) word reading in Chinese-English bilinguals in an attempt to specify the source of translation activation (phonology: segment and/or tone, and/or orthography). Before the experimental investigations into translation activation, Chapter 2 presented Chinese translation norms for 1429 English words so that a good set of stimuli could be selected for the following chapters. In the norming study (Experiment 1), Chinese-English bilinguals were asked to provide the first Chinese translation that came to mind for 1429 English words. The results revealed that 71% of the English words received more than one correct translation indicating the large amount of translation ambiguity when translating from English to Chinese. The relationship between translation ambiguity and word frequency, concreteness and language proficiency was investigated. Results revealed that English word frequency was positively correlated with the number of alternative translations, whereas English word concreteness was negatively correlated with the number of translations. Importantly, regression analyses showed that the number of Chinese translations was predicted by word frequency and concreteness.

Furthermore, an interaction between these predictors revealed that the number of translations was more affected by word frequency for more concrete words than for less concrete words. In addition, mixed-effects modelling revealed that word frequency, concreteness and English language proficiency were all significant predictors of whether or not a dominant translation was provided. Finally, correlations between the word frequencies of English words and their Chinese dominant translations were higher for translation-unambiguous pairs than for translation-ambiguous pairs. The translation norms reported in Chapter 2 have been made available online for other researchers together with lexical information about the words (Wen & van Heuven, 2016). All the stimuli used in the following chapters were based on English words and their dominant Chinese translation selected from this translation norming study.

Chapter 3 reported two behavioural experiments, which investigated the activation of segmental and tonal information of Chinese translations of English word pairs in an attempt to clarify the source of hidden repetition priming observed in Chinese-English bilinguals (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a). In addition to manipulating the segmental and tonal information of the Chinese translations in the English word pairs, the type of task was also manipulated to investigate whether Chinese phonological activation during English word reading is task-dependent. Experiment 2 used an unmasked hidden priming paradigm with a semantic relatedness judgment task (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a), whereas Experiment 3 combined the masked priming paradigm with a lexical decision task (Zhang et al., 2011). Unexpectedly, both experiments revealed no significant hidden repetition priming effects. To better understand these unexpected results, Chapter 4 used electrophysiological measures to further investigate segmental and tonal activation of Chinese translations.

Chapter 4 presented two electrophysiological experiments during which Chinese-English bilinguals reading English (Experiment 4) or Chinese (Experiment 5) word pairs. Experiments involved a semantic relatedness judgment task with word pairs in which the repetition of segmental and tonal information of the hidden Chinese translations (Experiment 4) or of the Chinese words (Experiment 5) was systematically manipulated in semantically and orthographically unrelated word pairs. In the critical experiment (Experiment 4), a hidden segmental repetition effect was observed in the

behavioural data (faster responses for segmental repetition), the ERP data (the reduced N400), and the time-frequency data (the gamma-band power reduction for segmental repetition). This converging behavioural and electrophysiological evidence suggests the dominant role of Chinese segmental information activation of Chinese translations of English words during English word reading.

To further investigate why no behavioural effects were found in Chapter 3, Chapter 5 investigated fast and automatic translation in three behavioural experiments using the masked priming paradigm with a lexical decision task. Experiment 6 investigated whether Chinese phonology was the source of the hidden repetition priming effects observed in Zhang et al. (2011), and Experiment 7 investigated whether these effects were affected by morpheme complexity and prime duration. Unexpectedly, both experiments failed to find any behavioural evidence of a hidden repetition priming effect in the masked priming paradigm. An exact replication of Zhang et al. (2011) was conducted in Experiment 8, which again revealed no evidence of hidden repetition priming. In contrast, cross-language priming data obtained with the same group of participants did reveal masked translation priming and crucially character repetition priming with masked Chinese primes, indicating that automatic Chinese translation is limited to English target words in the masked priming paradigm.

However, based on the findings of Experiment 8, it is still unclear whether or not the masked English (L2) prime words were processed up to the semantic level. One potential clue is to look at the findings of L2-L1 masked translation priming in the literature. Although the non-cognate masked translation priming paradigm has been widely used in the last 25 years to investigate word processing in bilinguals, studies using this paradigm reported mixed findings for L2-L1 translation priming (more discussion see Section 1.3 in Chapter 1). To investigate whether L2-L1 translation priming in the masked translation priming paradigm reported the literature is reliable and to explore the possible factors modulating the translation priming effects, Chapter 6 used a meta-analysis of 64 masked priming lexical decision experiments across 24 studies. For the first time in the literature, this meta-analysis assessed the effect sizes of L1-L2 and L2-L1 non-cognate translation priming effects in bilinguals. The meta-analysis also investigated the influence of potential moderators of translation priming effects. The results provided a clear evidence of significant translation priming effects

for both directions, with L1-L2 translation priming significantly larger than L2-L1 translation priming (i.e., effect size of 0.86 vs. 0.31). The analyses also revealed that L1-L2 translation effect sizes were moderated by the interval between prime and target (ISI), whereas L2-L1 translation effect sizes were modulated by the number of items per cell. Importantly, significant L2-L1 translation priming effects revealed by this meta-analysis suggested that masked L2 prime words were processed up to the semantic level.

7.3 Methodological and Theoretical Implications

7.3.1 The Importance of Translation Norms

As discussed in Chapter 1, translation equivalents have been used extensively to investigate bilingual language processing. However, translation equivalents do not always have one-to-one mappings. Despite the prevalence of translation ambiguity (Prior et al., 2007; Tokowicz et al., 2002; Tseng et al., 2014), studies using translation equivalents often neglect this potential issue. Therefore, one important contribution of this thesis is the translation norms, which will be a useful resource for researchers investigating Chinese-English bilingual language processing. For example, English words with Chinese dominant translations could be used in translation priming studies. In fact, Task 2 in Experiment 8 used English words with its Chinese dominant translations in the masked translation priming paradigm, and found a 232 ms priming effect from L1 to L2. This is the first study using translation norms to investigate translation priming in Chinese-English bilinguals. Although previous studies have reported robust L1-L2 translation priming effects in Chinese-English bilinguals, the average L1-L2 translation priming effect from 10 lexical decision experiments with Chinese-English bilinguals (see Table 4 in Section 1.3.3: Chen et al., 2014a; Chen et al., 2014b; Jiang, 1999; Luo et al., 2013; Wang & Forster, 2015; Xia & Andrews, 2015) is just 58 ms (range: 18 -127 ms), which is much smaller than what was found in Experiment 8 (232 ms). This contrast seems to suggest that translation priming could be stronger when English words with Chinese dominant translations (based on translation norms) are used.

To further illustrate the potential methodological issue, Chinese translations of the English words in Wu and Thierry (2010) were checked against the translation norms presented in Chapter 2.[33] Wu and Thierry (2010) reported that 12 Chinese-English bilinguals provided more than 96% of correct translations for 400 English words used in their study. Of those 400 words used in Wu and Thierry (2010), 342 (85.5%) English words were also included in the translation norms presented in Chapter 2. In total, the Chinese translations of 201 English words (58.8% out of 342) were the dominant translations provided by our translation norms. In other words, the Chinese translations of 141 English words were not the dominant translations according to our translation norms. One potential explanation could be that the proficiency of the bilinguals in the two studies were different. As revealed in Chapter 2, English language proficiency significantly predicted whether or not a dominant translation was provided. Therefore, differences in proficiency across the studies may have impacted the outcome of the dominant translations. However, Wu and Thierry (2010) did not mention whether the correct translations were dominant translations (i.e., most frequent translations provided for that English word) or not.

It is clear that future studies that use translation equivalents to investigate bilingual language processing (e.g., translation recognition, masked translation priming, hidden repetition priming) should take the issue of translation ambiguity into consideration and carefully select translation pairs from translation norms. In addition, the meta-analysis (Chapter 6) revealed that L2-L1 translation effect sizes were modulated by the number of items per cell. Therefore, future studies should use a sufficient number of items to investigate the L2-L1 translation priming, which could easily be accomplished by using databases with translation norms (e.g., Prior et al., 2007; Tokowicz et al., 2002; Wen & van Heuven, 2016).

7.3.2 Measures of Second Language Proficiency

As revealed in Chapter 2, the objective LexTALE scores (Lemhöfer & Broersma, 2012)

[33] The stimulus list was not included in Wu and Thierry (2010). However, they were included in Wu's (2008) thesis (thesis was obtained through the British Library e-thesis online service).

seem to be a more reliable measure of English proficiency than self-rated proficiency scores because participants' translation performance significantly correlated with their LexTALE scores, but not their self-rated proficiency scores. To provide a good description of participants' L2 proficiency, the English proficiency of all Chinese-English bilinguals was measured by LexTALE as well as self-rated scores in all the experiments reported in this thesis. It is important to note that this is the first attempt in the investigation of hidden repetition priming effects to objectively and subjectively measure Chinese-English bilinguals' L2 proficiency at the time of testing, because Zhang et al. (2011) only used self-rated scores and Wu and Thierry (e.g., Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a, 2012b) just reported the means and ranges of the IELTS scores. It is important to provide the detailed information of participants' L2 proficiency because future research using the same paradigm may need to compare the results across studies.

As discussed in previous chapters (i.e., Table 1 in Chapter 2 and the discussion in Chapter 6), one potential issue in the field of bilingual language processing is the overwhelming use of self-assessed proficiency ratings as an estimate of L2 proficiency and various types of objective assessment of L2 proficiency (for more discussion, see de Bruin, Carreiras, & Duñabeitia, 2017; Hulstijn, 2012; Lemhöfer & Broersma, 2012). For example, across the 24 studies analysed in Chapter 6, self-assessed proficiency ratings were used in 12 studies, and two studies which reported objective L2 proficiency information used different measures (Dimitropoulou et al., 2011b; Nakayama et al., 2016). Therefore, the comparison across studies is often limited because it is unclear whether participants tested in different studies are comparable in L2 proficiency, and importantly, it is difficult to explore the relationship between L2 proficiency and the experimental effects in the literature. To move the field forward, one promising solution, as suggested by Lemhöfer and Broersma (2012), is to use LexTALE as a standard English proficiency test. Lemhöfer and Broersma (2012) found that LexTALE scores significantly correlated with the scores of Quick Placement Test (2001) and TOEIC (Test of English for International Communication), which demonstrated that the LexTALE scores are a valid measure of general English proficiency. Although it is unclear whether the LexTALE scores correlates with receptive vocabulary tests, e.g., British Picture Vocabulary Scale (BPVS, Dunn, Dunn, & Whetton, 1982), it is no doubt that LexTALE is a good measure of English vocabulary knowledge (e.g., Bultena,

Dijkstra, & van Hell, 2014; Diependaele, Lemhöfer, & Brysbaert, 2013; Khare et al., 2013). In addition, it takes less than 5 minutes to complete LexTALE, and it is freely available (http://www.lextale.com/).

To explore the potential impact of L2 proficiency on translation activation, additional analyses were conducted for Task 2 of Experiment 8 (Chapter 5), which found significant behavioural effects of masked translation priming and character repetition priming. In the mixed-effects modelling analysis, LexTALE scores was not a significant predictor, suggesting that masked translation priming and character repetition priming effects are not moderated by English proficiency. Although this result seems disappointing, it is indeed in line with the important conclusion of van Hell and Tanner (2012). In their excellent narrative review, van Hell and Tanner (2012) summarised the research which tested different groups of bilinguals in the same studies (e.g., beginning learners, intermediate learners, and advance learners) to investigate the impact of L2 proficiency and cross-language activation. Despite different measures of L2 proficiency were used across studies, the comparison across groups in the same studies revealed that cross-language activation occurred in all the proficient bilinguals, but not beginning learners. Because all the participants tested in Experiment 8 were proficient Chinese-English bilinguals and they were relatively homogenous in language profiles, it is not very surprising that their LexTALE scores were not a significant predictor of the observed priming effects.

7.3.3 The Issue of Effect Size and Power

The current thesis revealed that the hidden repetition priming effects observed behaviorally in Zhang et al. (2011) was not robust. As discussed in Chapter 5, the most plausible explanation of the replication failure in the three experiments (Experiment 6-8) is the small effect size (i.e., $d = 0.349$) and low power (i.e., 0.61) in the original study. Although the observed effect size is small, it is very likely to be an over-estimation of the true effect size. Evidence for this comes from the results of the meta-analysis presented in Chapter 6, which reported that the effect size of L2-L1 translation priming in masked priming paradigm was 0.31. Because the L2-L1 translation priming effect involves semantic overlap, it is not likely that a hidden repetition effect, which only involves form repetition through translation, would be larger than the L2-L1 translation

priming effect. Thus, it is likely that the true effect size of the hidden repetition effect is even smaller than 0.349. In other words, the true power of Zhang et al. (2011) should be even smaller than 0.61, which leads to overestimating the replicability of the findings.

To explore the influence of power and effect size on the replication studies, a simulation study was conducted. Vasishth and Nicenboim (2016) provided an R function (sim_data.R) which can stimulate reaction times data of an experiment using a Latin-square design. Using the relevant parameters from the results of Zhang et al. (2011) (i.e., standard deviations in by-item and by-subject analysis, the correlations of reaction times in two experimental conditions, and the intercept of the mixed-effects modelling analysis), 1,000 experiments were simulated with 32 participants, 10 items per cell for a true priming effect (in millisecond instead of Cohen's d) estimated at 4 ms, 8 ms, and 12 ms respectively.[34] The results of mixed-effects modelling are presented in Figure 40 with each point representing a significant effect and the dashed line depicting the true effect (e.g., 4 ms, 8 ms, and 12 ms). Figure 40 also shows the proportions of significant effects (numbers in black for facilitatory effects and numbers in grey for inhibitory effects). For example, if the true priming effect was 8 ms, 51% of the 1,000 simulated experiments had a significant facilitatory effect and 5% of the simulated experiments showed a significant inhibitory effect. Therefore, the true power with the true priming effect of 8 ms is estimated to be 0.51. In other words, if the true effect is as small as 8 ms, the chance of finding significant result in a replication experiment with 32 participant's and 10 items per cell would be 0.51. Although a limitation of this simulation is that the magnitude of the true effect is unknown, this simulation study provides a general picture of the relationship between effect size, power and replications. Critically, one important observation of this simulation study is that none of the simulated experiments observed a priming effect as large as 75 ms found in Zhang et al. (2011). Therefore, it seems to suggest that the priming effect observed in Zhang et al. (2011) is an over-estimation of the true effect.

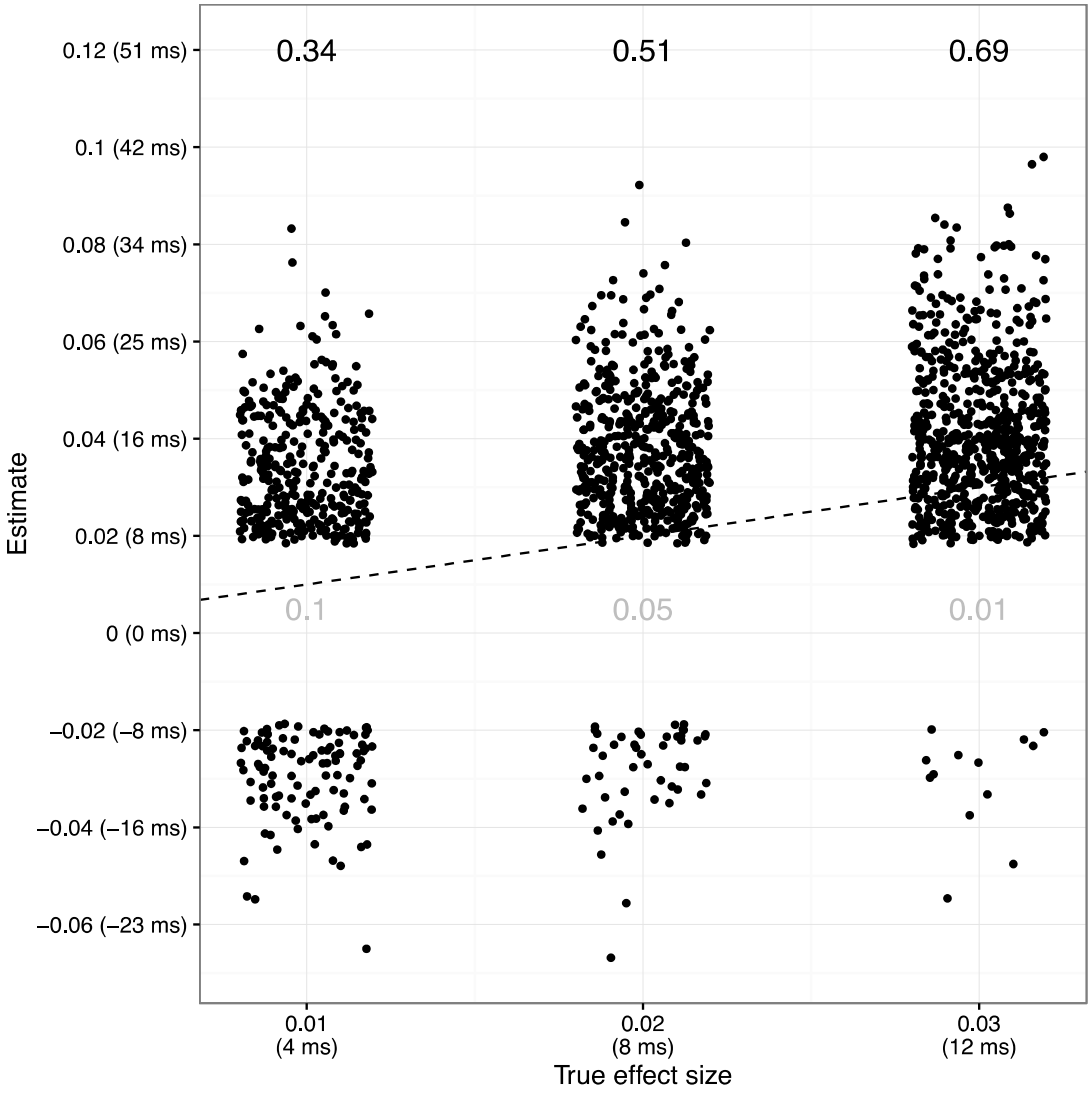
To further investigate the relationship between the statistical power and the magnitude

[34] Because the average L2-L1 translation priming effect was 11 ms based on all the experiments used in the meta-analysis, it is unlikely that the true effect of the hidden repetition priming is larger than 12 ms.

of the true effect, another set of simulations were conducted. Using the same parameters as the previous simulation, 1,000 experiments were simulated with a true priming effect estimated at 16 ms, 21 ms, and 25 ms respectively. The results of the mixed-effects modelling are presented in Figure 41. An important finding of these simulations is that the true power with a true priming effect of 16 ms is estimated to be larger than 0.8. It is unlikely that the true power is larger than the observed power of Zhang et al. (2011) (i.e., 0.61). Therefore, a true effect of 16 ms in unlikely.

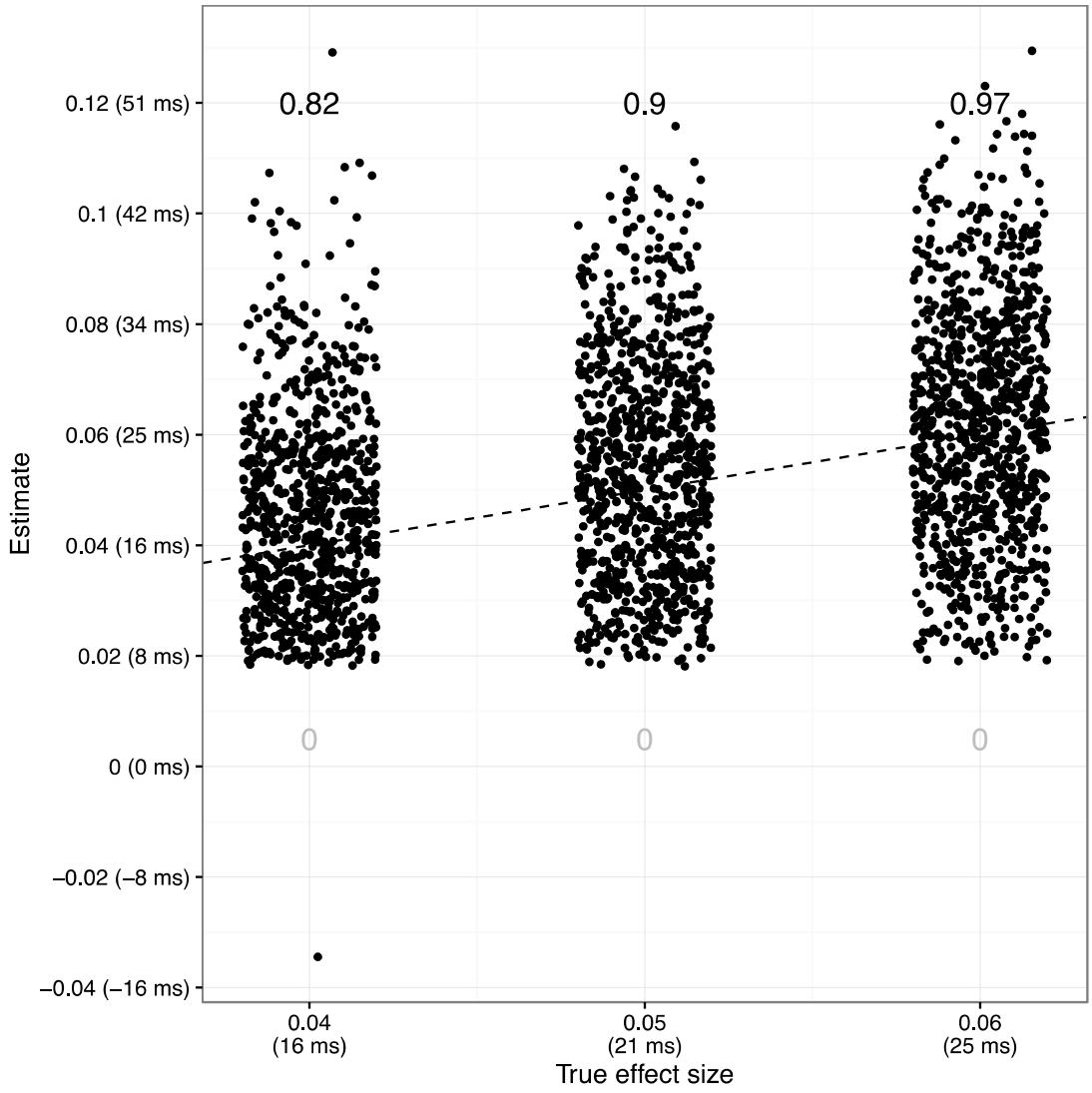
Taken together, the comparison between the hidden repetition priming effects, the meta-analysis of translation priming and the simulation study further illustrate the concerns over the issue of effect size and power in Zhang et al. (2011).

Figure 40. Results of simulated experiments with a true priming effect of 4, 8 12 ms



Note. Each point representing a significant effect and the dashed line depicting the true effect (e.g., 4 ms, 8 ms, and 12 ms); numbers in black are the proportions of significant facilitatory effects numbers in grey are the proportions of significant inhibitory effects.

Figure 41. Results of simulated experiments with a true priming effect of 16, 21 and 25 ms.



Note. Each point representing a significant effect and the dashed line depicting the true effect (e.g., 16 ms, 21 ms, and 25 ms); numbers in black are the proportions of significant facilitatory effects numbers in grey are the proportions of significant inhibitory effects

7.3.4 The Relationship between Electrophysiological and Behavioral Evidence

Although the present thesis failed to replicate the hidden repetition priming effects observed behaviorally in Zhang et al. (2011), Chapter 4 reported a successful replication of Wu and Thierry's (2010) ERP study. This contrast between the findings of behavioural and electrophysiology experiments does not necessarily mean that there is a lack of alignment between the behavioural evidence and the electrophysiological evidence of the hidden repetition priming effects. First of all, it is important to note that different paradigms were used in electrophysiological and behavioural investigations presented in Chapter 4 and Chapter 5. The electrophysiological experiment (Experiment 4) in Chapter 4 used the semantic relatedness judgment task with both the prime word and target word visible to bilinguals, whereas the three behavioural experiments (Experiment 6 to 8) presented in Chapter 5 used in the masked priming paradigm with only the target word visible. Furthermore, Task 2 of Experiment 8 found a character repetition effect, which suggested that Chinese translations of the visible English targets were activated. This interpretation is indeed in line with the results of the electrophysiological experiment, which revealed behavioural (faster responses for segmental repetition) and electrophysiological (the reduced N400 and the gamma-band power reduction for segmental repetition) evidence of Chinese translation activation for visible English primes and targets. Taken together, the behavioural and electrophysiological evidence of translation activation are consistent.

However, the results of Chinese semantic relatedness judgment experiment (Experiment 5) did reveal a mismatch between behavioural data and electrophysiological data, because no significant hidden repetition priming effects were found in the behavioural data whereas the electrophysiological data revealed significant effects (a larger LPC for segmental repetition, and the gamma-band power changers for tonal and segmental repetition). Although the lack of alignment between behavioural data and neuroimaging data (e.g., ERP, fMRI) is not uncommon, there is no consensus in the literature for how to interpret this issue, and its precise origin is still debated (for more discussion, see Li & Tokowicz, 2012; Poeppel, 2012; Valian, 2016). For example, to explain null effects in the behavioual data in contrast to significant effects in neuroimaging data, some researchers argue that neuroimaging measures may be more sensitive than behavioural measures in capturing processing differences (e.g., van Hell

& Kroll, 2013; van Heuven & Coderre, 2015a). Alternatively, another view is that differences in the neuroimaging measures can be purely attributed to differences in neural mechanism only when the behavioual data are comparable (Bialystok, 2016). Nevertheless, a different view is that the behavioual evidence is required to better understand any differences in the neuroimaging data (e.g., García-Pentón, Fernández García, Costello, Duñabeitia, & Carreiras, 2016; Paap, 2016; Paap, Johnson, & Sawi, 2015). It is clear that there is no simple answer to the alignment problem, and more studies are needed to understand the link between the behavioural and neuroimaging data.

7.3.5 The Time-frequency Analysis as an Insightful Approach to Investigate Visual Word Recognition

The present thesis has demonstrated that conducting the time-frequency analysis to investigate the event-related brain oscillations is an insightful approach that compliments the conventional analysis on the event-related brain potentials. For example, although the ERP data of the Chinese semantic relatedness judgment experiment (Experiment 5) only revealed the evidence of segmental repetition effects in terms of a larger LPC, the time-frequency data provided evidence for effects of both segmental and tonal repetition. A decrease in the gamma-band power was elicited by segmental repetition, whereas an increase in the gamma-band power was elicited by tonal repetition. Although it is unclear why segmental and tonal repetition lead to either an increase or decrease in the gamma-band power, the current findings indicate that there is a functional dissociation between Chinese segment and tone in the time-frequency data. Furthermore, the ERP data of the English semantic relatedness judgment experiment (Experiment 5) also revealed a decrease in the gamma-band power elicited by segmental repetition of Chinese translations in English word pairs. In other words, the converging evidence of segmental repetition in Chinese word pairs or Chinese translations of English word pairs has been observed in the time-frequency data. This converging evidence in the time-frequency data is in contrast to the different patterns in the ERP data because the segmental repetition in Chinese translations of English pairs elicited a reduced N400 instead of a larger LPC found in the segmental repetition of the Chinese word pairs. Taken together, it seems that the time-frequency analysis is an insightful approach to investigate monolingual and bilingual word

recognition. It is also important to highlight that this is one of the few studies on visual word recognition that conducted the time-frequency analysis (Matsumoto & Iidaka, 2008; Mellem et al., 2013) although the time-frequency analysis has been applied in many studies involving sentence reading (for a review, see Bastiaansen & Hagoort, 2006; Bastiaansen et al., 2011).

7.3.6 The Origin of Hidden Repetition Priming Effects

In the literature, it is widely accepted that hidden repetition priming effects are the evidence of translation activation (Morford et al., 2011; Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a; Zhang et al., 2011). The hidden repetition priming effect are consistent with the assumption of non-selective lexical access in bilinguals (Dijkstra & van Heuven, 2002; van Heuven et al., 1998). However, Costa, Pannunzi, Deco, and Pickering (2016) recently challenged the hidden repetition priming effects as empirical evidence of Chinese translation activation during English word reading, and argued that hidden repetition priming effects could be caused by transfer from Chinese representations to English representations as a consequence of learning. To support their alternative interpretation, Costa et al. (2016) created a computer model to simulate the acquisition process of English words (L2) after Chinese words (L1) were acquired. The model used the Hebbian learning model and included three English words (i.e., *ham*, *train*, *apple*) and their Chinese translation equivalents (e.g., 火腿, 火车, 苹果). During the training stage when connections between words were established, the model assumed that a given English word (e.g., *ham*) would activate its Chinese translation (e.g., 火腿, meaning "ham"), the form-related Chinese word through the Chinese network (e.g., 火车, meaning "train") and its English counterpart (e.g., *train*). After training, the model had learned three English words and their Chinese translations. In their simulations with the trained model, the input of an English word (e.g., *ham*) impacted the activity of the English word (e.g., *train*) if they were form-related in their Chinese translations. These results nicely captured the findings of Thierry and Wu (2007). Critically, Costa et al. (2016) then removed the three Chinese words from the model and ran the same simulation again. Their rational was that the removal of Chinese words would resemble a situation when Chinese is not activated. Interestingly, they found the same pattern again. Therefore, Costa et al. (2016) concluded that the impact of *ham* on *train* was not necessarily due to co-activation of the Chinese translations

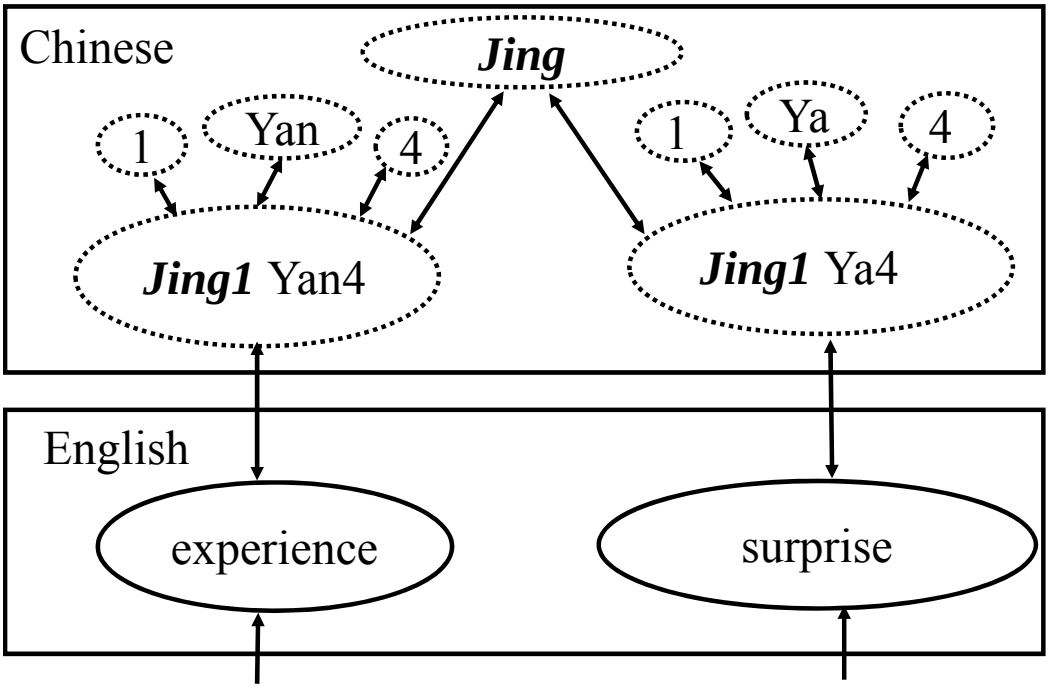
during English word reading, and that hidden repetition priming effects may originate from a direct connection between *ham* and *train*, which was transferred from the connection of their Chinese translations 火腿 (meaning "ham") and 火车(meaning "train").

Costa et al. (2016) modelling work is a neat approach that supports their alternative interpretation of the hidden repetition priming effects. Their approach is in line with the popular idea that "neurons that fire together wire together". However, the underlying logic of their approach seem problematic to me. It is important to note that Costa et al. (2016) directly connected 火腿 (meaning "ham") and 火车 (meaning "train") in the Chinese network, and this connection is critical for establishing the connection between *ham* and *train*. During the training stage, they assumed that the presented word *ham* would activate the Chinese word 火车 (meaning "train"), which is related in form to the Chinese translation of *ham* (i.e., 火腿). This might work for a simplified model containing only 3 Chinese words. However, it seems unlikely that such connections are actually established in real language use. As mentioned in Section 1.8.1, it is very common for many Chinese characters to appear in more than one Chinese word, for example, according to SUBTLEX-CH (Cai & Brysbaert, 2010), 238 Chinese words contain the character 火. In fact, the 99,121 Chinese words in the Chinese subtitle corpus (SUBTLEX-CH, Cai & Brysbaert, 2010) contain only 5,936 unique Chinese characters. Therefore, it seems unlikely that the word *ham* would activate all the Chinese words containing the character 火. Furthermore, the current chapter and previous studies (Wu et al., 2013; Wu & Thierry, 2010, 2012a, 2012b) have suggested the dominant role of Chinese phonological activation during English word reading. Following the logic used in Costa et al. (2016), it would suggest that *ham* would activate all the Chinese words with the homophone 火. As noted in Section 1.8.1, an important linguistic fact of Chinese is that there are only 404 different syllables in Mandarin Chinese (Duanmu, 2007) and 1,255 types of syllables when combined with tones (Duanmu, 2006). Consequently, it is common that a given syllable appears in many different Chinese words. Therefore, it is extremely unlikely that a given English word would activate all the Chinese words which share a homophonic character with its Chinese translation. Although Costa et al. (2016) admitted neglecting the predominant role of phonology, they might have underestimated the consequence of this limitation.

Taken together, the account of Costa et al. (2016) for hidden repetition effects fails to take into account the linguistic facts of the Chinese language and the empirical evidence of Chinese phonology as the major contributor in the hidden repetition effects. As a result, the rationale for the connection between *ham* and 火车 (meaning "train") during training is unlikely and this connection is not theoretically well-justified. More importantly, if the direct connection between *ham* and *train* exists as suggest by Costa et al. (2016), it is puzzling why the hidden repetition priming effects (e.g., *ham-train*) were not always observed in masked priming experiments (e.g., Experiment 6-8 in Chapter 5).

Therefore, consistent with the dominant view in the literature (Morford et al., 2011; Thierry & Wu, 2007; Wu & Thierry, 2010, 2012a; Zhang et al., 2011), I interpret hidden repetition priming effects as a result of Chinese translation activation during English word reading. A theoretical framework of hidden repetition priming effects is shown in Figure 42. This simplified model depicts the co-activation of Chinese and English during the semantic relatedness judgment task. When an English prime word (i.e., *experience*) is presented, the Chinese translation (represented by Pinyin, i.e., /Jing1Yan4/ which consists of segmental information /Jing/ /Yan/ and tonal information /1/ /4/) are also activated through semantic connections. In a similar vein, an English target word (i.e., *experience*) also activates its Chinese translation (i.e., /Jing1Ya4/), and receives feedback from the shared segment (i.e., /Jing/) through the Chinese translation. This simplified model highlights the dominant role of Chinese segment as the origin of the hidden repetition priming effects.

Figure 42. The origins of the phonological repetition priming effect in Chinese translations during English word reading.

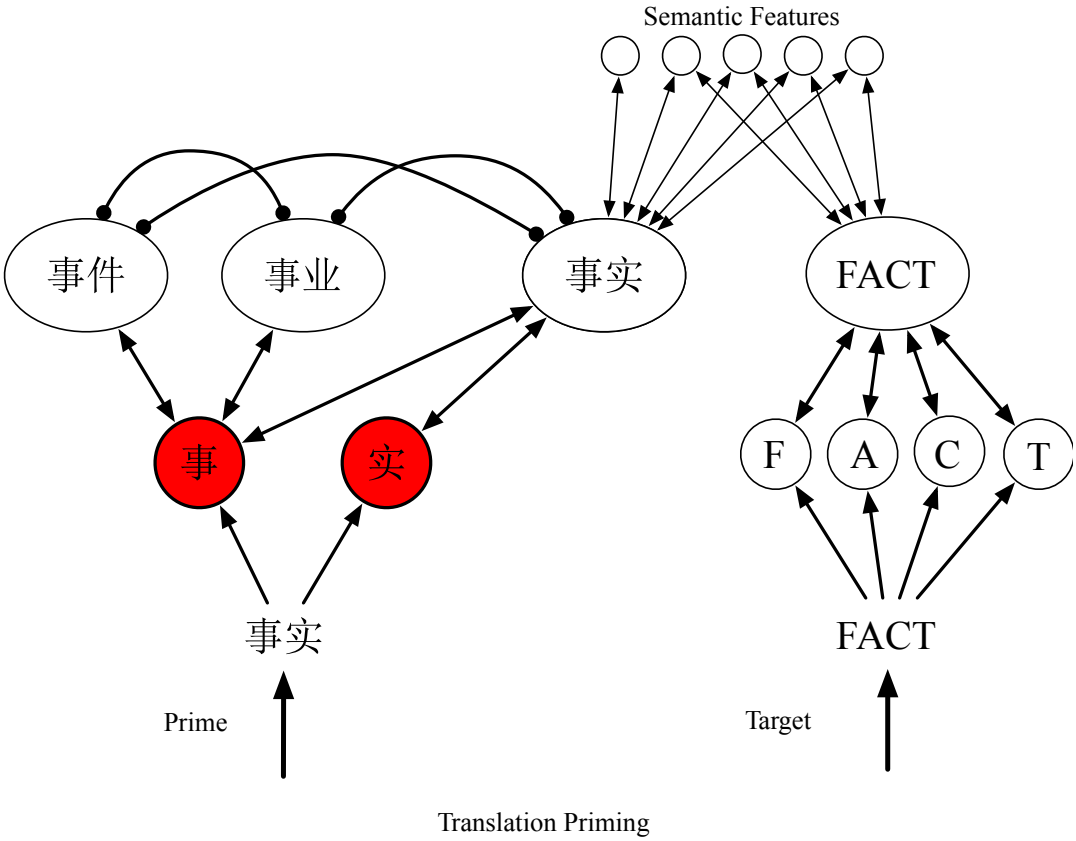


7.3.7 The Origin of Masked Translation Priming

Based on the discussion presented in the previous sections, it is clear that an important finding from both the behavioural and electrophysiological investigations in this thesis is that Chinese translation are activated when visible English words are presented to Chinese-English bilinguals. In this section, I will focus on the implications of this finding for masked translation priming effects. As discussed in Chapter 1, translation priming effects are generally assumed to originate from the semantic level (Finkbeiner et al., 2004; Schoonbaert et al., 2009), and the origin of priming effects are explained as the semantic features activated by primes facilitating the recognition of the translation-equivalent target. However, if the target words in L2 could automatic activate its L1 translation, it is possible that the translation activations of target words also receive feedback from the primes. This feedback is analogous to a resonance effect, which potentially contributes to the translation priming effects. In other words, the origin of translation priming effects could be a combination of the pre-activation semantics and through-translation lexical form overlap.

Following this rationale, a potential theoretical framework of L1-L2 translation priming is depicted in Figure 43 (arrows with triangle heads indicating activation and arrows with circular heads indicating inhibition). When the L1 prime (e.g., 事实 [fact]) is presented, its semantic features are quickly activated. At the same time, the presented L1 prime would inhibit L1 word candidates which that have different letters or characters (e.g., 事件 [incident], 事业 [career]). As discussed Section 1.3.4, a N250 effect was observed in L1-L2 translation priming, indicating that L2 translations are activated via connections from the pre-activated semantics (Hoshino et al., 2010; Midgley et al., 2009; Schoonbaert et al., 2011). When the L2 target (e.g., *fact*) is presented, its recognition is facilitated by the pre-activated semantics and lexical form overlap. Critically, the L2 target also activates its L1 translation, and thus receives feedback from the L1 prime. Taken together, the theoretical framework depicted in Figure 43 captures the origin of translation priming effects in terms of pre-activation semantics and through-translation lexical form overlap.

Figure 43. The origins of masked translation priming effects with L1 primes and L2 targets



Note. In the following example, the L1 prime is Chinese word 事实 [fact]. The activation of 事实 would facilitate its English translation equivalent *fact* and inhibit other L1 word candidates, e.g., 事件 [incident], 事业 [career].

7.4 Limitations and Future Directions

An obvious limitation of the present thesis is that the bilinguals tested in all the experiments were unbalanced bilinguals with higher proficiency in Chinese than English. Therefore, it is unclear whether the same patterns would be observed in very proficient Chinese-English bilinguals. Although no behavioural evidence of the hidden repetition effects was found in the masked priming paradigm, this finding does not rule out the possibility that hidden repetition effects might be obtained with very proficient Chinese-English bilinguals (e.g., professional interpreters). Due to the fast processing

of the L2 in highly proficient bilinguals, it is possible that L1 translations of both masked L2 (English) primes and L2 targets be quickly activated, which may generate hidden repetition effects. Therefore, future studies should test very proficient Chinese-English bilinguals to investigate the impact of L2 proficiency on translation activation. In addition, although the converging conclusion from the electrophysiological and behavioural experiments reported in Chapter 4 is that Chinese translations are activated when reading English words, it is unclear whether translation activation would occur if the English words are presented for a shorter duration (e.g., 250 ms). The relationship between presentation duration and translation activation should be investigated in the future studies as well. Another potential line of future research is to further explore the phonological activation of Chinese translations during English word reading. A potential auditory study could use the same design of Experiment 4 to investigate whether segmental activation of Chinese translation could elicit N400 reduction and gamma-band power changes. In other words, it is crucial to investigate whether the findings in visual word recognition tasks can be replicated with auditory word recognition tasks. Another potential study could compare the character repetition effects (e.g., 事实 - *career*, the Chinese translation of *career* is 事业, sharing the same character and pronunciation with 事实) and the phonological repetition effects (e.g., 经验 - *surprise*, the Chinese translation of *surprise* is 惊讶, whose first character is a homophone of 经) in the masked priming paradigm.

Importantly, although this thesis proposed a potential theoretical framework to explain the origin of the masked translation priming effects, this theoretical framework should be evaluated in a computation model, for example, using a localist-connectionist approach. In addition, more electrophysiological investigations are needed for masked translation priming effects. As discussed in Chapter 1, only a few electrophysiological studies have used the masked translation priming paradigm (Duñabeitia et al., 2010a; Geyer, Holcomb, Midgley, & Grainger, 2011; Hoshino et al., 2010; Midgley et al., 2009; Schoonbaert et al., 2011) and only one study tested different-script bilinguals (Hoshino et al., 2010). Therefore, more electrophysiological investigation into masked translation priming effects are needed, especially studies involving different-script languages. Furthermore, future electrophysiological studies may conduct time-frequency analyses as well to further explore translation priming effects.

7.5 Conclusions

This thesis investigated Chinese translation activation during English word reading. In a series of behavioural and electrophysiological experiments using carefully selected stimuli, English words were subliminally or visibly presented to Chinese-English bilinguals in an attempt to seek the source of automatic unconscious translation (segment/tone, phonology/orthography) and to explore to what extent translation activation is automatic. The behavioural investigations revealed that automatic translation only occurs for target words, which were visible to bilinguals, but not for the subliminally presented masked primes. In addition, in the electrophysiological study, the event-related brain potentials and event-related brain oscillations provided evidence for the predominant role of Chinese segmental (consonants and vowels) activation of Chinese translations during English word reading.

References

Allen, D. B., & Conklin, K. (2013). Cross-linguistic similarity and task demands in Japanese-English bilingual processing. *PLoS ONE, 8*(8), e72631. doi: 10.1371/journal.pone.0072631

Allen, D. B., Conklin, K., & van Heuven, W. J. B. (2015). Making sense of the Sense Model: Translation priming with Japanese-English bilinguals. *The Mental Lexicon, 10*(1), 32-52. doi: 10.1075/ml.10.1.02all

Altarriba, J., & Basnight-Brown, D. M. (2007). Methodological considerations in performing semantic-and translation-priming experiments across languages. *Behavior Research Methods, 39*(1), 1-18. doi: 10.3758/BF03192839

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *British Medical Journal, 311*, 485. doi: 10.1136/bmj.311.7003.485

Alvarez, P. R, Holcomb, P. J., & J, Grainger. (2003). Accessing word meaning in two languages: An event-related brain potential study of beginning bilinguals. *Brain and Language, 87*(2), 290-304. doi: 10.1016/S0093-934X(03)00108-1

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th Ed.). Washington, DC: American Psychological Association.

Antón-Méndez, I., & Gollan, T. H. (2010). Not just semantics: Strong frequency and weak cognate effects on semantic association in bilinguals. *Memory & Cognition, 38*(6), 723-739. doi: 10.3758/MC.38.6.723

Aparicio, X., & Lavaur, J.-M. (2015). Masked translation priming effects in visual word recognition by trilinguals. *Journal of Psycholinguistic Research*, 1-20. doi: 10.1007/s10936-015-9409-8

Athanasopoulos, P. (2015). Conceptual representation in bilinguals: The role of language specificity and conceptual change. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing*. Cambridge: Cambridge University Press.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Changes in theta and beta oscillations as signatures of novel word consolidation. *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn_a_00801

Barker, C. (2016). from http://semarch.linguistics.fas.nyu.edu/barker/Syllables/index.txt

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi: 10.1016/j.jml.2012.11.001

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition, 35*(5), 953-965. doi: 10.3758/BF03193468

Bastiaansen, M. C. M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research, 159*, 179-196. doi: 10.1016/S0079-6123(06)59012-0

Bastiaansen, M. C. M., & Hagoort, P. (2015). Frequency-based Segregation of Syntactic and Semantic Unification during Online Sentence Level Language Comprehension. *Journal of Cognitive Neuroscience*, 1-13. doi: 10.1162/jocn_a_00829

Bastiaansen, M. C. M., Mazaheri, A., & Jensen, O. (2011). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 31-49). New York: Oxford University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language, 26*(6), 658-672. doi: 10.1016/0749-596X(87)90108-2

Bialystok, E. (2016). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism* doi: 10.1075/lab.15040.bia

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences, 16*(4), 240-250. doi: 10.1016/j.tics.2012.03.001

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes, 22*(5), 633-660. doi: 10.1080/01690960601000746

Boada, R., Sanchez-Casas, R, Gavilán, J. M., García-Albea, J. E., & Tokowicz, N. (2013). Effect of multiple translations and cognate status on translation recognition performance of balanced bilinguals. *Bilingualism: Language and Cognition, 16*(01), 183-197. doi: 10.1017/S1366728912000223

Borenstein, M., Hedges, L. V. , Higgins, J. P.T. , & Rothstein, H. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*(1), 34-44. doi: 10.1162/jocn.1993.5.1.34

Brysbaert, M. (2013). LEXTALE_FR: A fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica, 53*(1), 23-37. doi: 10.5334/pb-53-1-23

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition, 13*(3), 359-371. doi: 10.1017/S1366728909990344

Brysbaert, M., Verreyt, N., & Duyck, W. (2010). Models as hypothesis generators and models as roadmaps. *Bilingualism: Language and Cognition, 13*(3), 383-384. doi: 10.1017/S1366728910000167

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911. doi: 10.3758/s13428-013-0403-5

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *The Quarterly Journal of Experimental Psychology, 67*(6), 1214-1241. doi: 10.1080/17470218.2013.853090

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5*(6), e10729. doi: 10.1371/journal.pone.0010729

Chen, B., Liang, L., Cui, P., & Dunlap, S. (2014a). The priming effect of translation equivalents across languages for concrete and abstract words. *Acta psychologica, 153*, 147-152. doi: 10.1016/j.actpsy.2014.10.004

Chen, B., Zhou, H., Gao, Y., & Dunlap, S. (2014b). Cross-language translation priming asymmetry with Chinese-English bilinguals: A test of the sense model. *Journal of Psycholinguistic Research, 43*(3), 225-240. doi: 10.1007/s10936-013-9249-3

Chen, H.-C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin & Review, 8*(3), 511-518. doi: 10.3758/BF03196186

Chen, J.-Y., & Chen, T.-M. (2007). Form encoding in Chinese word production does not involve morphemes. *Language and Cognitive Processes, 22*(7), 1001-1020. doi: 10.1080/01690960701190249

Coderre, E. L., & van Heuven, W. J. B. (2014). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology, 5*(1070). doi: 10.3389/fpsyg.2014.01070

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. Cambridge: MIT Press.

Colomé, A. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45*(4), 721-736. doi: 10.1006/jmla.2001.2793

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*, 497-505. doi: 10.1080/14640748108400805

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In D. S. (Ed.), *Attention and Performance IV* (pp. 535-555). Hillsdale,NJ: Erlbaum.

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*(3), 453-467. doi: 10.1177/0267658316637401

Coren, S. (1992). *The Left-Hander Syndrome: The Causes and Consequences of Left-Handedness*. Free Press: New York.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283-1296. doi: 10.1037/0278-7393.26.5.1283

Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2016). Do bilinguals automatically activate their native language when they are not using it? *Cognitive Science*, n/a-n/a. doi: 10.1111/cogs.12434

Davis, C., Sanchez-Casas, R., Garcia-Albea, J. E., Guasch, M., Molero, M., & Ferré, P. (2010). Masked translation priming: Varying language experience and word type with Spanish–English bilinguals. *Bilingualism: Language and Cognition, 13*(2), 137-155. doi: 10.1017/S1366728909990393

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST Dataset of Language Proficiency. *Frontiers in Psychology, 8*(522). doi: 10.3389/fpsyg.2017.00522

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1001-1018. doi: 10.1037/0278-7393.18.5.1001

de Groot, A. M. B. (2011). *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York: Psychology Press.

de Groot, A. M. B., & Comijs, H. (1995). Translation recognition and translation production: Comparing a new and an old tool in the study of bilingualism. *Language Learning, 45*(3), 467-509. doi: 10.1111/j.1467-1770.1995.tb00449.x

de Groot, A. M. B., Dannenburg, L., & van hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language, 33*(5), 600-629. doi: 10.1006/jmla.1994.1029

de Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language, 30*(1), 90-123. doi: 10.1016/0749-596X(91)90012-9

de Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning, 47*(2), 215-264. doi: 10.1111/0023-8333.71997007

DeFrancis, J. (1984). *The Chinese language: Fact and fantasy*: University of Hawaii Press.

Degani, T., Prior, A., Eddington, C. M., Arêas da Luz Fontes, A., & Tokowicz, N. (2013). Beyond semantic ambiguity: Effects of translation ambiguity in a single-language context. *Manuscript in revision*.

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE, 10*(3). doi: 10.1371/journal.pone.0121945

Dien, J. (2012). Applying principal components analysis to event-related potentials: A Tutorial. *Developmental Neuropsychology, 37*(6), 497-517. doi: 10.1080/87565641.2012.697503

Dien, J. (2017). Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. *International Journal of Psychophysiology, 111*, 42-56. doi: 10.1016/j.ijpsycho.2016.09.006

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*. doi: 10.3389/fpsyg.2014.00781

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology, 66*(5), 843-863. doi: 10.1080/17470218.2012.720994

Dijksta, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 179-201). New York: Oxford University.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 178-201). New York: Oxford University.

Dijkstra, T. (2007). The multilingual lexicon. In M. G. Gaskell (Ed.), *Handbook of Psycholinguistics* (pp. 251-265). Oxford, UK: Oxford University Press.

Dijkstra, T., De Bruijn, E., Schriefers, H., & Ten Brinke, S. (2000). More on interlingual homograph recognition: language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition, 3*(1), 69-78.

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41*(4), 496-528. doi: 10.1006/jmla.1999.2654

Dijkstra, T., & Rekké, S. (2010). Towards a localist-connectionist model of word translation. *The Mental Lexicon, 5*(3), 401-420. doi: 10.1075/ml.5.3.08dij

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175-197. doi: 10.1017/S1366728902003012

Dijkstra, T., Van Jaarsveld, H., & Brinke, S. T. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*(01), 51-66. doi: 10.1017/S1366728998000121

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011a). Masked translation priming effects with low proficient bilinguals. *Memory & Cognition, 39*(2), 260-275. doi: 10.3758/s13421-010-0004-9

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011b). Two words, one meaning: Evidence of automatic co-activation of translation equivalents. *Frontiers in Psychology, 2*. doi: 10.3389/fpsyg.2011.00188

Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition, 8*(03), 221-238. doi: 10.1017/S1366728905002270

Dong, Y., & Lin, J. (2013). Parallel processing of the target language during source language comprehension in interpreting. *Bilingualism: Language and Cognition, 16*(3), 682-692. doi: 10.1017/S1366728913000102

Duñabeitia, J. A., Dimitropoulou, M., Dowens, M. G., Molinaro, N., & Martin, C. (2016). The electrophysiology of the bilingual brain. In R. R. Heredia, J. Altarriba & A. B. Cieślicka (Eds.), *Methods in Bilingual Reading Comprehension Research* (pp. 265-312). New York: Springer.

Duñabeitia, J. A., Dimitropoulou, M., Uribe-Etxebarria, O., Laka, I., & Carreiras, M. (2010a). Electrophysiological correlates of the masked translation priming effect with highly proficient simultaneous bilinguals. *Brain Research, 1359*, 142-154. doi: 10.1016/j.brainres.2010.08.066

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010b). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology, 57*(2), 98-107. doi: 10.1027/1618-3169/a000013

Duanmu, S. (2006). Chinese (Mandarin): Phonology. In K. Brown (Ed.), *Encyclopedia of language and linguistics (Second Edition)* (pp. 3651-3355). Oxford: Elsevier.

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford: Oxford University Press.

Dunn, L. M., Dunn, L. M., & Whetton, C. (1982). *British Picture Vocabulary Scale*. Windsor, England: NFER-Nelson.

Duyck, W., Assche, E. V., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 663-679. doi: 10.1037/0278-7393.33.4.663

Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance, 30*(5), 889-906. doi: 10.1037/0096-1523.30.5.889

Duyck, W., & Brysbaert, M. (2008). Semantic access in number word translation: The role of crosslingual lexical similarity. *Experimental Psychology, 55*(2), 102-112. doi: 10.1027/1618-3169.55.2.102

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: New evidence from Dutch-French bilinguals. *Experimental Psychology, 56*(3), 173-179. doi: 10.1027/1618-3169.56.3.173

Eddington, C. M., & Tokowicz, N. (2013). Examining English–German translation ambiguity using primed translation recognition. *Bilingualism: Language and Cognition, 16*(02), 442-457. doi: 10.1017/S1366728912000387

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. doi: 10.3758/BRM.41.4.1149

Ferré, P., Sánchez-Casas, R., & Guasch, M. (2006). Can a horse be a donkey? Semantic and form interference effects in translation recognition in early and late proficient and nonproficient Spanish-Catalan bilinguals. *Language Learning, 56*(4), 571-608. doi: 10.1111/j.1467-9922.2006.00389.x

Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *Quarterly Journal of Experimental Psychology: Section A, 45*(3), 353-372.

Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society, 31*(2), 119-122. doi: 10.3758/BF03334157

Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology Section A, 47*(2), 365-382. doi: 10.1080/14640749408401116

Feynman, R. P. (1974). Cargo cult science. *Engineering & Science, 37*, 10-13.

Finkbeiner, M., Forster, K. I., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51*(1), 1-22. doi: 10.1016/j.jml.2004.01.004

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152-170. doi: 10.1111/j.1469-8986.2007.00602.x

Forster, K. I. (2013). How many words can we read at once? More intervenor effects in masked priming. *Journal of Memory and Language, 69*(4), 563-573. doi: 10.1016/j.jml.2013.07.004

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680-698. doi: 10.1037/0278-7393.10.4.680

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116-124. doi: 10.3758/BF03195503

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software, 8*(15), 1-27.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*(1), 71-99. doi: 10.1037/0033-2909.123.1.71

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 104-115. doi: 10.1037/0096-1523.13.1.104

García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition and Neuroscience, 31*(3), 303-327. doi: 10.1080/23273798.2015.1068944

Geyer, A., Holcomb, P. J., Midgley, K. J., & Grainger, J. (2011). Processing words in two languages: An event-related brain potential study of proficient bilinguals.

Journal of Neurolinguistics, 24(3), 338-351. doi: 10.1016/j.jneuroling.2010.10.005

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1122-1139. doi: 10.1037/0278-7393.23.5.1122

Grabner, R. H., Brunner, C., Leeb, R., Neuper, C., & Pfurtscheller, G. (2007). Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Research Bulletin, 72*(1), 57-65. doi: 10.1016/j.brainresbull.2007.01.001

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*(1), 1-35. doi: 10.1080/01690960701578013

Grainger, J. , Midgley, K. J., & Holcomb, P. J. . (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickman (Eds.), *Language Acquisition across linguistic and cognitive systems* (pp. 267-284). Philadelphia: John Benjamins.

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In H. Richard Jackson (Ed.), *Advances in Psychology* (Vol. Volume 83, pp. 207-220): North-Holland.

Grainger, J., & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes, 13*(6), 601-623. doi: 10.1080/016909698386393

Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The time course of orthographic and phonological code activation. *Psychological Science, 17*(12), 1021-1026. doi: 10.1111/j.1467-9280.2006.01821.x

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67-81. doi: 10.1017/S1366728998000133

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language, 27*(2), 210-223. doi: 10.1016/0093-934X(86)90016-7

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*(12), 1711-1725. doi: 10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology, 48*(12), 1726-1737. doi: 10.1111/j.1469-8986.2011.01272.x

Grosjean, F. (1998). Transfer and language mode. *Bilingualism: Language and Cognition, 1*(175-176). doi: 10.1017/S1366728998000285

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.

Grosjean, F., & Li, P. (2013). *The psycholinguistics of bilingualism*. New York: John Wiley & Sons.

Grossi, G., Savill, N., Thomas, E., & Thierry, G. (2012). Electrophysiological cross-language neighborhood density effects in late and early English-Welsh bilinguals. *Frontiers in Psychology, 3*(408). doi: 10.3389/fpsyg.2012.00408

Gu, Y. (2006). Chinese. In K. Brown (Ed.), *Encyclopedia of language and linguistics (Second Edition)*. Oxford: Elsevier.

Guasch, M., Sánchez-Casas, R., Ferré, P., & García-Albea, J. E. (2008). Translation performance of beginning, intermediate and proficient Spanish-Catalan bilinguals: Effects of form and semantic relations. *The Mental Lexicon, 3*(3), 289-308. doi: 10.1075/ml.3.3.03gua

Guo, T., Misra, M., Tam, J. W., & Kroll, J. F. . (2012). On the time course of accessing meaning in a second language: An electrophysiological investigation of translation recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1165-1186. doi: 10.1037/a0028076

Guo, T., Peng, D., & Liu, Y. (2005). The role of phonological activation in the visual semantic retrieval of Chinese characters. *Cognition, 98*(2), B21-B34. doi: 10.1016/j.cognition.2005.02.004

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438-441. doi: 10.1126/science.1095455

Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2016.08.004

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6*(2), 107-128. doi: 10.3102/10769986006002107

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*, 490-499. doi: 10.3102/10769986018003271

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition, 1*(3), 213-229. doi: 10.1017/S1366728998000364

Herrmann, C. S., Grigutsch, M., & Busch, N. A. (2005). EEG oscillations and wavelet analysis. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 229-259). Cambridge: MIT Press.

Herrmann, C. S., Rach, S., Vosskuhl, J., & Strüber, D. (2014). Time–frequency analysis of event-related potentials: A brief tutorial. *Brain Topography, 27*(4), 438-450. doi: 10.1007/s10548-013-0327-5

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician, 55*(1), 19-24. doi: 10.1198/000313001300339897

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience, 18*(10), 1631-1643. doi: 10.1162/jocn.2006.18.10.1631

Hornby, A. S. (2010). *Oxford Advanced Learner's English-Chinese Dictionary* The Commercial Press.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106*(1), 501-511. doi: 10.1016/j.cognition.2007.02.001

Hoshino, N., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2010). An ERP investigation of masked cross-script translation priming. *Brain Research, 1344*, 159-172. doi: 10.1016/j.brainres.2010.05.005

Hoshino, N., & Thierry, G. (2012). Do Spanish–English bilinguals have their fingers in two pies – or is it their toes? An electrophysiological investigation of semantic access in bilinguals. *Frontiers in Psychology, 3*(9). doi: 10.3389/fpsyg.2012.00009

Huang, C. T. J., Li, Y. H. A., & Simpson, A. (2014). *The Handbook of Chinese Linguistics*: Wiley Blackwell.

Hulstijn, J. H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition, 15*(2), 422-433. doi: 10.1017/S1366728911000678

Institute of Linguistic from Chinese Academy of Social Science. (2005). *Xiandai Hanyu Cidian (Contemporary Chinese Dictionary)*: The Commercial Press.

Ioannidis, J. P. A., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials, 4*(3), 245-253. doi: 10.1177/1740774507079441

Ito, A., Martin, A. E., & Nieuwland, M. S. (2016). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, 1-12. doi: 10.1080/23273798.2016.1242761

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language, 44*, 21-31. doi: 10.1006/jmla.2000.2747

Jia, X., Wang, S., Zhang, B., & Zhang, J. X. (2013). Electrophysiological evidence for relation information activation in Chinese compound word comprehension. *Neuropsychologia, 51*(7), 1296-1301. doi: 10.1016/j.neuropsychologia.2013.03.024

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition, 2*(1), 59-75.

Jiang, N. (2015). Six decades of research on bilingual presenattion. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 29-84). Cambridge: Cambridge University Press.

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language, 44*(1), 32-51. doi: 10.1006/jmla.2000.2737

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944-946. doi: 10.1111/2041-210X.12225

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*(1). doi: 10.1146/annurev-psych-122414-033702

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(02), 163-178. doi: 10.1111/1469-8986.3720163

Küssner, M. B., de Groot, A. M. B., Hofman, W. F., & Hillen, M. A. (2016). EEG beta power but not background music predicts the recall scores in a foreign-vocabulary learning task. *PLoS ONE, 11*(8), e0161387. doi: 10.1371/journal.pone.0161387

Kaan, E., Wayland, R., & Keil, A. (2013). Changes in oscillatory brain networks after lexical tone training. *Brain Sciences, 3*(2), 757. doi: 10.3390/brainsci3020757

Kappenman, E.S., & Luck, S. J. . (2012). ERP components the ups and downs of brainwave recordings. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 3-30). New York: Oxford University Press.

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., . . . Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology, 51*(1), 1-21. doi: 10.1111/psyp.12147

Kerkhofs, R., Dijkstra, T., Chwilla, D. J., & de Bruijn, E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research, 1068*(1), 170-183. doi: 10.1016/j.brainres.2005.10.087

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287-304. doi: 10.3758/s13428-011-0118-4

Khare, V., Verma, A., Kar, B., Srinivasan, N., & Brysbaert, M. (2013). Bilingualism and the increased attentional blink effect: Evidence that the difference between bilinguals and monolinguals generalizes to different levels of second language proficiency. *Psychological Research, 77*(6), 728-737. doi: 10.1007/s00426-012-0466-4

Kielar, A., Meltzer, J. A., Moreno, S., Alain, C., & Bialystok, E. (2014). Oscillatory responses to semantic and syntactic violations. *Journal of Cognitive Neuroscience, 26*(12), 2840-2862. doi: 10.1162/jocn_a_00670

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology, 124*(10), 2062-2063. doi: 10.1016/j.clinph.2013.03.024

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language, 49*(4), 484-499. doi: 10.1016/S0749-596X(03)00093-7

Kinoshita, S., & Lupker, S. J. (2003). *Masked priming: The state of the art*. Hove: Psychology Press.

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language, 100*(3), 257-261. doi: 10.1016/j.bandl.2006.07.006

Kondrak, G. (2013). Word similarity, cognation, and translational equivalence. In L. Borin & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 375-386). Berlin: De Gruyter Mouton.

Kong, L., Zhang, J. X., Kang, C., Du, Y., Zhang, B., & Wang, S. (2010). P200 and phonological processing in Chinese word recognition. *Neuroscience Letters, 473*(1), 37-41. doi: 10.1016/j.neulet.2010.02.014

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language, 109*(2), 68-74. doi: 10.1016/j.bandl.2008.06.002

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12*(5), 535-540. doi: 10.1038/nn.2303

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology, 25*(5), 497-514. doi: 10.1080/20445911.2013.799170

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition, 9*(2), 119-135. doi: 10.1017/S1366728906002483

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris & R. Sykes (Eds.), *Practical aspects of memory* (Vol. 2, pp. 389-395). London: Wiley.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming evidence for asymmetric connections between bilingual memory

representations. *Journal of Memory and Language, 33*(2), 149-174. doi: 10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 49-71). Malden: Blackwell.

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381. doi: 10.1017/S136672891000009X

Kuperman, Vi., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 802-823. doi: 10.1037/a0030859

Kutas, M., & Federmeier, Kara D. (2011). Thirty years and counting: Finding meaning in the N400 component of the eent-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potential reflect semantic incongruity. *Science, 207*, 203-205. doi: 10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163. doi: 10.1038/307161a0

La Heij, W., de Bruyn, Eddy, Elens, Erik, Hartsuiker, Rob, Helaha, Debbie, & van Schelven, Lie. (1990). Orthographic facilitation and categorical interference in a word-translation variant of the Stroop task. *Canadian Journal of Psychology/Revue canadienne de psychologie, 44*(1), 76. doi: 10.1037/h0084236

La Heij, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward translation: Evidence for concept mediation. *Journal of Memory and Language, 35*(5), 648-665. doi: 10.1006/jmla.1996.0034

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. doi: 10.3389/fpsyg.2013.00863

Laxén, J., & Lavaur, J. (2010). The role of semantics in translation recognition: Effects of number of translations, dominance of translations and semantic relatedness

of multiple translations. *Bilingualism: Language and Cognition, 13*(2), 157-183. doi: 10.1017/S1366728909990472

Lehmann, D. (1987). Principles of spatial analysis. In A. S. Gevins & A. Rémond (Eds.), *Handbook of electroencephalography and clinical neurophysiology: Revised series, Vol. 1. Analysis of electrical and magnetic signals* (pp. 390-354). Amsterdam: Elsevier.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48*(6), 609-621. doi: 10.1016/0013-4694(80)90419-8

Lemhöfer, K. , & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language specific and generalized lexical decision. *Memory & Cognition, 32*(4), 533-550. doi: 10.3758/BF03195845

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*(2), 325-343. doi: 10.3758/s13428-011-0146-0

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 12-31. doi: 10.1037/0278-7393.34.1.12

Li, C., Lin, C.-Y., Wang, M., & Jiang, N. (2013). The activation of segmental and tonal information in visual word recognition. *Psychonomic Bulletin & Review, 20*(4), 773-779. doi: 10.3758/s13423-013-0395-2

Li, C., Wang, M., & Lin, C. Y. (2016). Recent language experience influences cross-language activation in bilinguals with different scripts. *International Journal of Bilingualism*. doi: 10.1177/1367006916635837

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In E. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 59-85). North-Holland: Elsevier Science.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17*(8–9), 1345-1362. doi: 10.1016/j.neunet.2004.07.004

Li, P., & Tokowicz, N. (2012). Psycholinguistics of second language acquisition. In P. Robinson (Ed.), *Routledge encyclopedia of second language* (pp. 526-535). London: Routledge.

Li, P., & Zhao, X. (2015). Computational modeling of bilingual language acquisition and processing: Conceptual and methodological considerations. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 85-107). Cambridge: Cambridge University Press.

Libben, M. R., & Titone, D. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 381. doi: 10.1037/a0014875

Lin, Y.-H. (2007). *The sounds of Chinese*. Cambridge: Cambridge University Press.

Linck, J. A., Kroll, J. F., & Sunderman, G. L. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507-1515. doi: 10.1111/j.1467-9280.2009.02480.x

Liu, P. D., & McBride-Chang, C. (2010). Morphological processing of Chinese compounds from a grammatical view. *Applied Psycholinguistics, 31*(04), 605-617. doi: 10.1017/S0142716410000159

Liu, Y., Perfetti, C. A., & Hart, L. (2003). ERP evidence for the time course of graphic, phonological, and semantic information in Chinese meaning and pronunciation decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1231. doi: 10.1037/0278-7393.29.6.1231

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7*(4), 618-630. doi: 10.3758/BF03212999

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT press

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146-157. doi: 10.1111/psyp.12639.

Luo, X., Cheung, H., Bel, D., Li, Li, Chen, L., & Mo, L. (2013). The roles of semantic sense and form-meaning connection in translation priming. *The Psychological Record, 63*(1), 193.

Lupker, S. J., Perea, M., & Nakayama, M. (2015). Non-cognate translation priming effects in the same-different task: Evidence for the impact of “higher level”

information. *Language, Cognition and Neuroscience, 30*(7), 781-795. doi: 10.1080/23273798.2015.1015430

Ma, B., Wang, X., & Li, D. (2015). The Processing of visual and phonological configurations of Chinese one-and two-Character words in a priming task of semantic categorization. *Frontiers in Psychology, 6*. doi: 10.3389/fpsyg.2015.01918

Macizo, P., Bajo, T., & Martín, M. C. (2010). Inhibitory processes in bilingual language comprehension: Evidence from Spanish–English interlexical homographs. *Journal of Memory and Language, 63*(2), 232-244. doi: 10.1016/j.jml.2010.04.002

Mainy, N., Jung, J., Baciu, M., Kahane, P., Schoendorff, B., Minotti, L., . . . Lachaux, J.-P. (2008). Cortical dynamics of word recognition. *Human Brain Mapping, 29*(11), 1215-1230. doi: 10.1002/hbm.20457

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(02), 173-193. doi: 10.1017/S0142716403000092

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97-115. doi: 10.1017/S1366728903001068

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86*(1), 70-82. doi: 10.1016/S0093-934X(02)00535-7

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190. doi: 10.1016/j.jneumeth.2007.03.024

Matsumoto, A., & Iidaka, T. (2008). Gamma band synchronization and the formation of representations in visual word processing: Evidence from repetition and homophone priming. *Journal of Cognitive Neuroscience, 20*(11), 2088-2096. doi: 10.1162/jocn.2008.20136

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review, 88*(5), 375-405. doi: 10.1037/0033-295X.88.5.375

Mellem, M. S., Friedman, R. B., & Medvedev, A. V. (2013). Gamma-and theta-band synchronization during semantic priming reflect local and long-range lexical–

semantic networks. *Brain and Language, 127*(3), 440-451. doi: 10.1016/j.bandl.2013.09.003

Metting van Rijn, A.C., Peper, A., & Grimbergen, C. A. (1990). High-quality recording of bioelectric events. *Medical and Biological Engineering and Computing, 28*(5), 389-397.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., . . . Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176-182. doi: 10.1126/science.1199644

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2009). Masked repetition and translation priming in second language learners: A window on the time-course of form and meaning activation using ERPs. *Psychophysiology, 46*(3), 551-565. doi: 10.1111/j.1469-8986.2009.00784.x

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience, 23*(7), 1634-1647. doi: 10.1162/jocn.2010.21463

Midgley, K. J., Holcomb, P. J., van Heuven, W. J. B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research, 1246*, 123-135. doi: 10.1016/j.brainres.2008.09.078

Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word processing in Hindi–English sequential bilinguals: An eye tracking visual world study. *Reading and Writing, 27*(1), 129-151. doi: 10.1007/s11145-013-9436-5

Misra, M., & Holcomb, P. J. (2003). Event–related potential indices of masked repetition priming. *Psychophysiology, 40*(1), 115-130. doi: 10.1111/1469-8986.00012

Moldovan, C. D., Demestre, J., Ferré, P., & Sánchez-Casas, R. (2016). The role of meaning and form similarity in translation recognition in highly proficient balanced bilinguals: A behavioral and ERP study. *Journal of Neurolinguistics, 37*, 1-11. doi: 10.1016/j.jneuroling.2015.07.002

Moldovan, C. D., Sánchez-Casas, R., Demestre, J., & Ferré, P. (2012). Interference effects as a function of semantic similarity in the translation recognition task in bilinguals of Catalan and Spanish. *Psicologica: International Journal of Methodology and Experimental Psychology, 33*(1), 77-110.

Moon, J., & Jiang, N. (2012). Non-selective lexical access in different-script bilinguals. *Bilingualism: Language and Cognition, 15*(01), 173-180. doi: 10.1017/S1366728911000022

Morey, R. D., & Rouder, J. N. (2016). BayesFactor 0.9.12-2 Comprehensive R Archive Network. from https://cran.r-project.org/web/packages/BayesFactor/index.html

Morford, J. P., Wilkinson, E., Villwock, A., Piñar, P., & Kroll, J. F. (2011). When deaf signers read English: Do written words activate their sign translations? *Cognition, 118*(2), 286-292. doi: 10.1016/j.cognition.2010.11.006

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105-125. doi: 10.1037/1082-989X.7.1.105

Myers, J. (2006). Processing Chinese compounds: A survey of the literature. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 169-196). Oxford: Oxford University Press.

Nakayama, M., Ida, K., & Lupker, S. J. (2016). Cross-script L2-L1 noncognate translation priming in lexical decision depends on L2 proficiency: Evidence from Japanese–English bilinguals. *Bilingualism: Language and Cognition*, 1-22. doi: 10.1017/S1366728915000462

Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger, J. S. Nairne, I. Neath & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 69-93). Washington, DC, US: American Psychological Association.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402-407. doi: 10.3758/BF03195588

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas-Part II. *Language and Linguistics Compass, 10*, 591-613. doi: 10.1111/lnc3.12207

Nuzzo, R. (2015). How scientists fool themselves-and how they can stop. *Nature, 526*(7572), 182-185. doi: 10.1038/526182a

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese

but with segments in English. *Cognition, 115*(2), 282-302. doi: 10.1016/j.cognition.2010.01.001

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2013). Close but not proximate: The significance of phonological segments in speaking depends on their functional engagement. *Proceedings of the National Academy of Sciences, 110*(1), E3-E3. doi: 10.1073/pnas.1217032110

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2010). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 1-9. doi: 10.1155/2011/156869

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi: 10.1126/science.aac4716

Paap, K. R. (2016). The neuroanatomy of bilingualism: Will winds of change lift the fog? *Language, Cognition and Neuroscience, 31*(3), 331-334. doi: 10.1080/23273798.2015.1082607

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265-278. doi: 10.1016/j.cortex.2015.04.014

Paulmann, S., Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2006). Is bilingual lexical access influenced by language context? *NeuroReport, 17*(7), 727-731. doi: 10.1097/01.wnr.0000214400.88845.fa

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language, 68*(4), 315-332. doi: 10.1016/j.jml.2012.12.003

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10. doi: 10.3389/neuro.11.010.2008

Peng, D., Liu, Y., & Wang, C. (1999). How is access representation organized? The relation of polymorphemic words and their morphemes in Chinese. In J. Wang, A. Inhoff & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 65-89). Mahwah, NJ: Lawrence Erlbaum.

Perea, M., Dunabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language, 58*(4), 916-930. doi: 10.1016/j.jml.2008.01.003

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model some implications of research on Chinese for general theories of reading. *Psychological Review, 112*(1), 43-59. doi: 10.1037/0033-295X.112.1.43

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 101-118. doi: 10.1037/0278-7393.24.1.101

Perfetti, C. A., & Tan, L. H. . (1999). The constituency model of Chinese word identification. In A. Inhoff, H.-S. Chen & J. Wang (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 115-134). Mahwah, NJ.

Perfetti, C. A., & Zhang, S. (1991). Phonological processes in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(4), 633-643.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 24-33. doi: 10.1037/0278-7393.21.1.24

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., . . . Rugg, M. D. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology, 37*(02), 127-152. doi: 10.1111/1469-8986.3720127

Poeppel, D. (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology, 29*(1-2), 34-55. doi: 10.1080/02643294.2012.710600

Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23-38. doi: 10.1016/S0022-5371(84)90489-4

Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods, 39*(4), 1029-1038. doi: 10.3758/BF03193001

Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. *Psychological Science, 27*(7), 1036-1042. doi: 10.1177/0956797616645672

Qasem, M., & Foote, R. (2010). Cross-language lexical activation: A test of the revised hierarchical and morphological decomposition models in Arabic-English bilinguals. *Studies in Second Language Acquisition, 32*(01), 111-140. doi: 10.1017/S0272263109990271

Qi, Y. (2011). *Cognate Status and Cross-script Priming with Chinese-English Bilinguals and English-Chinese Bilinguals.* (MA Thesis), National University of Singapore.

Qin, R., Maurits, N., & Maassen, B. (2016). N170 tuning in Chinese: Logographic characters and phonetic pinyin script. *Scientific Studies of Reading, 20*(5), 363-374. doi: 10.1080/10888438.2016.1199554

Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences of the USA, 109*(35), 14265–14270. doi: 10.1073/pnas.1200632109

Qu, Q., Damian, M. F., & Kazanina, N. (2013). Reply to O'Seaghdha et al.: Primary phonological planning units in Chinese are phonemically specified. *Proceedings of the National Academy of Sciences, 110*(1), E4-E4. doi: 10.1073/pnas.1217601110

Quick Placement Test. (2001). Oxford: Oxford University Press.

R Core Team. (2016). R: A Language and Environment for Statistical Computing. from http://www.R-project.org/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372. doi: 10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506. doi: 10.1080/17470210902816461

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition, 35*(02), 213-235. doi: 10.1017/S0272263112000861

Rodriguez-Fornells, A., De Diego Balaguer, R., & Münte, T. F. (2006). Executive Control in Bilingual Language Processing. *Language Learning, 56*, 133-190. doi: 10.1111/j.1467-9922.2006.00359.x

Rodriguez-Fornells, A., Rotte, M., Heinze, H.-J., Nosselt, T., & Munte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature, 415*(6875), 1026-1029. doi: 10.1038/4151026a

Rodriguez-Fornells, A., Schmitt, B. M., Kutas, M., & Munte, T. F. (2002). Electrophysiological estimates of the time course of semantic and phonological encoding during listening and naming. *Neuropsychologia, 40*(7), 778-787. doi: 10.1016/S0028-3932(01)00188-9

Rosenthal, R. (1991). *Meta-analytical procedures for social research*. Newbury Park, CA: Sage.

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review, 21*(2), 301-308. doi: 10.3758/s13423-014-0595-4

Rousselet, G., & Pernet, C. (2011). Quantifying the time course of visual object processing using ERPs: It's time to up the game. *Frontiers in Psychology, 2*(107). doi: 10.3389/fpsyg.2011.00107

Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W.-J., Bick, A., Frost, S. J., . . . Duñabeitia, J. A. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences, 112*(50), 15510-15515. doi: 10.1073/pnas.1509321112

Rugg, M. D., & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. *Brain and Language, 32*(2), 336-361. doi: 10.1016/0093-934X(87)90132-5

Sáchez-Casas, R. M., García-Albea, J. E., & Davis, C. W. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology, 4*(4), 293-310. doi: 10.1080/09541449208406189

Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition, 17*(3), 542-555. doi: 10.1017/S1366728913000643

Schepens, J., Dijkstra, T., Grootjen, F., & van Heuven, W. J. B. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLoS ONE, 8*(5), e63006. doi: 10.1371/journal.pone.0063006

Schneider, W. , Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition, 37*(5), 569-586. doi: 10.3758/MC.37.5.569

Schoonbaert, S., Holcomb, P. J., Grainger, J., & Hartsuiker, R. J. (2011). Testing asymmetries in noncognate translation priming: Evidence from RTs and ERPs. *Psychophysiology, 48*(1), 74-81. doi: 10.1111/j.1469-8986.2010.01048.x

Schutter, D. J. L. G., Leitner, C., Kenemans, J. L., & van Honk, J. . (2006). Electrophysiological correlates of cortico-subcortical interaction: A cross-frequency spectral EEG analysis. *Clinical Neurophysiology, 117*(2), 381-387. doi: 10.1016/j.clinph.2005.09.021

Shen, D., & Forster, K. I. (1999). Masked phonological priming in reading Chinese words depends on the task. *Language and Cognitive Processes, 14*(5-6), 429-459. doi: 10.1080/016909699386149

Sholl, A., Sankaranarayanan, A., & Kroll, J. F. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science, 6*(1), 45-49. doi: 10.1111/j.1467-9280.1995.tb00303.x

Shook, A., & Marian, V. (2016). The influence of native-language tones on lexical access in the second language. *The Journal of the Acoustical Society of America, 139*(6), 3102-3109. doi: 10.1121/1.4953692

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 0956797611417632. doi: 10.1177/0956797611417632

Singh, L. (2013). One world, two Languages: Cross‐language semantic priming in bilingual toddlers. *Child Development*. doi: 10.1111/cdev.12133

Singh, N., & Mishra, R. K. (2015). Unintentional activation of translation equivalents in bilinguals leads to attention capture in a cross-modal visual task. *PLoS ONE, 10*(3), e0120131. doi: 10.1371/journal.pone.0120131

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 776-784. doi: 10.1037/a0022531

Snodgrass, J. G. (1993). Translating versus picture naming: Similarities and differences. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 83-114). Amsterdam/Philadelphia: John Benjamins.

Spinks, J. A. , Liu, Y., Perfetti, C. A. , & Tan, L. H. (2000). Reading Chinese characters for meaning: The role of phonological information. *Cognition, 76*(1), B1-B11. doi: 10.1016/S0010-0277(00)00072-X

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281-284. doi: 10.1111/1467-9280.00151

Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, amu028. doi: 10.1093/applin/amu028

Stevens, M., Mandera, P., Keuleers, E., & Brysbaert, M. (2015). *How spurious can you get? Charting the sensitivity and specificity of typical psycholinguistic experiments with megastudy data.* Paper presented at the 19th Conference of the European Society for Cognitive Psychology.

Strijkers, K., Costa, A., & Thierry, G. (2009). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*. doi: 10.1093/cercor/bhp153

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. doi: 10.1037/h0054651

Sunderman, G. L., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28*(03), 387-422. doi: 10.1017/S0272263106060177

Sunderman, G. L., & Priya, K. (2012). Translation recognition in highly proficient Hindi–English bilinguals: The influence of different scripts but connectable phonologies. *Language and Cognitive Processes, 27*(9), 1265-1285. doi: 10.1080/01690965.2011.596420

Taft, M., & Chen, H.-C. (1992). Judging homophony in Chinese the influence of tones. *Advances in Psychology, 90*, 151-172. doi: 10.1016/S0166-4115(08)61891-9

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2*(01), 45-58.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *The Journal of Neuroscience, 16*(13), 4240-4249.

Tan, L. H., Liu, H. L., Perfetti, C. A., Spinks, J. A., Fox, P. T., & Gao, J. H. (2001). The neural system underlying Chinese logograph reading. *Neuroimage, 13*(5), 836-846. doi: 10.1006/nimg.2001.0749

Tan, L. H., & Perfetti, C. A. (1997). Visual Chinese character recognition: Does phonological information mediate access to meaning? *Journal of Memory and Language, 37*(1), 41-57. doi: 10.1006/jmla.1997.2508

Thierry, G. (2016). Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning, 66*(3), 690-713. doi: 10.1111/lang.12186

Thierry, G., & Wu, Y. J. (2004). Electrophysiological evidence for language interference in late bilinguals. *Neuroreport, 15*(10), 1555-1558. doi: 10.1097/01.wnr.0000134214.57469.c2

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign language comprehension. *Proceedings of the National Academy of Sciences of the USA, 104*(30), 12530-12535. doi: 10.1073/pnas.0609927104

Thomas, M.S.C., & van Heuven, W. J. B. . (2005). Computational models of bilingual comprehension. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 202-225). Oxford: Oxford University Press.

Tokowicz, N. (2014). *Translation ambiguity affects language processing, learning, and representation.* Paper presented at the Selected Proceedings of the 2012 Second Language Research Forum.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes, 22*(5), 727-779. doi: 10.1080/01690960601057068

Tokowicz, N., Kroll, J. F., de Groot, A. M. B., & van Hell, J. G. (2002). Number-of-translation norms for Dutch-English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, & Computers, 34*(3), 435-451. doi: 10.3758/BF03195472

Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science*. Chichester, UK: Wiley-Blackwell.

Tsang, Y.-K., & Chen, H.-C. (2010). Morphemic ambiguity resolution in Chinese: Activation of the subordinate meaning with a prior dominant-biased context. *Psychonomic Bulletin & Review, 17*(6), 875-881. doi: 10.3758/PBR.17.6.875

Tsang, Y.-K., & Chen, H.-C. (2013a). Early morphological processing is sensitive to morphemic meanings: Evidence from processing ambiguous morphemes. *Journal of Memory and Language, 68*(3), 223-239. doi: 10.1016/j.jml.2012.11.003

Tsang, Y.-K., & Chen, H.-C. (2013b). Morpho-semantic processing in word recognition: Evidence from balanced and biased ambiguous morphemes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1990-2001. doi: 10.1037/a0033701

Tsang, Y.-K., & Chen, H.-C. (2014). Activation of morphemic meanings in processing opaque words. *Psychonomic bulletin & review, 21*(5), 1281-1286. doi: 10.3758/s13423-014-0589-2

Tsang, Y.‐K., & Chen, H.-C. (2012). Eye movement control in reading: Logographic Chinese versus alphabetic scripts. *PsyCh Journal, 1*(2), 128-142. doi: 10.1002/pchj.10

Tse, C.-S., Yap, M. J., Chan, Y.-L., Sze, W.-P., Shaoul, C., & Lin, D. (2016). The Chinese lexicon project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*. doi: 10.3758/s13428-016-0810-5

Tseng, A. M., Chang, L.-Y., & Tokowicz, N. (2014). Translation ambiguity between English and Mandarin Chinese: The roles of proficiency and word characteristics. In J. W. Schwieter & A. Ferreira (Eds.), *The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (pp. 107-165). Cambridge: Cambridge Scholars Publishing.

Valian, V. (2016). Putting together bilingualism and executive function. *Linguistic Approaches to Bilingualism, 6*(5), 565-574. doi: 10.1075/lab.15044.val

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence

reading. *Journal of Memory and Language, 64*(1), 88-107. doi: 10.1016/j.jml.2010.08.006

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science, 20*(8), 923-927. doi: 10.1111/j.1467-9280.2009.02389.x

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin, 135*(3), 452-477. doi: 10.1037/a0015329

van Hell, J. G. , & De Groot, A, M. B. (1998a). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*(3), 193-211. doi: 10.1017/S1366728998000352

van Hell, J. G., & de Groot, A. M. B. (1998b). Disentangling context availability and concreteness in lexical decision and word translation. *The Quarterly Journal of Experimental Psychology Section A, 51*(1), 41-63. doi: 10.1080/713755752

van Hell, J. G., & Kroll, J. F. (2013). Using electrophysiological measures to track the mapping of words to concepts in the bilingual brain: A focus on translation. In J. Altarriba & L. Isurin (Eds.), *Memory, Language, and Bilingualism: Theoretical and Applied Approaches*. New York: Cambridge University Press.

van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning, 62*, 148-171. doi: 10.1111/j.1467-9922.2012.00710.x

van Heuven, W. J. B. (2005). Bilingual interactive activation models of word recognition in a second language. In C. V. & B. Bassetti (Eds.), *Second language writing systems* (pp. 260-288). Clevedon, UK: Multilingual Matters.

van Heuven, W. J. B., & Coderre, E. L. (2015a). A call for sophisticated statistical approaches and neuroimaging techniques to study the bilingual advantage. *Cortex, 73*, 330-331. doi: 10.1016/j.cortex.2015.08.006

van Heuven, W. J. B., & Coderre, E. L. (2015b). Orthographic processing in bilinguals. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 308-326). Cambridge: Cambridge University Press.

van Heuven, W. J. B., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews, 64*(1), 104–122. doi: 10.1016/j.brainresrev.2010.03.002

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458-483. doi: 10.1006/jmla.1998.2584

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190. doi: 10.1080/17470218.2013.850521

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex, 18*(11), 2706-2716. doi: 10.1093/cercor/bhn030

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas-Part I. *Language and Linguistics Compass, 10*(8), 349-369. doi: 10.1111/lnc3.12201

Verhagen, J., & Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General, 143*(4), 1457-1475. doi: 10.1037/a0036731

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition 35*(5), 938-952. doi: 10.3758/BF03193467

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology, 113*(4), 569-586. doi: 10.1016/j.jecp.2012.08.001

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779-804. doi: 10.3758/bf03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*(6), 632-638. doi: 10.1177/1745691612463078

Wang, C., & Peng, D. (2000). 重复启动作业中词的语义透明度的作用 [The role of semantic transparencies in the processing of compound words]. *Acta Psychologica Sinica, 32*, 127-132.

Wang, K., Mecklinger, A., Hofmann, J., & Weng, X. (2010). From orthography to meaning: An electrophysiological investigation of the role of phonology in accessing meaning of Chinese single-character words. *Neuroscience, 165*(1), 101-106. doi: 10.1016/j.neuroscience.2009.09.070

Wang, M., Li, C., & Lin, C. Y. (2015a). The contributions of segmental and suprasegmental information in reading Chinese characters aloud. *PLoS ONE, 10*(11), e0142060. doi: 10.1371/journal.pone.0142060

Wang, X. (2013). Language dominance in translation priming: Evidence from balanced and unbalanced Chinese–English bilinguals. *The Quarterly Journal of Experimental Psychology, 66*(4), 727-743. doi: 10.1080/17470218.2012.716072

Wang, X., & Forster, K. I. (2010). Masked translation priming with semantic categorization: Testing the Sense Model. *Bilingualism: Language and Cognition, 13*(3), 327-340. doi: 10.1017/S1366728909990502

Wang, X., & Forster, K. I. (2015). Is translation priming asymmetry due to partial awareness of the prime? *Bilingualism: Language and Cognition, 18*(4), 651-669. doi: 10.1017/S1366728914000650

Wang, X., Liu, A., Wu, Y., & Wang, P. (2013a). Rapid extraction of lexical tone phonology in Chinese characters: A visual mismatch negativity study. *PLoS ONE, 8*(2), e56778. doi: 10.1371/journal.pone.0056778

Wang, X., Wu, Y., Liu, A., & Wang, P. (2013b). Spatio-temporal dynamics of automatic processing of phonological information in visual words. *Scientific Reports, 3*. doi: 10.1038/srep03485

Wang, X., Yang, J., Yang, J., Mencl, W, E., Shu, H., & Zevin, J. D. (2015b). Language differences in the brain network for reading in naturalistic story reading and lexical decision. *PLoS ONE, 10*(5), e0124388. doi: 10.1371/journal.pone.0124388

Wen, Y., & van Heuven, W. J. B. (2016). Chinese translation norms for 1429 English words. *Behavior Research Methods*. doi: 10.3758/s13428-016-0761-x

Williams, J. N. (1994). The relationship between word meanings in the first and second language: Evidence for a common, but restricted, semantic code. *European Journal of Cognitive Psychology, 6*(2), 195-220. doi: 10.1080/09541449408520143

Witzel, N. O., & Forster, K. I. (2012). How L2 words are stored: The episodic L2 hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1608-1621. doi: 10.1037/a0028072

Wong, A. C.-N., Qu, Z., McGugin, R. W., & Gauthier, I. (2011). Interference in character processing reflects common perceptual expertise across writing systems. *Journal of Vision, 11*(15), 1-8. doi: 10.1167/11.1.15

Wong, A. W.-K., Wu, Y., & Chen, H.-C. (2014). Limited role of phonology in reading Chinese two-character compounds: Evidence from an ERP study. *Neuroscience, 256*, 342-351. doi: 10.1016/j.neuroscience.2013.10.035

Wu, Y. J. (2008). *The nature of cross-language activation in late Chinese-English bilinguals: A behavioural and event-related potential investigation.* (Ph.D.), Bangor University.

Wu, Y. J., Cristino, F., Leek, C., & Thierry, G. (2013). Non-selective lexical access in bilinguals is spontaneous and independent of input monitoring: Evidence from eye tracking. *Cognition, 129*(2), 418-425. doi: 10.1016/j.cognition.2013.08.005

Wu, Y. J., & Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *The Journal of Neuroscience, 30*(22), 7646-7651. doi: 10.1523/JNEUROSCI.1602-10.2010

Wu, Y. J., & Thierry, G. (2011). Event-related brain potential investigation of preparation for speech production in late bilinguals. *Frontiers in Psychology, 2*, 114. doi: 10.3389/fpsyg.2011.00114

Wu, Y. J., & Thierry, G. (2012a). How reading in a second language protects your heart. *The Journal of Neuroscience, 32*(19), 6485-6489. doi: 10.1523/JNEUROSCI.6119-11.2012

Wu, Y. J., & Thierry, G. (2012b). Unconscious translation during incidental foreign language processing. *NeuroImage, 59*(4), 3468-3473. doi: 10.1016/j.neuroimage.2011.11.049

Wu, Y. J., & Thierry, G. (2013). Fast modulation of executive function by language context in bilinguals. *The Journal of Neuroscience, 33*(33), 13533-13537. doi: 10.1523/JNEUROSCI.4760-12.2013

Xia, V., & Andrews, S. (2015). Masked translation priming asymmetry in Chinese-English bilinguals: Making sense of the Sense Model. *The Quarterly Journal of Experimental Psychology, 68*(2), 294-325. doi: 10.1080/17470218.2014.944195

Xu, S. (1985). *Shuowen Jiezi*. Beijing: Chunghwa Book Co.

Xu, Y., Pollatsek, A., & Potter, M. C. (1999). The activation of phonology during silent Chinese word reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 837-857. doi: 10.1037/0278-7393.25.4.838

Yang, Z., Wang, C., Feng, L., Li, S., & Tao, S. (2016). Unsuccessful letter–sound integration in English reading by native Chinese speakers: Evidence from an event related potentials study. *Science Bulletin*, 1-10. doi: 10.1007/s11434-016-1078-y

Yu, E. C., Sprenger, A. M., Thomas, R. P., & Dougherty, M. R. (2014). When decision heuristics and science collide. *Psychonomic Bulletin & Review, 21*(2), 268-282. doi: 10.3758/s13423-013-0495-z

Yu, M., Mo, C., Li, Y., & Mo, L. (2015). Distinct representations of syllables and phonemes in Chinese production: Evidence from fMRI adaptation. *Neuropsychologia, 77*, 253-259. doi: 10.1016/j.neuropsychologia.2015.08.027

Yu, M., Mo, C., & Mo, L. (2014). The role of phoneme in Mandarin Chinese production: Evidence from ERPs. *PLoS ONE, 9*(9), e106486. doi: 10.1371/journal.pone.0106486

Yudes, C., Macizo, P., & Bajo, T. (2010). Cognate effects in bilingual language comprehension tasks. *NeuroReport, 21*(7), 507-512. doi: 10.1097/WNR.0b013e328338b9e1

Zhang, J. X., Fang, Z., Du, Y., Kong, L., Zhang, Q., & Xing, Q. (2012). Centro-parietal N200: An event-related potential component specific to Chinese visual word recognition. *Chinese Science Bulletin, 57*(13), 1516-1532. doi: 10.1007/s11434-011-4932-y

Zhang, Q., Zhang, J. X., & Kong, L. (2009). An ERP study on the time course of phonological and semantic activation in Chinese word recognition. *International Journal of Psychophysiology, 73*(3), 235-245. doi: 10.1016/j.ijpsycho.2009.04.001

Zhang, S., Perfetti, C. A., & Yang, H. (1999). Whole word, frequency-general phonology in semantic processing of Chinese characters. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 25(4), 858. doi: 10.1037/0278-7393.25.4.858

Zhang, T. (2013). *Automatic translation from English to Chinese and subsequent processing of the Chinese translations.* (Unpublished Ph.D. Thesis), University of Nottingham.

Zhang, T., van Heuven, W. J. B., & Conklin, K. (2011). Fast automatic translation and morphological decomposition in Chinese-English bilinguals. *Psychological Science, 22*(10), 1237-1242. doi: 10.1177/0956797611421492

Zhao, X., & Li, P. (2009). An online database of phonological representations for Mandarin Chinese. *Behavior Research Methods, 41*(2), 575-583. doi: 10.3758/BRM.41.2.575

Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism, 13*(5), 505-524. doi: 10.1080/13670050.2010.488284

Zhao, X., & Li, P. (2013). Simulating cross-language priming with a dynamic computational model of the lexicon. *Bilingualism: Language and Cognition, 16*(2), 288-303. doi: 10.1017/S1366728912000624

Zhou, B., & Krott, A. (2016). Bilingualism enhances attentional control in non-verbal conflict tasks – evidence from ex-Gaussian analyses. *Bilingualism: Language and Cognition*. doi: 10.1017/S1366728916000869

Zhou, H., Chen, B., Yang, M., & Dunlap, S. (2010). Language nonselective access to phonological representations: Evidence from Chinese–English bilinguals. *The Quarterly Journal of Experimental Psychology, 63*(10), 2051-2066. doi: 10.1080/17470211003718705

Zhou, X., & Marslen-Wilson, W. (1999a). Phonology, orthography, and lexical semantic activation in reading Chinese. *Journal of Memory and Language, 41*, 579-606. doi: 10.1006/jmla.1999.2663

Zhou, X., & Marslen-Wilson, W. (1999b). Sublexical processing in reading Chinese characters. In J. Wang, A. Inhoff & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 37-64). NJ: Lawrence Erlbaum.

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1245. doi: 10.1037//0278-7393.26.5.1245

Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999a). Morphology, orthography, and phonology reading Chinese compound words. *Language and Cognitive Processes, 14*(5-6), 525-565. doi: 10.1080/016909699386185

Zhou, X., Shu, H., Bi, Y., & Shi, D. (1999b). Is there phonologically mediated access to lexical semantics in reading Chinese? In J. Wang, A. Inhoff & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 135-172). Mahwah, NJ: Lawrence Erlbaum.

Appendix

Appendix 1. Database based on Experiment 1 in Chapter 2

The database with English-Chinese translation norms can be downloaded from http://link.springer.com/article/10.3758%2Fs13428-016-0761-x or http://www.psychology.nottingham.ac.uk/ECTN. The database is available in two files. The first file (ECTN_all_translations.xlsx) contains 23 columns:

1. Word Number (nr)

2. The English word (word)

3. The number of correct alternative translations provided by participants (num_corr_resp)

4-23. The correct translations provided by participants and number of participants who provided each of the correct translations (e.g., trans1_count is the number for participant provided trans1)

The second file (ECTN_words_results.xlsx) contains 21 columns with the following information:

1. Word Number (nr)

2. The English word (word)

3. The Length of the English word (word_length)

4. The SUBTLEX-UK Zipf value of the English word (word_UK_Zipf)

5. The concreteness ratings of the English word (concreteness_Eng)

6. The part-of-speech for the English word (POS_Eng)

7. The dominant part-of-speech for the English word (doPOS_Eng)

8. The dominant Chinese translation for the English word (dom_trans)

9. The number of characters in the dominant Chinese translation (dom_trans_nchar)

10. The number of strokes in the dominant Chinese translation (dom_trans_stroke)

11. The SUBTLEX-CH Zipf value of the dominant Chinese translation (dom_trans_Zipf)

12. The part of speech for the dominant Chinese translation (dom_trans_POS)

13. The dominant part of speech for the dominant Chinese translation (dom_trans_doPOS)

14. Pinyin with tone of the dominant Chinese translation (dom_trans_pinyin)

15. IPA of the dominant Chinese translation (dom_trans_IPA)

16. The cognate/loan word status (dom_trans_cognate)

17. The percentage of the dominant translation (percent_dom_trans)

18. The percentage of correct translations (percent_correct)

19. The percentage of incorrect translations (percent_incorrect)

20. The percentage of omitted responses (percent_omitted)

21. The number of correct alternative translations provided by participants (num_corr_resp)

Note: The SUBTLEX-UK Zipf values, part-of-speech, and dominant part-of-speech of the English words were taken from van Heuven et al. (2014); concreteness ratings of the English words were taken from Brysbaert et al. (2014); SUBTLEX-CH Zipf values, part-of-speech and dominant part of speech of the dominant Chinese translations were taken from Cai and Brysbaert (2010); IPA of the dominant Chinese translation was taken from Zhao and Li (2009).

Appendix 2. Stimuli used in Experiment 2

	Prime				Target
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*	
1	edge	pence	exit	past	bat
2	unit	cake	noon	camel	worry
3	winter	verb	rest	mouse	thing
4	opera	format	method	lover	pigeon
5	relation	champion	biology	harmony	audience
6	machine	guitar	stress	lady	egg
7	gene	record	window	length	muscle
8	shoulder	suggestion	spring	activity	prison
9	spirit	mirror	height	sports	economy
10	manager	police	fever	candle	whale
11	seventy	climate	forest	adverb	wife
12	shop	god	news	blog	harm
13	life	triumph	editor	purpose	voice
14	tool	contribution	autumn	novel	princess
15	output	number	fridge	pocket	schoolbag
16	input	maths	kite	panda	vegetable
17	driver	forty	table	overlap	silk
18	smile	scarf	zebra	grape	danger
19	pollution	weapon	birthday	waste	roof
20	hope	habit	aunt	room	knee
21	request	key	black	member	invitation
22	clothes	chair	plane	train	hospital
23	factor	bank	grey	turkey	music
24	hero	coin	body	neck	infant
25	detective	pillow	flat	dolphin	truth
26	support	plant	park	meeting	knowledge
27	reader	gambling	basket	score	poison
28	fox	nurse	tooth	mask	pepper
29	butter	lie	shoe	north	emperor

	Prime				Target
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*	
30	structure	sister	picture	crystal	festival
31	divorce	history	document	summary	centimetre
32	tomorrow	destiny	school	effort	name
33	frequency	spelling	service	industry	poverty
34	potential	pencil	city	palace	front
35	time	event	reason	card	food
36	stone	soldier	prince	safety	experiment
37	luggage	gender	culture	content	shape
38	research	tear	student	error	colour
39	onion	oxygen	human	seat	balcony
40	stamp	humour	white	expert	game
41	theory	instance	law	amount	gift
42	origin	flag	teacher	article	penguin
43	capital	income	biscuit	welcome	arm
44	fifty	physics	corner	skiing	lunch
45	actor	tobacco	rocket	control	eye
46	report	parcel	earth	metal	violence
47	moral	director	toilet	company	arrival
48	empire	enemy	author	aspect	place
49	customer	bone	bread	analysis	story
50	chemistry	garden	excuse	screen	painter
51	memory	chance	mistake	worker	season
52	plan	opportunity	college	weekend	skill
53	value	family	goal	rule	shelf
54	spy	scissors	policy	slipper	health
55	clone	cough	summer	frog	guest
56	lawyer	travel	wing	surface	green
57	adventure	sweater	address	ancestor	trade
58	organ	expectation	theatre	mosquito	balloon
59	design	tongue	battery	banana	society
60	world	lion	rabbit	cream	market

	Prime				Target
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*	
61	fact	fashion	sock	farmer	century
62	speed	soda	secret	apple	plastic
63	future	tail	bill	seed	position
64	cell	watermelon	task	rainbow	drama
65	system	sacrifice	photo	kitchen	detail
66	clue	sir	test	ear	phenomenon
67	leaf	picnic	church	insect	night
68	meaning	doctor	weight	feeling	art
69	budget	language	strategy	balance	corn
70	band	appointment	price	exam	moon
71	wisdom	spider	concept	dentist	quality
72	nature	purple	animal	tennis	dictionary

Appendix 3. Stimuli used in Experiment 3

	Prime				Target
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*	
1	winter	verb	rest	mouse	thing
2	potential	pencil	city	palace	front
3	lawyer	travel	wing	surface	green
4	spirit	mirror	height	sports	economy
5	unit	cake	noon	camel	worry
6	relation	champion	biology	harmony	audience
7	shoulder	suggestion	spring	activity	prison
8	report	parcel	earth	metal	violence
9	theory	instance	law	amount	gift
10	shop	god	news	blog	harm
11	manager	police	fever	candle	whale
12	value	family	goal	rule	shelf
13	butter	lie	shoe	north	emperor
14	onion	oxygen	human	seat	balcony
15	hero	coin	body	neck	infant
16	empire	enemy	author	aspect	place
17	time	event	reason	card	food
18	world	lion	rabbit	cream	market
19	seventy	climate	forest	adverb	wife
20	spy	scissors	policy	slipper	health
21	clothes	chair	plane	train	hospital
22	life	triumph	editor	purpose	voice
23	band	appointment	price	exam	moon
24	capital	income	biscuit	welcome	arm
25	tool	contribution	autumn	novel	princess
26	fox	nurse	tooth	mask	pepper
27	stone	soldier	prince	safety	experiment
28	organ	expectation	theatre	mosquito	balloon
29	winter	verb	rest	mouse	thing

	Prime				Target
	+Segment +Tone	*+Segment -Tone*	*-Segment +Tone*	*-Segment -Tone*	
30	opera	format	method	lover	pigeon
31	reader	gambling	basket	score	poison
32	stamp	humour	white	expert	game
33	customer	bone	bread	analysis	story
34	tomorrow	destiny	school	effort	name
35	meaning	doctor	weight	feeling	art
36	memory	chance	mistake	worker	season
37	research	tear	student	error	colour
38	machine	guitar	stress	lady	egg
39	pollution	weapon	birthday	waste	roof
40	system	sacrifice	photo	kitchen	detail
41	cell	watermelon	task	rainbow	drama
42	frequency	spelling	service	industry	poverty
43	hope	habit	aunt	room	knee
44	input	maths	kite	panda	vegetable
45	budget	language	strategy	balance	corn
46	origin	flag	teacher	article	penguin
47	leaf	picnic	church	insect	night
48	factor	bank	grey	turkey	music
49	fact	fashion	sock	farmer	century
50	future	tail	bill	seed	position
51	actor	tobacco	rocket	control	eye
52	design	tongue	battery	banana	society
53	luggage	gender	culture	content	shape
54	adventure	sweater	address	ancestor	trade
55	fifty	physics	corner	skiing	lunch
56	clone	cough	summer	frog	guest
57	structure	sister	picture	crystal	festival
58	speed	soda	secret	apple	plastic
59	moral	director	toilet	company	arrival
60	driver	forty	table	overlap	silk

Appendix 4. Stimuli used in Experiment 4

	+Segment +Tone		*+Segment -Tone*	
	Prime	Target	Prime	Target
1	lion	failure	soap	plane
2	pollution	roof	reply	grey
3	shop	harm	finger	income
4	egg	chance	kitchen	exit
5	unit	worry	mail	humour
6	smile	danger	tooth	stress
7	park	job	copyright	zebra
8	wife	seventy	scissors	prison
9	honey	kite	host	jewel
10	relation	audience	chair	hospital
11	winter	thing	baseball	help
12	factor	music	import	today
13	bat	edge	evidence	struggle
14	pigeon	song	cash	sir
15	economy	spirit	creation	window
16	name	tomorrow	glass	blog
17	boy	south	brother	panda
18	box	peace	muscle	guitar
19	festival	result	garden	skiing
20	hair	vote	superman	dynasty
21	success	orange	dinner	toy
22	front	potential	rainbow	property
23	tradition	captain	tail	scarf
24	fox	beard	train	activity
25	apple	bottle	comedy	habit
26	media	rose	address	enemy
27	screen	balance	lover	cancer
28	flood	red	question	culture
29	poison	reader	task	human

	+Segment +Tone		*+Segment -Tone*	
	Prime	Target	Prime	Target
30	beer	skin	excuse	structure
31	lunch	fifty	air	peacock
32	bullet	purple	religion	summary
33	ninety	alcohol	tragedy	north
34	gift	theory	orphan	bone
35	glove	capital	exchange	corner
36	nail	tissue	son	ear
37	toilet	strategy	yellow	lie
38	green	lawyer	emphasis	robbery
39	century	market	shoe	writing
40	arrival	moral	bank	engine
41	distance	sentence	bathroom	language
42	goal	wood	mirror	police
43	adverb	parent	patience	cream
44	leaf	night	wing	ruler
45	seat	author	efficiency	novel
46	story	customer	shoulder	building
47	plastic	speed	banana	necklace
48	design	society	rubbish	candle
49	corn	budget	business	god
50	health	spy	dust	meeting
51	report	violence	school	blood
52	hope	knee	shape	gender
53	infant	hero	decade	accident
54	life	voice	service	father
55	childhood	agreement	king	past
56	colour	research	reaction	crime
57	game	stamp	picnic	liquid
58	painter	chemistry	origin	balloon
59	season	plan	dolphin	fear
60	example	history	weapon	physics

	-Segment +Tone		*-Segment -Tone*	
	Prime	Target	Prime	Target
1	spring	method	classmate	sofa
2	grandson	loyalty	farm	wedding
3	aunt	cup	condition	rest
4	fever	tool	decision	perfume
5	schoolbag	angel	experiment	request
6	height	autumn	card	frog
7	street	opinion	turkey	cheque
8	black	editor	poster	metal
9	army	invitation	trip	expert
10	machine	insect	network	minute
11	pencil	biology	drama	aspect
12	worker	need	weight	sky
13	spelling	discovery	sock	industry
14	birthday	respect	writer	palace
15	weekend	manager	nature	security
16	stairs	eyebrow	experience	tongue
17	spoon	divorce	standard	river
18	existence	butter	flat	cough
19	lemon	wealth	silk	basketball
20	charity	male	singer	neck
21	milk	file	gambling	castle
22	basket	overlap	expression	time
23	colleague	sweater	morning	flag
24	harmony	wheel	coke	burning
25	luggage	bamboo	court	student
26	monkey	age	bread	ability
27	city	white	calendar	prince
28	dentist	country	war	title
29	blue	friend	summer	picture

	-Segment +Tone		*-Segment -Tone*	
	Prime	Target	Prime	Target
30	contact	revolution	signal	candy
31	woman	seafood	attitude	nation
32	female	hour	direction	jam
33	surface	effort	support	pink
34	lady	discussion	rule	penguin
35	evening	sister	heaven	friendship
36	law	rocket	science	volcano
37	seed	tear	mushroom	tie
38	exam	crystal	food	widow
39	protection	travel	beef	choice
40	pocket	fight	room	death
41	strawberry	daughter	democracy	oxygen
42	ambition	tennis	hunter	mouth
43	lip	teacher	content	pub
44	ocean	girl	vacation	courage
45	arm	feeling	film	scandal
46	champion	bean	detective	afternoon
47	maths	flour	noon	record
48	romance	skill	parcel	entrance
49	tent	forty	curtain	earth
50	difficulty	secret	input	hat
51	floor	camel	article	freedom
52	twenty	fishing	emperor	bee
53	profit	bedroom	stone	error
54	photo	cell	semester	meaning
55	suggestion	amount	grape	mask
56	guest	purpose	breakfast	soldier
57	cake	virus	tank	memory
58	victory	climate	cartoon	shooting
59	moon	professor	comparison	dream
60	number	battery	option	art

Appendix 5. Stimuli used in Experiment 5

	+Segment +Tone		*+Segment -Tone*	
	Prime	Target	Prime	Target
1	狮子	失败	肥皂	飞机
2	污染	屋顶	回复	灰色
3	商店	伤害	手指	收入
4	鸡蛋	机会	厨房	出口
5	单元	担心	邮件	幽默
6	微笑	危险	牙齿	压力
7	公园	工作	版权	斑马
8	妻子	七十	剪刀	监狱
9	蜂蜜	风筝	主人	珠宝
10	关系	观众	椅子	医院
11	冬天	东西	棒球	帮助
12	因素	音乐	进口	今天
13	蝙蝠	边缘	证据	挣扎
14	鸽子	歌曲	现金	先生
15	经济	精神	创造	窗户
16	名字	明天	玻璃	博客
17	男孩	南方	兄弟	熊猫
18	盒子	和平	肌肉	吉他
19	节日	结果	花园	滑雪
20	头发	投票	超人	朝代
21	成功	橙子	晚餐	玩具
22	前面	潜力	彩虹	财产
23	传统	船长	尾巴	围巾
24	狐狸	胡子	火车	活动
25	苹果	瓶子	喜剧	习惯
26	媒体	玫瑰	地址	敌人
27	屏幕	平衡	爱人	癌症
28	洪水	红色	问题	文化
29	毒药	读者	任务	人类

	+Segment +Tone		+Segment -Tone	
	Prime	Target	Prime	Target
30	啤酒	皮肤	借口	结构
31	午餐	五十	空气	孔雀
32	子弹	紫色	宗教	总结
33	九十	酒精	悲剧	北方
34	礼物	理论	孤儿	骨头
35	手套	首都	交换	角落
36	指甲	纸巾	儿子	耳朵
37	厕所	策略	黄色	谎言
38	绿色	律师	强调	抢劫
39	世纪	市场	鞋子	写作
40	到达	道德	银行	引擎
41	距离	句子	浴室	语言
42	目标	木头	镜子	警察
43	副词	父母	耐心	奶油
44	叶子	夜晚	翅膀	尺子
45	座位	作者	效率	小说
46	故事	顾客	肩膀	建筑
47	塑料	速度	香蕉	项链
48	设计	社会	垃圾	蜡烛
49	玉米	预算	商业	上帝
50	健康	间谍	灰尘	会议
51	报告	暴力	学校	血液
52	希望	膝盖	形状	性别
53	婴儿	英雄	十年	事故
54	生活	声音	服务	父亲
55	童年	同意	国王	过去
56	颜色	研究	反应	犯罪
57	游戏	邮票	野餐	液体
58	画家	化学	起源	气球
59	季节	计划	海豚	害怕
60	例子	历史	武器	物理

	-Segment +Tone		*-Segment -Tone*	
	Prime	Target	Prime	Target
1	春天	方法	同学	沙发
2	孙子	忠诚	农场	婚礼
3	阿姨	杯子	条件	休息
4	发烧	工具	决定	香水
5	书包	天使	实验	要求
6	高度	秋天	卡片	青蛙
7	街道	观点	火鸡	支票
8	黑色	编辑	海报	金属
9	军队	邀请	旅行	专家
10	机器	昆虫	网络	分钟
11	铅笔	生物	戏剧	方面
12	工人	需要	重量	天空
13	拼写	发现	袜子	工业
14	生日	尊重	作家	宫殿
15	周末	经理	自然	安全
16	楼梯	眉毛	经历	舌头
17	勺子	离婚	标准	河流
18	存在	黄油	公寓	咳嗽
19	柠檬	财富	丝绸	篮球
20	慈善	男性	歌手	脖子
21	牛奶	文件	赌博	城堡
22	篮子	重叠	表达	时间
23	同事	毛衣	早上	旗帜
24	和谐	轮子	可乐	燃烧
25	行李	竹子	法庭	学生
26	猴子	年龄	面包	能力
27	城市	白色	日历	王子
28	牙医	国家	战争	题目
29	蓝色	朋友	夏天	图片

	-Segment +Tone		*-Segment -Tone*	
	Prime	Target	Prime	Target
30	联系	革命	信号	糖果
31	女人	海鲜	态度	国家
32	女性	小时	方向	果酱
33	表面	努力	支持	粉色
34	女士	讨论	规则	企鹅
35	晚上	姐妹	天堂	友谊
36	法律	火箭	科学	火山
37	种子	眼泪	蘑菇	领带
38	考试	水晶	食物	寡妇
39	保护	旅行	牛肉	选择
40	口袋	打架	房间	死亡
41	草莓	女儿	民主	氧气
42	野心	网球	猎人	嘴巴
43	嘴唇	老师	内容	酒吧
44	海洋	女孩	假期	勇气
45	手臂	感觉	电影	丑闻
46	冠军	豆子	侦探	下午
47	数学	面粉	中午	记录
48	浪漫	技能	包裹	入口
49	帐篷	四十	窗帘	地球
50	困难	秘密	输入	帽子
51	地板	骆驼	文章	自由
52	二十	钓鱼	皇帝	蜜蜂
53	利润	卧室	石头	错误
54	照片	细胞	学期	意义
55	建议	数量	葡萄	面具
56	客人	目的	早餐	士兵
57	蛋糕	病毒	坦克	记忆
58	胜利	气候	卡通	射击
59	月亮	教授	比较	梦想
60	数字	电池	选择	艺术

Appendix 6. Stimuli used in Experiment 6

	Phonological & Orthographical Repetition			Phonological Repetition		
	PO control	*PO prime*	*Target*	*P control*	*P prime*	*Target*
1	author	action	luggage	host	hero	infant
2	button	banana	perfume	opera	onion	balcony
3	example	invention	fever	object	origin	penguin
4	comedy	culture	file	dinner	driver	silk
5	white	waste	romance	voice	value	shelf
6	parent	picnic	ambition	framework	frequency	poverty
7	dentist	detail	cell	toilet	theory	gift
8	technology	university	elephant	result	report	violence
9	hair	hour	novel	comment	capital	arm
10	strawberry	journalist	memory	semester	shoulder	prison
11	guitar	genius	heaven	authority	adventure	trade
12	orange	opinion	audience	apple	actor	eye
13	exhibition	experience	manager	date	unit	worry
14	tear	verb	animal	morning	meaning	art
15	bacon	angle	corner	summer	spirit	economy
16	address	ability	energy	woman	world	market
17	shape	score	minute	farmer	factor	music
18	patient	pumpkin	south	smile	stamp	game
19	cancer	career	fact	trousers	tomorrow	name
20	funeral	fashion	time	wealth	winter	thing
21	picture	program	necklace	nation	nature	dictionary
22	temper	truth	vacuum	professor	vegetable	input
23	bottle	bread	mask	basket	budget	corn
24	task	tool	salary	brush	butter	emperor
25	singer	signal	envelope	goal	hope	knee
26	habit	hobby	lover	soap	shop	harm
27	curtain	climate	balloon	selection	structure	festival
28	tooth	train	rocket	chair	clone	guest

	Phonological & Orthographical Repetition			Phonological Repetition		
	PO control	*PO prime*	*Target*	*P control*	*P prime*	*Target*
29	pressure	passport	nurse	flood	fifty	lunch
30	help	hell	map	beef	band	moon
31	monkey	method	direction	channel	plastic	speed
32	shoe	sofa	beach	religion	research	colour
33	mirror	member	success	son	spy	health
34	respect	reality	cash	student	seventy	wife
35	king	kite	risk	forest	future	position
36	food	flat	park	length	lawyer	green
37	purpose	princess	company	detective	potential	front
38	brother	battery	film	contract	customer	story
39	mixture	machine	chance	lion	leaf	night
40	biscuit	biology	life	expert	empire	place

Appendix 7. Stimuli used in Experiment 7

	Hidden One-character Prime			Hidden Two-character Prime		
	Control	*Repetition*	*Target*	*Control*	*Repetition*	*Target*
1	book	open	switch	author	action	luggage
2	cancer	bright	understand	option	object	physics
3	boy	big	size	bottle	bread	mask
4	city	send	happen	plant	picnic	ambition
5	game	fire	train	curtain	climate	balloon
6	hard	pass	tradition	guitar	genius	heaven
7	well	look	audience	shoe	sofa	beach
8	force	carry	sports	bacon	angle	corner
9	friend	father	parent	food	flat	park
10	check	cross	process	funeral	fashion	time
11	life	hand	operation	police	poster	dolphin
12	blow	wave	waste	death	doubt	pregnancy
13	colour	square	aspect	task	tool	salary
14	burn	star	week	dentist	detail	cell
15	compare	protect	nurse	corn	care	relation
16	door	idea	opinion	title	thief	novel
17	gain	left	around	biscuit	biology	birthday
18	passport	electric	phone	king	kite	risk
19	good	year	young	shape	score	minute
20	collect	combine	contract	patient	pumpkin	south
21	version	deliver	traffic	picture	program	necklace
22	gate	root	ability	bill	base	gene
23	district	electric	computer	comedy	culture	file
24	man	now	cash	singer	signal	envelope
25	blood	heavy	important	tooth	turkey	rocket
26	family	number	maths	help	hell	map
27	know	make	homework	exhibition	experience	manager
28	choice	square	policy	mirror	member	success

	Hidden One-character Prime			Hidden Two-character Prime		
	Control	*Repetition*	*Target*	*Control*	*Repetition*	*Target*
29	income	single	unit	purpose	princess	company
30	moment	reason	story	candle	career	fact
31	large	group	organize	temper	truth	vacuum
32	dollar	divide	distribute	button	banana	perfume
33	adult	horse	careless	example	invention	fever
34	night	water	level	habit	hobby	lover
35	day	old	boss	technology	university	elephant
36	people	little	careful	strawberry	journalist	memory
37	sound	watch	show	monkey	method	direction
38	baby	full	satisfy	tear	verb	animal
39	alarm	clock	hour	brother	battery	film
40	view	land	earth	mixture	machine	chance

Appendix 8. Stimuli used in Experiment 8 Task 1

	First Character			Second Character		
	Control	*Repetition*	*Target*	*Control*	*Repetition*	*Target*
1	book	open	switch	power	close	switch
2	cancer	bright	understand	order	white	understand
3	boy	big	size	three	small	size
4	city	send	happen	area	grow	happen
5	game	fire	train	war	car	train
6	hard	pass	tradition	clear	unite	tradition
7	well	look	audience	matter	public	audience
8	force	carry	sports	case	move	sports
9	friend	father	parent	enough	mother	parent
10	check	cross	process	visitor	journey	process
11	life	hand	operation	peace	skill	operation
12	blow	wave	waste	toy	fee	waste
13	colour	square	aspect	cent	face	aspect
14	burn	star	week	bird	term	week
15	compare	protect	nurse	holiday	soldier	nurse
16	door	idea	opinion	get	see	opinion
17	gain	left	around	woman	right	around
18	passport	electric	phone	hold	word	phone
19	good	year	young	money	light	young
20	collect	combine	contract	school	around	contract
21	version	deliver	traffic	because	through	traffic
22	gate	root	ability	fall	lead	ability
23	district	electric	computer	bread	brain	computer
24	man	now	cash	file	cash	cash
25	blood	heavy	important	work	want	important
26	family	number	maths	month	learn	maths
27	know	make	homework	building	industry	homework
28	choice	square	policy	sister	needle	policy
29	income	single	unit	pressure	location	unit

	First Character			Second Character		
	Control	*Repetition*		*Control*	*Repetition*	
30	moment	reason	story	child	thing	story
31	large	group	organize	spoon	weave	organize
32	dollar	divide	distribute	plane	cloth	distribute
33	adult	horse	careless	diary	tiger	careless
34	night	water	level	rock	flat	level
35	day	old	boss	dream	board	boss
36	people	little	careful	dress	heart	careful
37	sound	watch	show	man	now	show
38	baby	full	satisfy	past	foot	satisfy
39	alarm	clock	hour	hero	head	hour
40	view	land	earth	warm	ball	earth

Appendix 9. Stimuli used in Experiment 8 Task 2 and Task 3

	Prime			Target
	Translation Equivalent	*Character Repetition*	*Control*	
1	海豚	海报	包裹	dolphin
2	工资	工具	任务	salary
3	同事	同意	监狱	colleague
4	火箭	火鸡	牙齿	rocket
5	大象	大使	技术	elephant
6	频道	频率	健康	channel
7	口袋	口音	狮子	pocket
8	平衡	平行	学生	balance
9	成功	成员	律师	success
10	分钟	分数	形状	minute
11	电影	电池	兄弟	film
12	项链	项目	图片	necklace
13	香水	香蕉	按钮	perfume
14	爱人	爱好	习惯	lover
15	气球	气氛	窗帘	balloon
16	关系	关心	玉米	relation
17	结构	结果	歌剧	structure
18	地图	地狱	帮助	map
19	记忆	记者	草莓	memory
20	生日	生物	饼干	birthday
21	动物	动词	眼泪	animal
22	角落	角度	文化	corner
23	机会	机器	混合	chance
24	怀孕	怀疑	死亡	pregnancy
25	发烧	发明	例子	fever
26	信封	信号	歌手	envelope

	Prime			Target
	Translation Equivalent	*Character Repetition*	*Control*	
27	信封	信号	歌手	envelope
28	野心	野餐	椅子	ambition
29	原则	原子	客人	principle
30	教授	教堂	苹果	professor
31	意义	意识	国家	meaning
32	天堂	天才	手套	heaven
33	细胞	细节	牙医	cell
34	自然	自由	邮票	nature
35	风险	风筝	国王	risk
36	时间	时尚	葬礼	time

Appendix 10. Results of mixed-effects modelling analysis of RTs for Experiment 6-8

Experiment 6

Orthographic & Phonological Repetition			
	Estimate (SE)	***t value***	***p***
Fixed effects			
Intercept	2.955e+00 (1.896e-02)	155.871	<0.001
Priming	5.171e-03 (1.896e-02)	0.633	0.532
LexTALE	-4.590e-03 (1.813e-03)	-2.323	0.017
RT of previous trails	1.703e-05 (7.364e-06)	2.313	0.021
Item sequences	1.903e-04 (6.387e-05)	-2.979	0.003
Random effects	**Variance**		**SD**
Items ((Intercept)	0.005		0.072
Subjects (Intercept)	0.007		0.083
Subjects (Priming)	0.001		0.033
Residuals	0.009		0.095

Phonological Repetition			
	Estimate (SE)	***t value***	***p***
Fixed effects			
Intercept	2.940e+00 (1.764e-02)	166.656	<0.001
Priming	5.599e-04 (5.381e-03)	0.104	0.917
LexTALE	-4.643e-03(1.685e-03)	-2.756	0.009
RT of previous trails	2.080e-05 (7.009e-06)	2.968	0.003
Random effects	**Variance**		**SD**
Items ((Intercept)	0.005		0.071
Subjects (Intercept)	0.005		0.073
Residuals	0.009		0.093

Hidden One-character Prime

	Estimate (SE)	*t value*	*p*
Fixed effects			
Intercept	2.984e+00 (2.199e-02)	135.705	<0.001
Priming	1.039e-02 (9.136e-03)	1.137	0.256
Prime duration	9.294e-03 (9.131e-03)	1.018	0.309
Priming*Prime duration	-1.488e-02 (1.288e-02)	-1.155	0.248
RT of previous trails	1.128e-05 (5.738e-06)	1.966	0.0495
Random effects	**Variance**		**SD**
Items ((Intercept)	0.009		0.095
Subjects (Intercept)	0.007		0.083
Residuals	0.012		0.109

Hidden Two-character Prime

	Estimate (SE)	*t value*	*p*
Fixed effects			
Intercept	3.014e+00 (2.078e-02)	145.066	<0.001
Priming	-4.283e-03 (9.234e-03)	-0.464	0.643
Prime duration	-7.561e-03 (9.175e-03)	-0.824	0.410
Priming*Prime duration	-3.723e-03 (1.302e-02)	-0.286	0.775
RT of previous trails	1.823e-05 (5.577e-06)	3.270	0.0011
Random effects	**Variance**		**SD**
Items ((Intercept)	0.008		0.089
Subjects (Intercept)	0.006		0.078
Residuals	0.012		0.108

Experiment 8

Task 1

	Estimate (SE)	*t value*	*p*
Fixed effects			
Intercept	2.953e+00 (1.674e-02)	176.400	<0.001
Priming	1.593e-05 (6.273e-03)	0.003	0.998
Position	-5.457e-03 (6.287e-03)	-0.868	0.386
Priming*Position	6.915e-03 (8.881e-03)	0.779	0.436
RT of previous trails	3.105e-05 (5.900e-06)	5.263	<0.001
Random effects	**Variance**		**SD**
Items ((Intercept)	0.005		0.074
Subjects (Intercept)	0.006		0.079
Residuals	0.010		0.098

Task 2

	Estimate (SE)	*t value*	*p*
Fixed effects			
Intercept	3.011e+00 (1.582e-02)	190.350	<0.001
Priming: Translation	-1.145e-01 (9.959e-03)	-11.502	<0.001
Priming: Repetition	-2.223e-02 (6.965e-03)	-3.192	0.002
RT of previous trails	1.979e-05 (6.334e-06)	3.124	0.002
Random effects	**Variance**		**SD**
Items ((Intercept)	0.004		0.069
Subjects (Intercept)	0.005		0.071
Subjects (Priming: Translation)	0.003		0.056
Subjects (Priming: Repetition)	0.0005		0.021
Residuals	0.010		0.101

Task 3

	Estimate (SE)	*t value*	*p*
Fixed effects			
Intercept	2.840e+00 (1.269e-02)	223.775	<0.001
Priming: Repetition	4.501e-02 (6.074e-03)	7.409	<0.001
RT of previous trails	8.829e-05 (1.331e-05)	6.635	<0.001
Random effects	**Variance**		**SD**
Items ((Intercept)	0.002		0.039
Subjects (Intercept)	0.005		0.072
Residuals	0.011		0.104

Appendix 11. Results of additional analysis for Chapter 6

We also explored the impact of script in a subset of masked experiments with a prime duration of up to 60 ms and an ISI of up to 50 ms. This analysis included 23 experiments in L1-L2 direction (17 different-script studies and 6 same-script studies) and 20 experiments in L2-L1 direction (13 different-script studies and 7 same-script studies). For both directions, script type was not a significant moderator, *p*s > .80. The results of random-effects model conducted to estimate the effect sizes of L1-L2 and L2-L1 translation priming in same and different scripts are shown in Figure S1.

Figure S1 Overall effect sizes for L1-L2 and L2-L1 translation priming involving different script and same script in healily masked experiements (with their 95% confidence intervals)

