Processing of L2 Words in Bilingual Children and Adults: Predictors, Patterns, and Tendencies

TING ZHAO

WOLFSON COLLEGE



Department of Education

University of Oxford

July 2015

THIS THESIS IS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY. THIS THESIS IS ENTIRELY MY OWN WORK, AND EXCEPT WHERE OTHERWISE INDICATED, DESCRIBES MY OWN RESEARCH.

Table of Contents

Table of Contents I
Abstract VII
Acknowledgements VIII
List of Figures IX
List of Tables XII
List of Abbreviations XVI
Chapter 1 Introduction 1
1.1 Background 1
1.2 Position in the literature 8
1.2.1 Under-researched population: Child L2 learners 8
1.2.2 Under-researched or under-estimated influences: Lexical and learner variables 10
1.2.3 Processing patterns and tendencies 13
1.2.4 Methodological inadequacy 14
1.3 Purposes and significance of the present study 17
1.4 Structure of the thesis 21
Chapter 2 Literature Review 23
2.1 Defining key terms 23
2.1.1 Bilingual 23
2.1.2 Mental lexicon 24
2.1.3 Lexical processing 25
2.2 Lexical processing in memories 26
2.2.1 Memory system and semantic memories 26
2.2.2 Bilingual lexicons and lexical processing: Structures, mechanisms, and models 29
2.2.2.1 First phase: Pre-RHM 30
2.2.2.2 Second Phase: RHM 33
2.2.2.3 Third Phase: Post-RHM 43
2.2.3 Interim summary 53
2.3 Lexical processing in spoken production 54
2.3.1 Monolingual speaker's production of single words: Components, stages, and models 55
2.3.1.1 Components and stages 55
2.3.1.2 Models 59
2.3.2. Bilingual speakers' production of single words 63
2.3.3 Interim summary 68
2.4 What is in a word? 69
2.4.1 Meaning level 69
2.4.1.1 Word concreteness 69
2.4.1.2 Word typicality 74
2.4.1.3 Semantic neighborhood density 79

2.4.1.4 Number of related senses 84
2.4.2 Form level 88
2.4.3 Usage level 94
2.4.3.1 Subjective word frequency 94
2.4.3.2 Age of acquisition 99
2.4.4 Interim Summary 104
2.5 Lexical Processing in Child and Adult L2 Learners 104
2.5.1 Memory system 105
2.5.2 L1 entrenchment 106
2.5.3 Acquisition process 108
2.6 Summary: Chapter 2 110
Chapter 3 Research Methodology 112
3.1 Methodological rationale 112
3.2 Research methods 114
3.2.1 RL-based experiments 115
3.2.1.1 Picture naming 116
3.2.1.2 Forward translation 119
3.2.2 Questionnaires and assessments 121
3.3 Research questions 122
3.4 Research design 126
3.5 Population and sample 131
3.5.1 Population 131
3.5.2 Sample, sample size, and sampling strategies used 133
3.5.2.1 Sample and sample size 133
3.5.2.2 Sampling strategies used 134
3.6 Research procedures 143
3.6.1 Pilot study phase 145
3.6.2 Preparation study phase 147
3.6.2.1 Sub-study 1: Encoding of pictures and characters 148
3.6.2.2 Sub-study 2: Stimulus pictures 150
3.6.3 Main study phase 153
3.6.3.1 RL-based experiments 153
3.6.3.2 Recording RAs and RLs 157
3.6.3.3 Participant-rated questionnaires 162
3.6.3.4 Databases 164
3.6.3.5 Assessments, and demographic and language background questionnaires 165
3.7 Analytical strategies used 176
3.7.1 Basic analysis strategies 176
3.7.1.1 Independent-measures ANOVA 176
3.7.1.2 Repeated-measures ANOVA and Mixed-model ANOVA 178
3.7.1.3 Multiple regression 180
3.7.2 Structural equation modeling 183
3.8 Ethical considerations 193

3.9 Validity, reliability, and generalizability 195
3.9.1 Validity 195
3.9.2 Reliability 198
3.9.3 Generalizability 199
3.10 Summary: Chapter 3 200
Chapter 4 Data Analyses and Results: Child Participants 203
4.1 Data Screening 205
4.2 Characteristics of participants and stimulus words 207
4.2.1 Sample characteristics 207
4.2.2 Characteristics of stimulus words 211
4.2.3 Relationships between meaning-related variables, form, and usage of L2 words: Child L2 learners 214
4.2.3.1 Hypothesized SEM model 214
4.2.3.2 Measurement model 218
4.2.3.3 Re-specified path model 220
4.2.3.4 Interim summary: Relationships between lexical variables 221
4.3 The predictive effects of learner- and word feature-related variables on L2 lexical processing among child L2 learners 222
4.3.1 Predictive effects of learner-related variables on L2 lexical processing among child L2 learners 222
4.3.1.1 Results: Predictive effects of learner-related variables on L2 lexical processing 226
4.3.1.2 Interim summary: Learner-related variables and L2 lexical processing 227
4.3.2 Predictive effects of word feature-related variables on L2 lexical processing among child L2 learners 230
4.3.2.1 Model specification 230
4.3.2.2 Evaluation of assumptions for the specified path models 232
4.3.2.3 Path model estimation and results 232
4.3.2.4 Interim summary: Predictive effects of lexical variables on processing efficiency 237
4.4 General pattern of L2 Lexical processing among child L2 learners 238
4.4.1 Speed and accuracy of encoding pictures and Chinese characters 239
4.4.1.1 Speed of encoding 239
4.4.1.2 Accuracy of encoding 241
4.4.2 General pattern of L2 lexical processing among child L2 learners 242
4.4.2.1 Response latencies 243
4.4.2.2 Response accuracies 248
4.4.2.3 Interim summary: General pattern of children's L2 lexical processing 252
4.5 Tendency toward a certain pattern of L2 lexical processing among child L2 learners 253
4.5.1 Learner-related variables and processing tendency 254
4.5.2 Word feature-related variables and processing tendency 257

4.5.3 Interim summary: Factors predicting processing tendency 259
4.6 Summary: Chapter 4 260
Chapter 5 Data Analysis and Results: Adult Participants 263
5.1 Data screening 264
5.2 Characteristics of participants and stimulus words 265
5.2.1 Sample characteristics 265
5.2.2 Characteristics of stimulus words 271
5.2.3 Relationships between meaning-related variables, Form level and Usage level of L2 words: Adult L2 learners 274
5.2.3.1 Hypothesized SEM model 274
5.2.3.2 Measurement model 277
5.2.3.3 Re-specified SEM model 279
5.2.3.4 Interim summary: Relationships between lexical variables 280
5.3 The predictive effects of learner- and word feature-related variables on L2 lexical processing among adult L2 learners 281
5.3.1 Predictive effects of learner-related variables on L2 lexical processing among adult L2 learners 281
5.3.1.1 Results: Predictive effects of learner-related variables on L2 lexical processing 285
5.3.1.2 Interim summary: Learner-related variables and L2 lexical processing 286
5.3.2 Predictive effects of word feature-related variables on L2 lexical processing among adult L2 learners 289
5.3.2.1 Model specification 289
5.3.2.2 Evaluation of assumptions for the specified SEM model 291
5.3.2.3 SEM model estimation and results 291
5.3.2.4 Interim summary: Predictive effects of lexical variables on processing efficiency 296
5.4 General pattern of L2 Lexical processing among adult L2 learners 297
5.4.1 Speed and accuracy of encoding pictures and Chinese characters 297
5.4.2 General pattern of L2 lexical processing among non-English majors 301
5.4.2.1 Response latencies: Non-English majors 301
5.4.2.2 Response accuracies: Non-English majors 306
5.4.2.3. Interim summary: General pattern of non-English majors' L2 lexical processing 310
5.4.3 General pattern of L2 lexical processing among English majors 310
5.4.3.1 Response latencies: English majors 311
5.4.3.2 Response accuracies: English majors 315
5.4.3.3 Interim summary: General pattern of English majors' L2 lexical processing 319
5.5 Tendency toward a certain pattern of L2 lexical processing among adult L2 learners 320
5.5.1 Learner-related variables and processing tendency 320

5.5.2 Word feature-related variables and processing tendency 324
5.5.3 Interim summary: Factors predicting processing tendency 326
5.6 Summary: Chapter 5 329
Chapter 6 General Discussion 332
6.1 Summary of results 332
6.2 Explaining the data: Learner and lexical characteristics 334
6.2.1 Learner effects 334
6.2.2 Lexical effects 340
6.2.2.1 Conceptual-semantic level 340
6.2.2.2 Form level 346
6.2.2.3 Usage level 347
6.2.3 Summary: Learner variables vs. lexical variables 350
6.3 Explaining the data: Mechanisms underlying L2 lexical processing 351
6.3.1 Child L2 learners 351
6.3.2 Adult L2 learners 359
6.4 Summary: Dynamics of bilingual lexical processing 367
Chapter 7 Conclusion 369
7.1 Limitations of the present study and recommendations for future research 369
7.1.1 Limitations of the present study 369
7.1.2 Recommendations for future research 373
7.2 Pedagogical implications 377
7.3 Conclusion 386
References 389
Appendices 422
Appendix 3.1 Participation information sheet for teachers 422
Appendix 3.2 Participation information sheet for students 424
Appendix 3.3 Consent form for teachers and students 426
Appendix 3.4 Participation information sheet for parents 427
Appendix 3.5 Consent form for parents 428
Appendix 3.6 Checklist of experimental procedures 429
Appendix 3.7 Rating task format: word concreteness (adult participants) 431
Appendix 3.8 Rating task format: word concreteness and subjective word frequency (child participants) 434
Appendix 3.9 Rating task format: Word typicality (adult participants) 436
Appendix 3.10 Rating task format: Word typicality (child participants) 439
Appendix 3.11 Rating task format: Subjective word frequency (adult participants) 441
Appendix 3.12 Word associates test (child participants) 444
Appendix 3.13 Word associates test (adult participants) 447
Appendix 4.1 Characteristics of stimulus words: Child participants 451
Appendix 4.2 Comparing word features across blocks: Child participants 454
Appendix 4.3 Assumptions of SEM: Do Meaning and Form predict Usage among child L2 learners? 465
Appendix 4.4 Measurement model with standardized coefficients: Meaning, Form, and Usage of L2 words among child L2 learners (IBM SPSS AMOS output) 468

Appendix 4.5 Path model with standardized coefficients: Meaning, form and usage of L2 words among child L2 learners (IBM SPSS AMOS output) 469
Appendix 4.6 Results of associations between nominal learner-related variables and the response latency of PE and that of FT among child L2 learners (n = 39) 470
Appendix 4.7 Assumptions of MR: Predictive effect of learner-related variables on PE and FT response latencies among child L2 learners (n = 39) 471
Appendix 4.8 Assumptions of SEM model: Predictive effect of word feature-related variables on PE and FT response latencies 473
Appendix 4.9 Path models with standardized coefficients: Predictive effects of word feature-related variables on PE and FT response latencies (IBM SPSS AMOS output) 475
Appendix 4.10 Results of associations between nominal learner-related variables and PF-FT response latency differences among child L2 learners (n = 39) 476
Appendix 4.11 Assumptions of MR: Predictive effects of learner-related and word feature-related variables on PE-FT RL differences among child L2 learners 477
Appendix 5.1 Characteristics of stimulus words: Adult participants 479
Appendix 5.2 Comparing word features across blocks: Adult participants 485
Appendix 5.3 Assumptions of SEM: Do Meaning and Form predict Usage among adult L2 learners? 496
Appendix 5.4 Measurement model with standardized coefficients: Meaning, Form, and Usage of L2 words among adult L2 learners (IBM SPSS AMOS output) 499
Appendix 5.5 Re-specified SEM model with standardized coefficients: Do Meaning and Form predict Usage of L2 words among adult L2 learners (IBM SPSS AMOS output) 500
Appendix 5.6 Results of associations between nominal learner-related variables and the response latency of PE and that of FT among adult L2 learners (n = 94) 501
Appendix 5.7 Assumptions of MR: Predictive effect of learner-related variables on PE and FT response latencies among adult L2 learners (n = 94) 502
Appendix 5.8 Assumptions of SEM model: Predictive effect of word feature-related variables on PE and FT response latencies among adult L2 learners 504
Appendix 5.9 SEM model with standardized coefficients: Predict effects of word feature-related variables on PE and FT response latencies among adult L2 learners (IBM SPSS AMOS output) 506
Appendix 5.10 Results of associations between nominal learner-related variables and PF-FT response latency differences among adult L2 learners (n = 94) 508
Appendix 5.11 Assumptions of MR: Predictive effect of learner-related and word feature-related variables on PE-FT RL differences among adult L2 learners 509

Abstract

Within the context of foreign language learning, very little research has examined how learners process second language (L2) words in terms of which variables best predict their processing speed and which mechanisms best characterize bilingual lexical processing. The present study set out to address this gap by using a range of learner and lexical variables (such as vocabulary size, word length, and age of acquisition) as points of reference against which to identify the best predictors of children's and adults' L2 lexical processing, and by comparing response latencies across stimulus conditions so as to identify the processing pattern specific to each age group. Thirty-nine primary-age and 94 university-level Chinese learners of English performed a picture-naming task in English and then in Chinese, and then completed a Chinese-to-English task. The researcher analyzed and estimated how those learner and lexical variables predicted the recorded response latencies by means of multiple regression and structural equation modeling, and made cross-condition comparisons with the use of analysis of variance. The results suggested that different aspects of vocabulary knowledge contributed significantly to predicting children's and adults' processing speed, and that shorter processing time was significantly and directly predicted by the younger age at which an L2 word was learned and its higher degree of word typicality. Both lexical association and conceptual mediation were present in L2 lexical processing irrespective of learners' age but in general the later an L2 word is learned, the greater the likelihood that the word is lexically accessed and processed. These results illustrate the crucial role that language experiences and conceptual structures play in influencing the ease or difficulty with which L2 lexical items are retrieved, and reflect the complexities and dynamics involved in processing bilingual lexicons. These findings will be discussed within the context of the role of research and theory in developing evidence-based pedagogical practice with a specific reference to vocabulary acquisition in children and adults learning foreign languages within input-limited contexts.

Acknowledgements

It has been an incredible experience reading all exciting research on bilingual lexical processing, conducting the fieldwork with primary-age children and university-level students, and having insightful discussions with experts in the fields examined. I have learned a tremendous amount throughout the course of writing this thesis and have been greatly enriched by encounters with a diverse pool of academic talents, all of which will have a far-reaching impact upon my career in the years to come.

The completion of this thesis would not have been possible without the invaluable support and help of many people. First of all, I would like to thank my supervisor Professor Victoria Murphy, whose intellectual inspiration and guidance have been crucial to my academic growth over the past three years. She not only aroused my research interest in language development but also taught me what real scholarship can and should be. I am also grateful to Professor Ernesto Macaro, who supervised my MSc research study and gave me chances to broaden my academic interests. In acknowledgement, a debt of gratitude goes to the many other people who guided me throughout my academic journey in Oxford, including Dr Xin Wang, Dr Lars-Erik Malmberg, Professor Pam Sammons, Dr James Hall, Julie Dearden, Dr Catherine Walter, and Professor Kathy Sylva. Additionally, many thanks go to all members of the ALSLA group and the REAL group, for their great support and advice.

Lastly, and on a personal note, I would like to thank my friends and family for their unconditional support and consistent encouragement. I am particularly indebted to my husband Zhuo for his constant support in all my endeavors.

List of Figures

Figure 2. 1 A taxonomy of long-term memory systems (adapted from Squire & Zola, 1996) 27
Figure 2. 2 Three types of possible bilingual representations 30
Figure 2. 3 Word Association Model 31
Figure 2. 4 Concept Mediation Model 31
Figure 2. 5 Logic behind Potter et al.'s experiment (Potter et al., 1984, p. 25) 32
Figure 2. 6 Revised Hierarchical Model (Adapted from Kroll & Stewart, 1994) 33
Figure 2. 7 Distributed Feature Model (adapted from De Groot, 1992b) 45
Figure 2. 8 Modified Hierarchical Model (Pavlenko, 2009, p. 147) 49
Figure 2. 9 Three-stage Vocabulary Acquisition Model (Jiang, 2004, p. 418) 51
Figure 2. 10 Levelt's model of speech production: A blueprint of the speaker (Levelt, 1999, p. 87) 62
Figure 2. 11 Bilingual spoken word production: Spread Activation Model (adapted from Hermans, 2000; Poulisse & Bongaerts, 1994) 65
Figure 2. 12 Schematic Representation of Bilingual System (Costa, 2005, p. 314) 66
Figure 2. 13 Concept selection model (adapted from La Heij, 2005) 68
Figure 2. 14 Bilingual version of the dual-coding system (Paivio & Desrochers, 1980, p. 391) 71
Figure 2. 15 The BIRD category (Aitchison, 2012) 74
Figure 2. 16 Key issues reviewed in this chapter 111

Figure 3. 1 Processing components involved in picture naming (Indefrey & Levelt, 2004, p. 109) 117
Figure 3. 2 Pathways of conceptual mediation and word association in picture naming (Potter et al., 1984, p. 25) 118
Figure 3. 3 Processing components involved in visual word recognition (Indefrey & Levelt, 2004, p. 109) 120
Figure 3. 4 Pathways of conceptual mediation and word association in forward translation (Potter et al., 1984, p. 25) 121
Figure 3. 5 Latin square design adopted in the present study 127
Figure 3.6 Sampling Strategies used in the present study: Adult participants 139
Figure 3. 7 Sampling Strategies used in the present study: Child participants 142
Figure 3. 8 Overview of research procedures 144
Figure 3. 9 Structure of a trial: Encoding pictures and Chinese characters 150
Figure 3. 10 Structure of a trial: Name agreement, and image agreement and variability 152
Figure 3. 11 Logic behind the experiment (Potter et al., 1984) 154
Figure 3. 12 Structure of a trial: Comparing picture naming and forward translation 157
Figure 3.13 Sample of an acoustic waveform for the production of 'frog' 159
Figure 3. 14 Description of the Wordsmyth entry for the word 'apple' 164

Figure 3. 15 Presentation of a lexical item in X_Lex 168
Figure 3. 16 Result of an X_Lex test 169
Figure 3. 17 Example of Matrix Reasoning test (Wechsler, 1999) 174
Figure 3. 18 The partitioning of variance for a repeated-measures experiment (Gravetter & Wallnau, 2007, p. 442) 179
Figure 3. 19 Variance in MR 181
Figure 3. 20 Exploratory vs. Confirmatory factor analyses (Blunch, 2012, p. 128) 184
Figure 3. 21 A full model: Measurement models plus structural models 185
Figure 3. 22 Flow chart of basic steps of SEM (Kline, 2011, p. 92) 188
Figure 3. 23 Domains of quality (Oancea & Furlong, 2007, p. 133) 193
Figure 3. 24 Types of validity and threats to validity (based on Ary et al., 2010; Cohen et al., 2007; Goodwin, 2009) 198
Figure 3. 25 Validity and reliability: Four possibilities 199
Figure 3. 26 Flowchart of Chapter 3 202

Figure 4. 1 Overview of research questions and methods of data analysis: Child participants 204
Figure 4. 2 Hypothesized SEM model: Do Meaning and Form predict the Usage of L2 words among child L2 learners? 215
Figure 4. 3 Measurement model: Meaning, Form and Usage of L2 words among bilingual children 218
Figure 4. 4 Re-specified path model: Form and Usage of L2 words among bilingual children 220
Figure 4. 5 Re-specified path model: Do meaning and form predict the usage of L2 words among child L2 learners? 220
Figure 4. 6 Hypothesized path models: Predictive effects of word feature-related variables on PE and FT response latencies in child L2 learners 232
Figure 4. 7 Final path model with standardized coefficients: Predictive effects of word features on PE response latencies among child L2 learners 233
Figure 4. 8 Final path model with standardized coefficients: Predictive effects of word features on FT response latencies among child L2 learners 236
Figure 4. 9 Comparisons (by participants) across stimulus conditions and groups: response latencies among child L2 learners 245
Figure 4. 10 Comparisons across stimulus conditions and blocks: Response latencies by items among child L2 learners 247
Figure 4. 11 Comparisons (by participants) across stimulus conditions and groups: Response accuracies (in percentage) among child L2 learners 249
Figure 4. 12 Comparisons across stimulus conditions and blocks: Response accuracies by items among child L2 learners 252
Figure 4. 13 Summary of Chapter 4: Research questions and results 261

Figure 5.1 Overview of research questions and methods of data analysis 263
Figure 5. 2 Distributions of sample characteristics 267

Figure 5. 3 Hypothesized SEM model: Do Meaning and Form predict the Usage of L2 words among adult L2 learners? 274
Figure 5. 4 Measurement model: Meaning, Form and Usage of L2 words among adult L2 learners 277
Figure 5. 5 Re-specified SEM model: Do Meaning and Form predict the Usage of L2 words among adult L2 learners? 279
Figure 5. 6 Final SEM model with standardized coefficients: Do Meaning and Form predict the Usage of L2 words among adult L2 learners? 280
Figure 5. 7 Hypothesized SEM model: Predictive effects of word feature-related variables on PE and FT response latencies in adult L2 learners 290
Figure 5. 8 Final SEM model with standardized coefficients: Predictive effects of word features on PE response latencies among adult L2 learners..... 292
Figure 5. 9 Final SEM model with standardized coefficients: Predictive effects of word features on FT response latencies among adult L2 learners..... 294
Figure 5. 10 Comparisons (by participants) across stimulus conditions and groups: Response latencies among non-English majors..... 303
Figure 5. 11 Comparisons across stimulus conditions and blocks: Response latencies by items among non-English majors 305
Figure 5. 12 Comparisons (by participants) across stimulus conditions and groups: Response accuracies (in percentage) among non-English majors 308
Figure 5. 13 Comparisons across stimulus conditions and blocks: Response accuracies by items among non-English majors..... 310
Figure 5. 14 Comparisons (by participants) across stimulus conditions and groups: Response latencies among non-English majors..... 313
Figure 5. 15 Comparisons across stimulus conditions and blocks: Response latencies by items among English majors..... 315
Figure 5. 16 Comparisons (by participants) across stimulus conditions and groups: Response accuracies (in percentage) among English majors 317
Figure 5. 17 Comparisons across stimulus conditions and blocks: Response accuracies by items among English majors 319
Figure 5. 18 Summary of Chapter 5: Research questions and results 330

Figure 6. 1 Structure of the discussion chapter 334
Figure 6. 2 Functional-Developmental Model of Bilingual Lexicons: Child L2 Learners 356
Figure 6. 3 Functional-Developmental Model of Bilingual Lexicons: Adult L2 Learners 365
Figure 6. 4 Dynamics of bilingual lexical processing 368

List of Tables

Table 2. 1 Overview of studies examining the RHM 35
Table 2. 2 Selected literature examining word concreteness effects 73
Table 2. 3 Selected literature examining word typicality effects 77
Table 2. 4 Selected literature examining semantic neighborhood effects 82
Table 2. 5 Selected literature examining number-of-related-senses effects 86
Table 2. 6 Summary of multiple regression studies examining the contribution of form variables to the fluency of lexical processing 91
Table 2. 7 Selected literature examining word length effects 92
Table 2. 8 Selected literature examining word frequency effects 97
Table 2. 9 Selected literature examining age-of-acquisition effects 103

Table 3. 1 Overview of research questions, data sources and analyses 125
Table 3. 2 Overview of research design: pilot, preparation and main study phases 129
Table 3. 3 Values of word feature-related variables: Ratings and databases 161
Table 3. 4 Descriptions of associates and distractors used in the revised WAT172
Table 3. 5 Measures of effect size, their uses, and cut-off criteria (based on Field, 2013; Tabachnick & Fidell, 2012) 193

Table 4. 1 Summary of valid and deleted responses 206
Table 4. 2 Demographics, and language and academic backgrounds of child participants (n = 39) 208
Table 4. 3 Pearson correlations (r values) between test scores: Child participants (n = 39) 210
Table 4. 4 Word features of each block: Child participants (n = 105) 212
Table 4. 5 Comparing stimulus words between blocks: Child participants (n = 105) 213
Table 4. 6 Correlation matrix of word feature-related variables: Child participants (n = 105) 217
Table 4. 7 Correlations between learner-related variables, and PE and FT performance in child L2 learners (n = 39) 224
Table 4. 8 Standard MR: Predictive effects of learner-related variables on PE response latencies among child L2 learners (n = 39) 228
Table 4. 9 Standard MR: Predictive effects of learner-related variables on FT response latencies among child L2 learners (n = 39) 229
Table 4. 10 Correlation matrix of lexical variables and child participants' response latencies (n = 104) 231
Table 4. 11 Decomposition of standardized effects for path diagrams of Figures 4.7 235
Table 4. 12 Decomposition of standardized effects for path diagrams of Figures 4.8 237
Table 4. 13 Descriptive statistics of encoding speed (in ms) by child participants

between conditions 239
Table 4. 14 Descriptive statistics of encoding speed (in ms) by items between conditions: Child participants 240
Table 4. 15 Descriptive statistics of accuracy of encoding (in percentage) by child participants between conditions 241
Table 4. 16 Descriptive statistics of accuracy of encoding (in percentage) by items between conditions: Child participants 241
Table 4. 17 Descriptive statistics of response latencies (in ms) by participants across stimulus conditions in each group: Child L2 learners 244
Table 4. 18 Post-hoc pairwise comparisons between stimulus conditions in each group (n = 39): Response latencies (in ms) among child L2 learners 245
Table 4. 19 Descriptive statistics of response latencies (in ms) by items across stimulus conditions in each block: Child L2 learners 246
Table 4. 20 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 105): Response latencies (in ms) by items among child L2 learners 247
Table 4. 21 Descriptive statistics of response accuracies (in percentage) by participants across stimulus conditions in each group: Child L2 learners 248
Table 4. 22 Pairwise comparisons between stimulus conditions in each group (n = 39): Response accuracies (in percentage) among child L2 learners 249
Table 4. 23 Descriptive statistics of response accuracies (in percentage) by items across stimulus conditions in each block: Child L2 learners 250
Table 4. 24 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 105): Response accuracies by items among child L2 learners. 251
Table 4. 25 Correlation matrices: Relationships between learner-related variables, and PE-FT response latency differences among child L2 learners (n = 39) 255
Table 4. 26 Standard multiple regression of Chinese course scores on the PE-FT response latency difference (n = 39) 256
Table 4. 27 Correlation matrix: Correlations between word feature-related variables, and PE-FT response latency differences among child L2 learners (n = 105) 258
Table 4. 28 Standard multiple regression of age of acquisition on the PE-FT response latency difference among child L2 learners (n = 105) 259

Table 5. 1 Summary of valid and deleted responses 265
Table 5. 2 Distributions of sample characteristics: Adult participants (n = 94) 266
Table 5. 3 Descriptive statistics of test scores among non-English majors and among English-majors 270
Table 5. 4 Pearson correlations (r values) between test scores: Adult participants (n = 94) 270
Table 5. 5 Summary statistics: Lexical features of each block (n = 198) 272
Table 5. 6 Comparing stimulus words between blocks: ANOVA results (n = 198) 273

Table 5. 7 Correlation matrix of word feature-related variables: Adult participants (n = 198)..... 276
Table 5. 8 Correlations between learner-related variables, and PE and FT performance in adult L2 learners (n = 94)..... 283
Table 5. 9 Standard MR: Predictive effects of learner-related variables on PE response latencies among adult L2 learners (n = 94) 287
Table 5. 10 Standard MR: Predictive effects of learner-related variables on FT response latencies among adult L2 learners (n = 94) 288
Table 5. 11 Decomposition of standardized effects for structural models of Figure 5.8..... 293
Table 5. 12 Decomposition of standardized effects for structural models of Figure 5.9..... 296
Table 5. 13 Descriptive statistics of encoding speed (in ms) by adult participants between conditions 298
Table 5. 14 Descriptive statistics of encoding speed (in ms) by items between conditions: Adult participants..... 299
Table 5. 15 Descriptive statistics of accuracy of encoding (in percentage) by adult participants between conditions..... 299
Table 5. 16 Descriptive statistics of accuracy of encoding (in percentage) by items between conditions: Adult participants..... 300
Table 5. 17 Descriptive statistics of response latencies (in ms) by participants across stimulus conditions in each group: Non-English majors..... 302
Table 5. 18 Post-hoc pairwise comparisons between stimulus conditions in each group (n = 60): Response latencies (in ms) among non-English majors... 303
Table 5. 19 Descriptive statistics of response latencies (in ms) by items across stimulus conditions in each block: Non-English majors 304
Table 5. 20 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response latencies (in ms) by items among non-English majors 305
Table 5. 21 Descriptive statistics of response accuracies (in percentage) by participants across stimulus conditions in each group: Non-English majors 306
Table 5. 22 Pairwise comparisons between stimulus conditions in each group (n = 60): Response accuracies (in percentage) among non-English majors .. 307
Table 5. 23 Descriptive statistics of response accuracies (in percentage) by items across stimulus conditions in each block: Non-English majors 308
Table 5. 24 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response accuracies by items among non-English majors 309
Table 5. 25 Descriptive statistics of response latencies (in ms) by participants across stimulus conditions: English majors..... 311
Table 5. 26 Post-hoc pairwise comparisons between stimulus conditions in each group (n = 34): Response latencies (in ms) among English majors 312
Table 5. 27 Descriptive statistics of response latencies (in ms) by items across

stimulus conditions in each block: English majors 313
Table 5. 28 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response latencies (in ms) by items among English majors 314
Table 5. 29 Descriptive statistics of response accuracies (in percentage) by participants across stimulus conditions in each group: English majors 315
Table 5. 30 Pairwise comparisons between stimulus conditions in each group (n = 34): Response accuracies (in percentage) among English majors 316
Table 5. 31 Descriptive statistics of response accuracies (in percentage) by items across stimulus conditions in each block: English majors 317
Table 5. 32 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response accuracies by items among English majors.... 318
Table 5. 33 Correlation matrices: Relationships between learner-related variables, and PE-FT response latency differences among adult L2 learners (n = 94) 322
Table 5. 34 Standard multiple regression of academic achievement on the PE-FT response latency difference (n = 94)..... 323
Table 5. 35 Correlation matrix: Correlations between word feature-related variables, and PE-FT response latency differences among adult L2 learners (n = 198) 325
Table 5. 36 Standard multiple regression of word feature-related variables on the PE-FT response latency difference among adult L2 learners (n = 198).... 328

Table 6. 1 Participant and stimulus information about studies using Potter et al.'s experimental paradigm 361

List of Abbreviations

AoA	Age of Acquisition
AEV	Average Extracted Variances
AMOS	Analysis of MOment Structures
ANOVA	Analysis of Variance
BT	Backward Translation
CET	College English Test
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CMM	Concept Mediation Model
CR	Construct Reliability
CUREC	Central University Research Ethics Committee
DMF	Distributed Feature Model
DV	Dependent Variable
EFA	Exploratory Factor Analysis
EFL	English as a Foreign Language
ESL	English as a Second Language
FT	Forward Translation
IV	Independent Variable
L1	First Language
L2	Second Language
MHM	Modified Hierarchical Model
MR	Multiple Regression
MS	Million Second
NoL	Number of Letters
NoP	Number of Phonemes
NoRS	Number of Related Senses
NoS	Number of Syllables
NoSN	Number of Semantic Neighbors
PC	Picture Naming in Chinese

PE	Picture Naming in English
PN	Picture Naming
PND	Phonological Neighborhood Density
RA	Response Accuracy
RHM	Revised Hierarchical Model
RL	Response Latency
RMSEA	Root Mean Square Error of Approximation
RT	Reaction Time
SC	Semantic Categories
SEM	Structural Equation Modeling
SIC	Interconstruct Correlations
SWF	Subjective Word Frequency
VIF	Variance Inflation Factor
VLT	Vocabulary Levels Test
WAM	Word Association Model
WASI	Wechsler Abbreviated Scale of Intelligence
WAT	Word Associates Test
WC	Word Concreteness
WI	Word Imageability
WL	Word Length
WT	Word Typicality
YLE	Cambridge Young Learners English
MS	Million Seconds
SD	Standard Deviations
MD	Mean Difference
SE	Standard Error
OWF	Objective Word Frequency
ICC	Intraclass Correlation Coefficient

Chapter 1 Introduction

This chapter first sets up the background for the present research. Next, research purposes are specified, and the significance of this study is discussed. After that, the structure of the whole thesis is presented.

1.1 Background

Human language is a rule-governed system that we know and use to conduct dialogical interactions, perform social acts, express individual identities, and understand the world in which we live. All these purposes are achieved mainly through our comprehension and production of languages, two types of everyday cognitive activities that are underpinned by distinct but interrelated processing mechanisms. The present study focused on language production, in part because many native Chinese learners have particular difficulty with speaking the second language (English), as evidenced by a much lower IELTS (International English Language Testing System) score of speaking (on average, 5.3) compared to those of reading (on average, 5.9) and listening (on average, 5.7) based on 2010 test data (IELTS annual review, 2010).

One goal of second language (L2) learning is to gain fluency in speech production so as to achieve smoothness in oral communication with others. Compared to word production in the first language (L1), producing L2 utterances requires more processing effort, which is caused by speakers' incomplete L2 knowledge in terms of

lexical, syntactic, morphological and phonological respects, and high cognitive demands needed for coping with L1 activation (Kormos, 2006). Not surprisingly, in most if not all circumstances, the speech rate of the L2 has been found to be much slower than that of the L1 (Raupach, 1984). In order to advance our knowledge of the psycholinguistic process underlying L2 spoken production, it is necessary to take participants and individual lexical items as units of analysis against which to examine the ease with which L2 words are accessed and processed, as these factors are widely deemed to be crucial to determining how fluent a speech stream can be delivered.

Language production can be examined simply through spoken naming, which is 'a natural corollary of referring' (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979, p. 337) fundamental to human language use (Terrace, 1985). Beyond that, naming is one of the first linguistic functions mastered by young children, enabling them to build up a common vocabulary base for simple communications between parents and children (Bates et al., 1979). Research into how people retrieve and name objects has a long tradition, which dates back to the 19th century work by Cattell (1886). Picture naming is one of the commonly used tasks to decompose operations in language production and to sketch the architecture of production processes (e.g. Glaser, 1992; Levelt et al., 1991; Potter, So, Eckardt, & Feldman, 1984). Pictures are thought of as symbols to represent the referents they stand for but also these visual representations approximate individuals' imagistic mental representations. In this regard, the process of analyzing pictures is regarded as similar to what occurs in the

initial stage of natural spoken production—mental conceptualization.

Another way to study language production is through translation, which has been more recently used relative to picture naming (Jiang, 2012). Any individuals who are competent of two languages, both receptively and productively, are usually able to perform forward and backward translation. Previous research studies involving the use of translation have mainly addressed what occurs in bilingual spoken production by specifying the structure of bilingual lexicons and by working out the mechanisms underlying L2 lexical processing (e.g. De Groot, 1992b; De Groot, Dannenburg, & Vanhell, 1994; Kroll & Stewart, 1994). In addition, translation has long been used in L2 classrooms to assist instructors in delivering teaching content and assessing language learners' L2 performance. As such, researching translation likely provides insight into how a common instructional activity proceeds at individual levels within formal educational contexts.

Monolingual speech production proceeds through conceptualization, grammatical encoding, articulation, and self-monitoring (Levelt, 1989), a process to which bilingual speech production should be similar if L2 speakers were able to switch off the L1 from the very start. A consistent body of research has shown that the two languages of a bilingual are active to different extents (e.g. Hoshino & Kroll, 2008; Kroll, Bobb, & Wodniecka, 2006; Moon & Jiang, 2012), and the degree to which each language is active varies along a continuum between two extremes: one is *concept mediation* (i.e. the presence of a direct link between an L2 word and its

concept) and the other is *word association* (i.e. the presence of a direct link between an L2 word and its L1 equivalent), largely depending on a bilingual's experiences, task demands, and word types. Over the past few decades, much effort has been devoted to addressing the complexities and dynamics involved in L2 lexical processing by modeling the mechanisms underlying L2 lexical processing.

Drawing on Kuhn's theory (Kuhn, 1962), the development of these models can be divided into three stages: (1) a '*pre-paradigm*' stage in which attempts were made to explore all possible structures of bilingual lexicons (e.g. the models proposed by Weinreich, 1953) but no consensus has been reached on any particular models, (2) a '*normal science*' stage in which a highly influential model—*Revised Hierarchical Model* (RHM, Kroll & Stewart, 1994)—was proposed, having inspired a large body of empirical research to assess its validity and generalizability, and (3) a *transition* stage between 'normal science' and 'revolutionary science' in which continuous efforts are made to accommodate the 'anomalies' that are beyond the explanatory reach of the RHM and simultaneously to put forward new models with a higher problem-solving ability in certain respects. In order to ensure that future models would be more generalizable to different types of bilingual learners and words, it is deemed necessary to closely examine two variables: one is lexical characteristics and the other is learners' age.

We may probably all know that every human being is blessed with a set of unique qualities, which make each of us different. To a certain extent this is also the case

with words. All words have their own physical forms, phonological elements, syntactic functions, semantic properties, and pragmatic implications. A lexical entry typically consists of a lemma (i.e. semantic and syntactic information) and a lexeme (i.e. the morphological, phonological, and orthographical information) (Levelt, 1989), both of which serve to set words apart as a distinct item in the lexical repertoire but also have some impact upon our comprehension and production activities. Given that words can be described and defined multi-dimensionally, one approach to understand the process of spoken production is to vary the lexical characteristics that stimulus words have so as to identify which characteristic have the most impact upon a speaker's choice of a particular word and the time taken to produce that word.

Among a variety of lexical characteristics, of particular interest to the present study are conceptual-semantic variables, word forms, and word usage. To be specific, the conceptual-semantic variables include: (1) word typicality, which refers to how well a lexical item represents its category, (2) word concreteness through which words can be labeled as concrete or abstract, (3) semantic neighborhood density, which is measured by how many words occur within a specified radius, and (4) the number of related senses, which is defined by the number of related lexical entries that a target word has. At the form level, there are three variables under current examination, namely: the number of letters, the number of phonemes, and the number of syllables. Together with the above lexical variables, two usage-related ones—subjective word frequency and age of acquisition—were included as well.

These variables were selected because most of them correspond to different stages of spoken word production (e.g. conceptualization, formulation, and articulation). Moreover, these variables have been found to influence the fluency of L1 and/or L2 lexical processing in previous studies, as will be reviewed in Chapter 2 shortly (e.g. Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; D'Amico, Devescovi, & Bates, 2001; De Groot, 1992b; De Groot, Borgwaldt, Bos, & Van den Eijnden, 2002; Greenberg & Bjorklund, 1981; Hino & Lupker, 1996).

In addition to examining the roles that word features likely play in L2 lexical processing, age is another factor that deserves our careful consideration. In the present study, two types of successive bilinguals, child and adult L2 learners, are the populations of interest. One of the striking differences between these two populations is that adult L2 learners have had more L1 experience compared to child L2 learners. In addition, adult learners are more inclined to fit new information into existing models, whereas children are more inclined to restructure existing schemes through accommodation. Given this, these two age cohorts may rely on different mechanisms to process L2 words, and in effect, the structure of their bilingual lexicons may not be the same.

Other age-related differences concern the process of L2 acquisition, which are socio-culturally and psychologically shaped. Both populations of interest learn English as a compulsory subject within instructional contexts in a non-native English speaking country, but, comparatively speaking, children are more intrinsically

motivated to learn the L2, whereas adults are more instrumentally stimulated to learn the L2 (Zhou, Gao, & Zang, 2010). Additionally, in the context of the present study, the main aim of primary-level English teaching is to develop students' communicative competence through a variety of listening and speaking activities (*National English Curriculum Standards*, 2011), whereas the main aim of university-level English teaching is to prepare students to succeed on the College English Test (*Syllabus for College English Test*, 2006). Given the above age-related differences, it is necessary to examine child L2 learners as a distinct group. The present study thus sought to extend previous examinations by examining both child and adult L2 learners so as to determine what, if any, processing similarities or differences exist between these two age cohorts in terms of patterns and tendencies.

An alternative way of examining age-related effects is to consider at what age a particular word is learned (i.e. *age of acquisition*). A widely accepted view is that early life experiences have a lasting impact upon our behavior in adulthood, and by the same token, our exposure to language in childhood is believed to have a persisting effect on how we perform in adulthood. Age of acquisition can therefore be seen as a simulation of learners' experience sequentially encountering and practicing L2 words. In this regard, examining the effect of age of acquisition on learners' processing efficiency would certainly provide insight into the basic developmental trend of acquisition processes, in and of themselves. Additionally, this specific-item lexical characteristic could also serve as a point of reference against which to assess its

power in predicting L2 lexical processing tendencies toward concept mediation or word association.

To sum up, the present study addresses the bilingual lexicon and L2 lexical processing in relation to word characteristics and learners' age. Only after having a clear understanding of this issue can we reach a confident conclusion about the nature of the L2 mental lexicon, and armed with this knowledge, we will be able to facilitate, as far as possible, L2 vocabulary learning, and to advance empirical and pedagogical implications for current conceptualizations of L2 teaching in input-limited settings.

1.2 Position in the literature

The unique characteristics of this study are discussed by pointing out the literature gaps that the current research intends to fill in.

1.2.1 Under-researched population: Child L2 learners

At the present young children are one of the fastest growing populations who are learning an L2 around the world (Kan & Kohnert, 2005). As regards the context of the present study, many Chinese state-run primary schools have included English in the curriculum from grade three onwards. Moreover, compared to adult L2 learners, child L2 learners are cognitively less mature; and compared to monolingual children, bilingual children have experienced learning an L2. Against this background, it is more appropriate to examine child L2 learners as a separate group rather than as identical to adult L2 learners or monolingual children.

Over the past decades there have been many attempts to examine the end state of L2 lexical processing reached by adult learners. For instance, the *Word Association Model* and the *Concept Mediation Model*, proposed by Potter et al. (1984), describe the possible relationships between the L1 lexicon and the L2 lexicon in advanced stages of L2 learning. It should be acknowledged that there do exist some models like the RHM, which assumes a developmental shift in the relative ease of accessing the conceptual route as learners' fluency increases (Kroll & De Groot, 1997). Note that these lines of research are almost exclusively based on between-subjects, or more exactly, between-adults comparisons, thus only reflecting a snapshot of processing differences between more and less proficient bilingual adults. Under this scenario, it seems hard to draw confident conclusions as to whether the same pattern of developmental shift can be observed across all age groups. Additionally, development should be defined not only by learning outcomes but also viewed as a life-span process, which means that L2 lexical processing is not only moderated by learners' proficiency levels but also regulated by the age. Examining the role of age in L2 lexical processing is thus necessary for better characterizing the bilingual lexicon in development.

However, little work has been done on children's emerging bilingual lexical representations and processing, especially those learning an L2 within input-limited contexts, possibly due to the fact that this population has only recently begun to receive focused attention in the research literature. In order to address this literature

gap, the present study recruited primary-level students (aged 10 or 11 years) to participate in the experiments.

Finally, one purpose of the present study is to assess the predictive effects of a set of lexical and learner variables on the processing efficiency of L2 words. Most studies addressing these issues have been targeted at bilingual adults, particularly university students who have English as an L2 (e.g. De Groot, 1992b; Hirsh, Morrison, Gaset, & Carnicer, 2003; Jin, 1990). It thus remains unclear the extent to which and in what manner these variables predict children's processing speed, let alone to compare the contribution made by each variable across age groups.

1.2.2 Under-researched or under-estimated influences: Lexical and learner variables

Representations of lexicons constitute both modal and amodal nodes, with which to manifest certain lexical characteristics and to reflect individuals' experiences with words. The processing mechanisms operate in ways that correspond to the types of lexical representations involved; in other words, there could be some interactions between lexical variables and processing mechanisms. Thus, it would be deemed necessary to examine in what circumstances and with what lexical characteristics concept mediation or word association more likely occurs. The past literature predominantly examined mechanism differences between concrete and abstract words, as evidenced by De Groot's proposal of *Distributed Feature Model* (1992a, 1993). But apart from this, we know much less about how other lexical variables (e.g. age of

acquisition, semantic neighborhood density, and word typicality) behave in influencing the likelihood that a certain pattern of lexical processing is in operation. This proposed study sought to assist in closing the current gap in the literature by incorporating a wide range of important lexical variables.

Moreover, there are two ways to normalize lexical variables: one is to use existing databases in which values of lexical variables are provided with regard to a large number of words. Typical examples include those comprehensive databases like the Medical Research Council (MRC) Psycholinguistic Database (Coltheart, 1981) and those data available in published research reports or journal articles (e.g. Alario & Ferrand, 1999). The second way is to require participants to rate lexical variables on scales through which subjective judgments on specific variables are made. The first approach has been more commonly used than the second approach to normalize characteristics of stimulus words in previous studies examining lexical effects (e.g. Alario et al., 2004; C. M. Morrison, Ellis, & Quinlan, 1992; Severens, Lommel, Ratinckx, & Hartsuiker, 2005).

When both options are theoretically available in the present study, our preference was given to the subjective approach on the following two grounds. First, as shown in the mega study of 2,428 single-syllable English words (Balota et al., 2004), subjective estimates were found to better predict lexical decision and naming performance above and beyond objective word frequencies within five different existing corpora. What this suggests is that subjectively rated values are likely to reflect, as closely as

possible, participants' real understanding and usage of target words, and thus to better capture the variance in their performance, even though collecting subjective estimates from a large number of participants is a relatively time-and resource-consuming endeavor. Second, most psycholinguistic databases are constructed with exclusive reference to monolingual users' resources or estimates. Given this, generalizing these monolinguals' data to examine and describe bilingual learners' performance should be carried out with caution because when objective norms do not correlate well with those generated by participants themselves, the magnitude of lexical effects could not be accurately estimated.

Another point to stress is that lexical characteristics have been examined as discrete variables in almost all previous studies, but in fact, some of them correspond to the same layer of speech production or the same aspect of lexical knowledge. It could therefore be argued that some related variables can be reasonably combined into a joint variable to represent their underlying construct; in so doing, it will be possible to better understand how lexical characteristics relate to each other at a more macro-level, a purpose that this study attempts to achieve.

Finally, a widely accepted view is that how L2 words are processed is inseparable from characteristics of individual learners. Previous studies have focused on examining the roles that proficiency levels or fluency levels play in bilingual lexical processing (e.g. Chen, 1990; De Groot & Poot, 1997; Francis, Tokowicz, & Kroll, 2014). For example, in the study by De Groot and Poot (1997), forward and backward

translations were compared across three groups of 20 unbalanced bilinguals whose proficiency or fluency was operationally defined by the grade in which the participants were enrolled. This line of research holds implications for the interaction between proficiency levels and L2 lexical processing but it should be noted that proficiency, in and of itself, is a multi-component construct mediated by learners' knowledge in terms of vocabulary, grammar, pronunciation, fluency and sociolinguistic respects, and the ability to use them (Luecht, 2003), and moderated by the amount of exposure, types of exposure, age of acquisition and other variables related to learning experiences.

Given the complexities involved in operationalizing language proficiency, it is necessary to examine how each specific aspect of linguistic knowledge likely affects learners' productive performance. Additionally, as mentioned above, learner variables should not be defined only in terms of learning attainments but should also include a set of learning process-related factors. Only after we have examined how every possible kind of learner variable influences bilingual lexical processing are we able to unify all relevant evidence to propose a model that will accommodate what have thus far been observed among different types of learners.

1.2.3 Processing patterns and tendencies

'Processing pattern' is an often-used term in the bilingualism literature to characterize the mechanisms responsible for accessing and retrieving L2 words. As mentioned above, there are two typical patterns of lexical processing in bilinguals: word

association and concept mediation, and the activation of each largely depends on learners' proficiency levels. Arguably, these patterns can be seen as two extremes of the continuum on which both patterns are manifested in L2 lexical processing but in different proportions.

Indeed, many studies examining the mechanisms underlying L2 lexical processing have given support to the simultaneous presence of both processing patterns in more-fluent and less-fluent L2 learners (for reviews of these studies, see Kroll, Bobb, Misra, & Guo, 2008; Kroll et al., 2006). Given this, we should not only address which model is supported, but also, under what circumstances L2 lexical processing tends to move toward word association or concept mediation, an issue that the present research attempts to explore.

This 'likeliest' view of lexical processing attaches importance to comparing the weights of these two processing alternatives through which to identify the reasons why one route is more dominant than the other, and emphasizes the dynamics of bilingual lexicons due to the influence arising from learners, lexical items, and tasks, and constant interactions between these factors.

1.2.4 Methodological inadequacy

In most studies, lexical effects have been examined using the analysis of variance (ANOVA). As pointed out by Balota et al. (2004), this approach allows for identifying the important variables that can modulate learners' performance, but there are several difficulties associated with using it. One main difficulty is that it seems

hard to ensure that an outcome can be attributed only to the variable being manipulated. This is in part because it is no easy task to design experimental materials that are matched on every lexical aspect except for the independent variable of interest (Cutler, 1981). As a consequence, no wonder that the history of psycholinguistics encounters 'the continual discovery of new confounds' (Cutler, 1981, p. 65). Consider the processing superiority of concrete over abstract words, which has once been explained by the imageable system that concrete words have (Paivio, 1986; Paivio & O'Neill, 1970). Years later, Schwanenflugel, Harnishfeger, and Stowe (1988) reported that imageability effects disappeared when concrete and abstract words were matched on context availability. What this suggests is that caution should be taken in identifying the presence or the absence of certain lexical effects on the basis of a limited number of items that could potentially vary with the independent variable. When the problem of contamination occurs, it would be difficult to work out whether the independent variable or confounds are responsible for the observed effect.

Other difficulties associated with the use of ANOVA include: (1) it would be easy to lose information when continuous independent variables are put into categories; (2) subjects may become implicitly or explicitly sensitive to the variable being manipulated, since test words all have values representative of the targeted lexical dimensions; and (3) results drawn from studies examining a small number of specific variables may hardly hold general implications for the lexical processing

system as a whole. Although it is necessary to identify whether or not a particular variable affects lexical processing, equally, if not more important, how much of the performance variance that can be accounted for by that variable needs to be investigated (for more details, see Balota et al., 2004).

The above-mentioned difficulties could be partially minimized by using multiple regression models, in which language rather than the experimenter defines the stimulus set to some extent. Using this analytical strategy enables us to produce a more general picture, although it may be accompanied by the lack of a specific focus on certain variables. There exist a large body of research using multiple regression to estimate the unique variance predicted directly by an extended set of lexical variables (e.g. Bonin, Chalard, Méot, & Fayol, 2002; Cuetos, Alvarez, González-Nosti, Méot, & Bonin, 2006; De Groot, 1992b); however, as mentioned above, some lexical variables may have both direct and indirect effects on the outcome variable, in which case the use of multiple regression runs the risk of glossing over these indirect effects.

Given all of the above considerations, structural equation modeling (SEM) that combines path analysis and confirmatory factor analysis is a better methodological choice to use in the present study. This analytical strategy allows for performing a complete and simultaneous analysis of the complex relationships among a large set of lexical variables and for estimating direct, indirect and total effects of one lexical variable on another.

In a word, this study extends previous relevant research in terms of the intended

target population, the variables of interest, and the application of research methodology.

1.3 Purposes and significance of the present study

There are three main purposes of this study. First, it aims to assess the predictive effects of learner and lexical variables on the ease with which target words are processed in picture naming and forward translation across age groups. To be specific, this study attempts to examine the complex relationships between lexical variables themselves and to identify significant direct and indirect predictors of participants' processing efficiency, if any, by using structural equation modeling, and to investigate which learner variables contribute most significantly to response times by performing multiple regression analyses. Moreover, the latency variance explained by learner variables is compared with that explained by lexical variables so as to infer which dimension better predicts L2 lexical processing: learners' experiences with a particular lexical item or their overall L2 linguistic knowledge. This study also aims to make a cross-age comparison so as to identify whether or not there exist any variables that contribute significantly to the outcome variable in both age cohorts. This comparison result would provide insight into the structure of bilingual lexicons and the processing mechanisms in development.

The second main purpose of this study is to investigate the general pattern of L2 lexical processing across three groups of English as a foreign language (EFL)

learners[1]: 5th-grade primary-level pupils, university students with the L2 exposure less than 20 hours per week, and university students with the L2 exposure more than 20 hours. Which processing pattern can more appropriately describe each group's performance is examined by comparing the speed of picture naming with that of forward translation. In addition, whether there exist pattern differences between age groups, and between higher and lower L2-exposure groups will be addressed. Note that it is theoretically untenable to compare the speed of naming and that of translation between child and adult L2 learners in a direct manner because these two populations are fundamentally incomparable in many respects, but it is reasonably possible to make a qualitative comparison between them in terms of processing patterns.

The third purpose of this study is to explore whether there are any learner and lexical variables that significantly predict the tendency toward a certain pattern of lexical processing by computing linear relationships between learner-and-lexical predictors and naming-versus-translation latency differences. In other words, the researcher intends to address in what circumstances word association or concept mediation is more likely to occur with what types of words and among what types of learners. Comparisons are also made across age groups so as to understand whether or not the variables of interest all behave in the same way irrespective of age. Armed

[1] The terms 'L2 learners' and 'EFL learners' are used interchangeably in the present study to refer to those who are learning English mainly through instruction in classrooms in non-native English-speaking countries.

with this knowledge, we are able to propose a modified model of bilingual lexicons, which specifies the conditions under which the lexical or the conceptual route is more predominantly activated in spoken production of L2 words.

The present study is significant in four major ways. First, as mentioned above, most previous studies have been targeted at bilingual adults, and accordingly little attention has been given to bilingual children. Child foreign language learners were recruited to participate in this study so as to obtain information about lexical representations and processing mechanisms at a given point in development. Conducting a parallel study of spoken production that involves both children and adults allows for comparing performance across the usual predictors of naming and translation speed and identifying differences, if any, in processing mechanisms between age groups. Hence, this study is believed to be both original and significant in terms of the target population being investigated.

Second, this research includes a wider range of learner and lexical predictors than have been used in previous studies of bilingual spoken production. For example, it included three different measures of word length (in letters, phonemes, and syllables), two different measures of word usage (subjective frequency ratings and age of acquisition), and a set of meaning-related variables such as word typicality, semantic neighborhood density, and the number of related senses, which have not been fully addressed before. As regards learner variables, in particular, two assessments of vocabulary (size and depth), onset age of L2 learning, and the amount of L2 exposure

were included in the present study. To the best of our knowledge, the predictive effects of these learner variables on the performance of bilingual speakers have not yet been examined in the literature.

Third, this study represents an initial attempt to take an exploratory approach to address what kinds of words are more likely to be lexically or conceptually processed in what types of learners. The results of which will add details and richness to the functional architecture of the bilingual lexical processing system, or more exactly, will illustrate the means by which box-and-arrow diagrams operate and under what circumstances a specific arrow flows between and in and out of a particular component of the system. Furthermore, the adoption of a holistic view of the conditions under which a certain processing tendency likely occurs paves the way for future examinations of each crucial variable in a separate manner.

Fourth, this study uses SEM instead of multiple regression analysis as a methodological improvement to assess direct, indirect and total effects of lexical predictors on response latencies. In addition to that, it is allowed to test the extent to which the model of interest is consistent with the observed data by examining its fit indices. It seems to me that this is the first study that uses SEM to capture complex relationships between lexical variables themselves and then to model their effects on the speed of spoken production. Additionally, latent variables are formed to manifest or to reflect the general dimensions that are underpinned its specific indicators. In so doing, this study moves away from the examination of each lexical variable to that of

the specified constructs themselves and structural relations between constructs, thus a better understanding of the nature of lexical characteristics can be gained at a more macro level.

To sum up, the present study set out to make empirical contributions to the bilingualism literature by taking age into account, by including as many predictor variables as possible, by examining determinants of L2 lexical processing tendencies, and by improving upon current analytical strategies. It is hoped that the results of this study will provide information about the nature of lexical processing specific to child and adult L2 learners and thus to hold implications for promoting age-appropriate pedagogical practice in input-limited foreign language learning contexts.

1.4 Structure of the thesis

The present research concerns areas that can be roughly described by five keywords: bilingual lexicons, bilingual lexical processing, spoken production, child foreign language learners, and adult foreign language learners. This thesis is structured as follows. Chapter 2 provides theoretical and empirical grounds on which to situate this study. The first section discusses the structures of bilingual lexicons and tracks the development of relevant models. The second section elaborates on the components of spoken production and then presents characteristics specific to bilingual production. After this, the lexical variables under examination are discussed in detail one by one, along with a review of their effects on processing efficiency. The last section discusses age-related differences in biological, cognitive and social respects.

Chapter 3 presents a detailed description of methodological issues, including methodological rationale, research methods, research design, research questions, population and sample, research procedures, analytical strategies used, ethical considerations, and validity, reliability, and generalizability of the study. In this chapter, a number of figures and tables are provided wherever appropriate to diagram research processes and steps.

Chapters 4 and 5 report on the results of data analysis obtained from the child participants and from the adult participants respectively. These two chapters are organized corresponding to the order of research questions. The most powerful learner and lexical predictors of the response latencies are identified, followed by an investigation of processing patterns based on by-subject and by-item analyses in both age groups. In what follows, processing tendencies serve as the outcome variable against which significant learner and lexical predictors are identified.

Chapters 6 and 7 provide possible explanations to account for the findings obtained in the present study. A bilingual-lexicon model that considers age and lexical effects is proposed, and its scope of applicability is then discussed. After this, limitation of the present study will be summarized and direction for future research proposed. In the following section, pedagogical implications for current conceptualizations of vocabulary teaching and learning within input-limited contexts will be presented. This chapter ends with a summary of the main findings in this thesis.

Chapter 2 Literature Review

This chapter defines the relevant key terms for this thesis and reviews the available literature on the structure of bilingual lexicons and the mechanisms underlying bilingual lexical processing. This is followed by a discussion of spoken production models and the processes whereby L2 words are produced. After that, a range of lexical characteristics and their effects on language processing are elaborated. The last section addresses age-related differences in relation to lexical processing.

2.1 Defining key terms

2.1.1 Bilingual

The term bilingual has been defined in many different ways, one of which relates to the proficiency levels that a language user has achieved in different languages. Under this approach, there exist two extreme views: one is represented by Bloomfield (1933), who considers bilinguals should have 'native-like control of two or more languages' (p. 55); and the other, represented by Macnamara (1969), regards bilinguals as possessing one of the four language skills, even to a minimal degree. Between these two extremes lie many variants that differ in the proficiency requirements for being labeled as a 'bilingual' (e.g. Haugen, 1953; Titone, 1972).

Rather than defining the term precisely, some scholars have developed a typology to capture the characteristics specific to each type of bilingual (see Li, 2000, for a summary). In line with this approach, the adult population involved in the current

study can be described mainly as *unbalanced bilingual*s, whereas the child population can be seen mainly as *emergent bilinguals*.

2.1.2 Mental lexicon

The mental lexicon is often compared to a '*mental dictionary*' (Aitchison, 2012), in that both contain information about a word's lemma (i.e. semantics and syntax) and lexeme (i.e. morphology and phonology/orthography) (Levelt, 1989). In contrast, these two terms are different in such a way that the mental lexicon has not only a *realistic referent* but also a *mentalist reference*–'language users' conceptualization of the world' (Jackendoff, 2011, p. 690), which can hardly be found in a dictionary. Another difference is that dictionary words often have definite and fixed meanings; whereas mental lexicons appear to be indefinable, standing on 'a family of meanings' (Wittgenstein, 1953/2009, p. 77) growing out of each individual's interactions with and perceptions of their physical and social worlds.

Taken all the above together, the mental lexicon, briefly termed, refers to what is stored in each individual's mind in relation to words' conceptual representations, semantic meanings, syntactic structures, morphological properties, phonological features, and orthographical shapes, a mixture of which is used by humans to connect with the outside world, and their experiences in turn shapes the content of their mental lexicons.

2.1.3 Lexical processing

Lexical processing is a term with many uses in the field of psycholinguistic and neurolinguistic studies. At the broad end, it embraces any description of what happens inside the brain when words are accessed, activated, stored and retrieved. Understanding such a process involves examining how lexical forms are mapped onto word meanings or vice versa, how word meanings are represented and organized in the mind, and what factors likely impact upon different stages of lexical processing. Such a broad definition, however, may hardly reflect the scope and characteristics of the present study. As such, a more specific conceptualization of lexical processing is adopted instead, that is, the cognitive mechanisms by which L2 words are accessed and processed in the spoken production of single words, typically in picture naming and forward translation.

At the empirical level, lexical processing has been researched from two main perspectives: one pertains to the state in which L1 and L2 specific information coexist or conflict with each other, as well as the state in which the L1 or the L2 dominates the lexical entry. The second perspective concerns the stages through which conceptual activation flows to articulation during L2 production, as well as the connections that may exist between each stage. Knowledge of how such processing systems work may allow us to infer the state of mental representations of L2 lexicons in bilingual learners. The literature concerning the state of bilingual memories and bilingual lexicons will be reviewed in the following sections, but before that, a brief

description of the memory system is presented first as it holds implications for the process of word retrieval.

2.2 Lexical processing in memories

2.2.1 Memory system and semantic memories

Various forms of memory have been identified by manipulating variables of interest (e.g. retrieval conditions, measurement of tests, and task demands) and by including natural variables (e.g. expertise, age, and memory deficits) (Lockhart, 2000). Although the exact number and nature of different memory subsystems remain contentious (Squire, Rolls, Johnson, & Buckner, 2007), it has been generally agreed on the existence of 'episodic, semantic, procedural and working memory, not to mention declarative versus non-declarative memory, and explicit versus implicit memory' (Schacter, 2007, p. 23). The memory system (see Figure 2.1) most relevant to the present study is the declarative one, which is associated with the processing of lexical items (Ullman, 2001, 2004). By definition, declarative memories refer to the capacity for conscious recollection of facts and events (Squire, 1992). The major distinction between declarative and non-declarative memories likely rests on the way by which mental contents are retrieved: in most cases, declarative memories are accessed consciously through verbal reports of facts, experiences and events or through tests on a 'knowing that' type of knowledge; whereas non-declarative memories are accessed unconsciously through changes in behavior or performance due to practice and repetition accumulated over time (Squire, 1992).

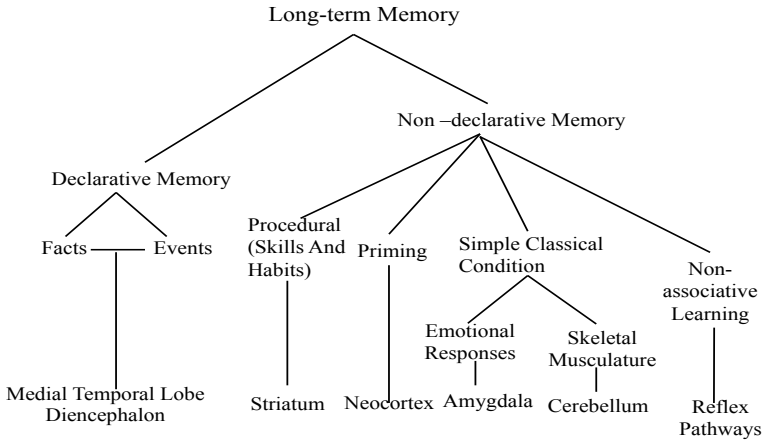


Figure 2. 1 A taxonomy of long-term memory systems (adapted from Squire & Zola, 1996)

The declarative memory system can be further divided into semantic and episodic memories. The former serves to store organized knowledge such as words, concepts, and relations, in contrast to episodic memories that 'store information about temporally dated episodes and events' (Tulving, 1972, p. 385). Functionally speaking, both memory forms are closely intertwined with language acquisition and processing, and each has its own distinguishing features. First, information stored in semantic memories can be retrieved in a relatively context-free manner as compared to that stored in episodic memories. In other words, semantic memory retrieval can be operationalized without referring to the specific contexts in which a lexical item is acquired, whereas episodic memory retrieval needs to rely on the specific language through which a concept is acquired.

Second, with particular episodes having merged into general categories, semantic memories show category specificity, as evidenced by the selective impairment of

knowledge stored in semantic memories (e.g. Farah & McClelland, 1991; Tyler, Moss, Durrant-Peatfield, & Levy, 2000; Warrington & Shallice, 1984). A case in point is that patients with Alzheimer's disease encounter more difficulties in processing lexical items that represent natural objects relative to those that represent manufactured objects (Garrard, Patterson, Watson, & Hodges, 1998). One plausible explanation for such a category-specific deficit is related to the neurally distributed networks in which sensory, motor and/or more abstract information that a lexical item possesses are stored in different regions of the brain, thus processing lexical items of a certain category is likely to be influenced by its own category-specific attributes and features (Farah & McClelland, 1991; Warrington & Shallice, 1984). Results of these studies hold important implications for the present research at least in the sense that lexical representations are inherently distributed with respect to lexical characteristics.

Third, semantic and episodic memories interact with each other in such a way that episodes can turn into relatively stabilized semantic memories with practice over time (Baars & Gage, 2010). What this suggests is that the degree to which memory episodes are entrenched in the brain reflects the ease or difficulty of retrieving semantic memories.

The above section has briefly discussed the involvement of declarative memories in lexical processing. The following section will focus on intersections of memory and language systems in bilingual learners.

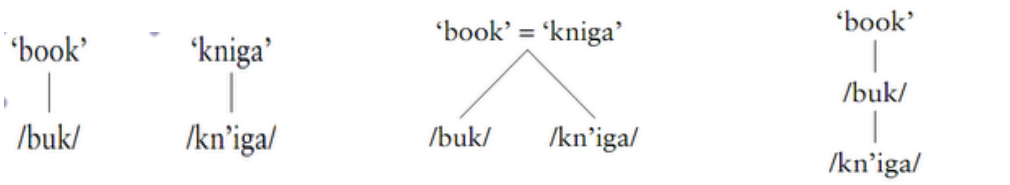
2.2.2 Bilingual lexicons and lexical processing: Structures, mechanisms, and models

It is possible for bilinguals to communicate effectively in a single language, but this does not mean that the language not in use is switched off. As a matter of fact, both languages remain active but to different extents. Over the past few decades, several models have therefore been proposed to specify the structure of bilingual lexicons and to delineate the mechanisms underlying L2 lexical processing. These models can be arranged in a chronological order: the pre-Revised Hierarchical Model (RHM) phase, the RHM phase, and the post-RHM phase.

The attempt to develop models, which accounts for both experimental and non-experimental data, has a history of work stretching from small models with a restricted range of application to more comprehensive models with a wider range of application. It should be pointed out that older models have not been replaced but have been modified by newer models, as the scope of problems they could solve became increasingly narrow or because they lacked explanatory power to account for certain kinds of observations. Relevant models will be reviewed and assessed in the following sections so as to provide a framework for understanding the structure of bilingual memory, as well as for understanding the mechanisms sustaining interactions between the L1 and the L2.

2.2.2.1 First phase: Pre-RHM

As early as half a century ago, Weinreich (1953) proposed three types of possible bilingual representations based upon the relationship between a signifier and its signified across languages (see Figure 2.2): *coordinate* (i.e. conceptual system is language dependent), *compound* (i.e. L1 and L2 lexicons share one conceptual system), and *subordinate* (i.e. L2 words are learned through their translation equivalents). Weinreich also pointed out that the transition from type C to type A might occur in language learning.



A. Coordinate B. Compound C. Subordinate

Figure 2. 2 Three types of possible bilingual representations (adapted from Weinreich, 1953)

Several studies supported such a coordinate-compound distinction (e.g. Lambert, Havelka, & Crosby, 1958; Osgood & Sebeok, 1954); however, the absence of this distinction was also found in a study by Kolers (1963). As Durgunoğlu and Roediger (1987) commented, these inconsistent results could be attributed to the use of different tasks, which reflected different levels of mental representation due to varying retrieval demands.

Decades later, two hierarchical models were proposed by Potter et al. (1984) — the *Word Association Model* (WAM, see Figure 2.3) and the *Concept Mediation*

Model (CMM, see Figure 2.4). In the WAM, the L1 mediates the relationship between the L2 and its concept; by contrast, in the CMM, the routes of L1-to-L2 and L2-to-L1 are both conceptually mediated.

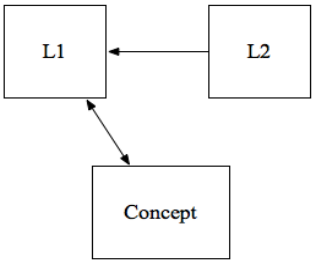


Figure 2. 3 Word Association Model (adapted from Potter et al. 1984)

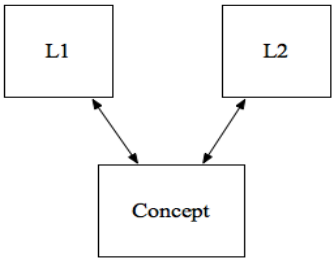
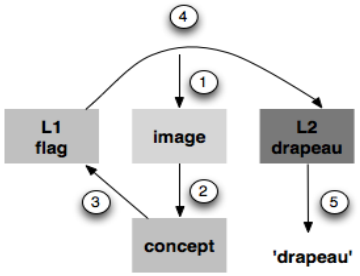
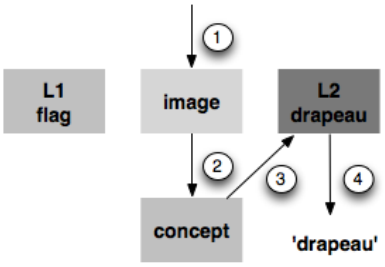


Figure 2. 4 Concept Mediation Model (adapted from Potter et al. 1984)

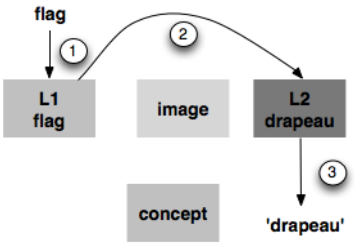
These two alternative models have been tested by Potter, et al. (1984). The logic behind their experiment is that: naming a picture can be seen as conceptually mediated. So, as illustrated in Figure 2.5a, if the L2 was directly connected to the L1 (i.e. WAM), time spent on naming a picture in the L2 should be more than time spent on the translation from the L1 to the L2. However, as shown in Figure 2.5.b, if the L2 was conceptually mediated (i.e. CMM), time spent on translation and on picture naming should be equally the same.



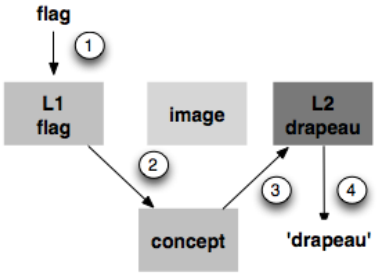
Picture naming in L2



Picture naming in L2



Translation into L2



Translation into L2

a. Word Association Model

b. Concept Mediation Model

Figure 2. 5 Logic behind Potter et al.'s experiment (Potter et al., 1984, p. 25)

Two experiments were conducted, one with proficient Chinese-English bilinguals and the other with non-fluent English-French bilinguals. The results supported the CMM and refuted the WAM, because there were no significant in-group differences in response latencies between picture naming and word translation for both proficient and non-proficient bilinguals.

However, subsequent empirical studies (e.g. Chen & Leung, 1989; Kroll & Curley, 1988) following Potter et al.'s logic reveal that the WAM is valid too. This is because the WAM best characterizes the performance of adult beginners. What this suggests is that learners' proficiency levels are likely to have an impact upon the structure of bilingual lexicons. Based on these findings, Kroll and De Groot (1997)

proposed the *Developmental Hypothesis*, claiming that there is a developmental shift from lexical to conceptual processing of L2 words with learners' fluency increasing. However, this shift cannot be predicted by the WAM and the CMM. Also, these two models fail to explain the observed asymmetry in translation (i.e. backward translation is faster than forward translation) (Kroll & Stewart, 1994).

2.2.2.2 Second Phase: RHM

In order to address these unresolved issues, Kroll and Stewart (1994) proposed the *Revised Hierarchical Model* in which a conceptual link and a lexical link coexist (see Figure 2.6). The RHM assumes a strong connection from the L2 to the L1 (represented by a solid line) but a relatively weaker connection from the L1 to the L2 (depicted by a dotted line). The link between words and concepts is hypothesized to be stronger for the L1 than for the L2. According to this model, an L2 word is first mapped onto its translation equivalent at the initial stage of learning. After this, the link between the word and its concept will be added onto the shared conceptual system. As learners' proficiency levels increase, the ease of accessing the conceptual link between an L2 word and its concept will be obtained.

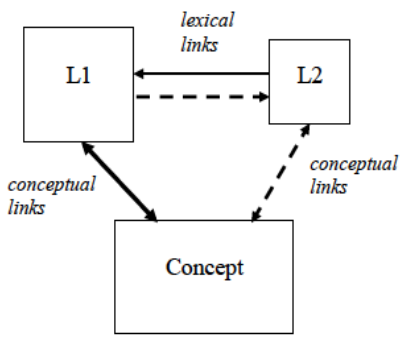


Figure 2. 6 Revised Hierarchical Model (Adapted from Kroll & Stewart, 1994)

The RHM has laid out a framework for understanding the relationship between the L1 and the L2 at lexical and conceptual levels, as well as the developmental progression of bilingual lexicons. The RHM is believed to be influential in many respects, not least for inspiring a considerable body of research that set out to assess its validity (for arguments supporting and refuting the RHM, see Brysbaert & Duyck, 2010; Kroll, Van Hell, Tokowicz, & Green, 2010). These studies, by and large, have contributed to addressing whether or not an asymmetric relationship exists between the L1 and the L2, and whether or not language non-selectivity occurs in accessing L2 words through a wide range of language comprehension and production tasks (see Table 2.1. for an overview of these studies). In order to make clear the contributions and limitations of the RHM, four strands of behavioral research examining the RHM will be reviewed below, namely: forward and backward translation, translation priming, processing of cognates, and the nature of semantics in L2 lexical processing.

Table 2. 1 Overview of studies examining the RHM

Author (year)	Task(s)	Participants	Conclusion(s)
Kroll and Stewart (1994)	Picture naming Word naming Word translation	English-Dutch bilinguals in L2 dominant environment	Asymmetry between L1 and L2 (backward translation was faster than backward translation) Category interference in L1-to-L2 direction
De Groot et al. (1994)	Word translation	Dutch-English unbalanced bilinguals in L1 environment	Meaning is more important for forward translation than for backward translation. Asymmetry between L1 and L2 particularly for non-cognates and concrete words (forward translation is faster than backward translation)
Gollan, Forster, and Frost (1997)	Masked translation priming Lexical decision	Hebrew-English bilinguals in L1 dominant environment (Exp. 1 and 3) English-Hebrew bilinguals in L2 environment (Exp. 2 and 4)	Asymmetric masked translation priming effects were obtained with different scripts.
Grainger and Frenck-Mestre (1998)	Masked translation priming Semantic decision Lexical decision	English-French bilinguals in L2 dominant environment	Semantic categorization was more sensitive to translation priming effects relative to lexical decision.
Jiang (1999)	Masked translation priming Lexical decision	Chinese-English bilingual graduates or visiting scholars in L2 dominant environment	Asymmetric masked translation priming effects (Exp. 1 and 2)
Talamas, Kroll, and Dufour (1999)	Translation recognition	English-Spanish undergraduates and graduates in L1 dominant environment (but some participants were Spanish native speakers before, but now were English dominant)	Support developmental hypothesis (from reliance on lexical form to reliance on semantic meaning)
Kim and Davis (2003)	Masked priming in lexical decision, naming and semantic categorization	Korean-English unbalanced bilingual university students	Task demands modulate priming results
Marian and Spivey (2003)	Eye tracking Spoken word recognition	Russian-English bilingual university students in L2 dominant environment	Parallel activation of two languages Between-language competition varies with stimuli, language background, and language mode.
Costa and Santesteban (2004)	Picture naming Language switching task	Highly proficient Spanish-Catalan bilinguals Korean-Spanish L2 learners Spanish-Catalan L2 learners	Asymmetrical switching costs for L2 learners

Table 2. 1 Overview of studies examining the RHM (Continued)

Author (year)	Task(s)	Participants	Conclusion(s)
Finkbeiner, Forster, Nicol, and Nakamura (2004)	Masked translation priming Semantic decision	Japanese-English bilinguals in L2 dominant environment	Prominent masked L2-L1 translation priming (Exp. 1) but not in lexical decision (Exp. 2)
Ferré, Sánchez‐Casas, and Guasch (2006)	Translation recognition	Spanish-Catalan bilinguals (early proficient learners, late proficient, and late non-proficient) in an L2 immersion context	Support developmental hypothesis Both lexical and conceptual are available to all bilinguals, regardless of their proficiency and age of acquisition.
Sunderman and Kroll (2006)	Translation recognition	English-Spanish bilinguals (a group of less proficient and the other group of more proficient) in L2 dominant environment	All learners, regardless of proficiency, showed interference for lexical neighbors and for meaning-related pairs.
Basnight-Brown and Altarriba (2007)	Masked translation priming Lexical decision (Exp. 2)	Highly proficient Spanish-English bilinguals in L2 dominant environment	Symmetric masked translation priming effects (Exp. 2)
Perea, Duñabeitia, and Carreiras (2008)	Masked translation priming Lexical decision	Highly proficient Basque-Spanish bilinguals in L2 immersion context	Symmetric masked translation priming effects (Exp. 1)
Hoshino and Kroll (2008)	Picture naming	Spanish-English bilinguals and Japanese-English bilinguals in L2 dominant environment	Cognate facilitation for different-script bilinguals
X. Wang and Forster (2010)	Masked translation priming Semantic decision	Chinese-English bilinguals in L2 dominant environment	Presence of L2-to-L1 priming for exemplars (Exp. 1)
Duñabeitia, Perea, and Carreiras (2010)	Masked priming Lexical decision	Highly proficient Spanish-Basque simultaneous bilinguals in L2 immersion context	Masked priming effects for both forward and backward directions

FORWARD AND BACKWARD TRANSLATION: WHICH IS FASTER?

The most direct evidence in support of the RHM comes from the observed asymmetry in translation latencies, that is, backward translation (from the L2 to the L1) is faster than forward translation (from the L1 to the L2) (e.g. Kroll, 1993; Kroll & Stewart, 1994, Experiment 3; Sholl, Sankaranarayanan, & Kroll, 1995). What these studies suggest is that forward translation is mediated through conceptual routes, whereas backward translation is mediated through lexical routes (see Figure 2.6 above). Although the RHM postulates that semantic components get more involved in L1 processing relative to in L2 processing, this does not mean that the latter is not able to access meaning at all (Kroll et al., 2010). This argument may explain why both directions of translation have been found to be sensitive to the manipulation of semantic variables (e.g. De Groot et al., 1994; Heij, Hooglander, Kerling, & Van Der Velden, 1996; Heredia, 1997).

Another point worthy of mention is that the translation asymmetry has not been observed in every piece of research. In fact, the opposite asymmetrical pattern—forward translation was faster than backward translation—has been observed with less proficient L2 users (e.g. De Groot et al., 1994, Experiment 2; Heij et al., 1996, Experiment 3). Additionally, the translation asymmetry has been found to be non-significant even for relatively proficient L2 users in some studies (e.g. De Groot et al., 1994, Experiment 1; Heij et al., 1996, Experiment 4).

It is difficult, therefore, to draw a definite conclusion on the basis of these

inconsistent results; consequently, the RHM has been assessed using other approaches, one of which is translation priming.

TRANSLATION PRIMING: ASYMMETRICAL OR SYMMETRICAL?

In a masked translation-priming task, a prime word is presented in one language very briefly and then obscured by a masking stimulus. This is followed by the presence of a target word in another language, based on which a lexical decision (e.g. Is the target word a real word or not?) or a semantic decision (e.g. Does the target word represent a part of a human body?) is then made by participants. One advantage associated with this type of task is that the participant is likely to be unaware of the prime word, and because of this, the likelihood of making strategic decisions based on anticipation can be reduced (Forster & Davis, 1984).

In the lexical decision task, the magnitude of the masked translation priming effect has been found to be larger in L1-to-L2 direction than in L2-to-L1 direction, leading to support the predictions postulated in the RHM (e.g. Dimitropoulou, Duñabeitia, & Carreiras, 2011; Gollan et al., 1997; Jiang, 1999, Experiment 1 and 2; Schoonbaert, Holcomb, Grainger, & Hartsuiker, 2011). For example, in the study by Jiang (1999), relatively proficient Chinese-English bilinguals were tested with two blocks of translation pairs, of unrelated pairs, and of non-word pairs in both directions of translation. As soon as the target stimulus appeared on the screen, the participants pressed a YES/NO button to specify whether or not it was a real word. The results showed that the translation-priming effect was stronger in forward relative to

backward direction regardless of the type of words tested (i.e. concrete or abstract).

Such an asymmetric priming effect, however, does not hold for all types of bilinguals, and one exception has to do with highly proficient balanced learners. In Basnight-Brown and Altarriba study (2007, Experiment 2), the performance of highly proficient Spanish-English bilinguals in lexical decision showed a symmetry of the masked translation priming effect, and a similar finding was reported by Perea et al. (2008, Experiment 1). These results can be explained by the developmental shift postulated in the RHM, that is, bilinguals with high L2 proficiency are able to access L2 words directly through the conceptual route rather via translation equivalents.

Furthermore, the pronounced L2-to-L1 translation priming effect has been observed in the semantic categorization task (e.g. Finkbeiner et al., 2004, Experiment 1, 2 and 3; Grainger & Frenck-Mestre, 1998; X. Wang & Forster, 2010). For example, in the study by Finkbeiner et al. (2004), a robust priming effect was found in backward translation among a group of Japanese-English bilinguals, but no such priming effect was obtained when the semantic categorization task was changed to the lexical decision task. This discrepancy could be due to the fact that semantic categorization is more sensitive to semantic variables as compared to lexical decision (Grainger & Frenck-Mestre, 1998). An alternative interpretation (i.e. the *Category Restriction Hypothesis*) could be that semantic information is likely to be restricted by the category when a lexical decision is made, and because of this restriction the effectiveness of an L2 prime is enhanced (Finkbeiner et al., 2004).

Taken together, previous cross-language priming studies have yielded varied results: asymmetrical priming effects and symmetrical priming effects due to different types of task used, different proficiency levels that the participants had, and different concepts of stimuli accessed and processed. These inconsistencies suggest that the bilingual memory system is not static but dynamic per se, depending on a variety of language-related, task-specific and learner-related factors. It therefore stands to reason that the nature of bilingual memory is more complex and dynamic than the RHM predicts, in terms of both processing mechanisms and developmental trends.

PROCESSING OF COGNATES: FACILITATIVE OR NOT?

Another important line of evidence supporting the RHM comes from the observed superiority of processing cognates (i.e. words that share meaning and spelling/pronunciation across two languages) over non-cognates. The logic underlying these studies is that if bilingual lexical processing is believed to be language-selective per se, conceivably, there should be no significant differences between the way that bilingual learners respond to cognates and non-cognates; otherwise, vice versa.

The study by Costa, Caramazza, and Sebastian-Galles (2000), for example, showed that Catalan-Spanish bilinguals named pictures in L2 22 milliseconds (ms) faster with cognate names relative to pictures with non-cognate controls. As the authors argued, the presence of this facilitative effect can be accounted for by the *Multiple Selection Hypothesis*, which assumes that both the target lexical node and its translation equivalent are active during bilingual lexical processing. Also, this result

indicates that the conceptual link may not substitute the lexical link even after learners have got extensive exposure to the L2, which is to say that the L1 may still play an important role when relatively fluent L2 learners process L2 lexical items.

It should be pointed out that the Costa et al. study examined the same-script language, which means that the stimulus pairs were phonologically and potentially orthographically similar. Several subsequent research studies have therefore attempted to examine whether or not cognate facilitation can be extended to different-script languages. In a study by Hoshino and Kroll (2008), for example, Spanish-English and Japanese-English bilinguals named pictures with cognates and pictures with non-cognates. Its result confirmed the advantage of processing cognates over non-cognates even for non-script languages, indicating that at least the phonological level of the native language remains active in processing L2 lexicons.

To account for the cognate effects, there exist other explanations in addition to the presence of co-activation of the two languages of a bilingual. Van Hell and De Groot (1998) reasoned that cognate translations are semantically more similar than non-cognate translations, and that the ease with which a lexical item is retrieved is sensitive to the frequency of usage. For these reasons, the semantic representations of cognate words are retrieved faster than those of non-cognate words. Another explanation relates to the learning process by which cognates are learned more easily than non-cognates, and such a difference may have a lasting effect on the speed with which L2 words are accessed at a later stage (Costa, Heij, & Navarrete, 2006).

Although more research is needed to identify the causal source of the cognate effects, what can be concluded from what we have reviewed so far is that the activation of non-response language reaches at least the phonological level in bilingual lexical processing.

THE NATURE OF SEMANTICS: SEMANTIC INVOLVEMENT IN L2-TO-L1 PROCESSING?

As reviewed above, there have been many empirical attempts giving support to the RHM but several challenges remain ahead. One of them relates to those studies showing an active involvement of semantic components in L2 processing, even among less proficient language learners. Heij et al. study (Experiments 3 and 4, 1996), for instance, showed that forward and backward translation of words (e.g. BOTTLE) that were accompanied by semantically related pictures (e.g. of a glass) was faster than those by semantically unrelated words. This result indicates that the semantic context is influential for both directions of translation, leading the authors to reconsider the validity of Kroll and Stewart's asymmetrical model. In another study by De Groot et al. (1994), meaning variables affected the speed of backward translation of non-cognates among unbalanced Dutch-English bilinguals, although to a lesser extent, than that of forward translation. This result serves to support a weak version of the asymmetrical relationship between the L1 and the L2 rather than a strong version of it.

The noticeable influence of semantic variables on L2 lexical processing has been observed in other experiments involving the use of word recognition, for instance, the

study examining Dutch-English bilinguals' letter-case judgment based on the semantic Simon paradigm (Duyck & De Houwer, 2008), Sunderman and Kroll's research into unbalanced English-Spanish bilinguals' performance on a translation verification task (2006), and the study investigating less fluent English-French bilinguals' performance on a semantic categorization task (Dufour & Kroll, 1995). It might be argued that the RHM underestimates the involvement of semantic components in L2 lexical processing among unbalanced bilinguals, but as Kroll et al. (2010) counter-argued that the RHM was not intended to model visual word recognition but to model spoken word production. In this sense, it should be made clear that those experiments using recognition rather than production tasks to assess the adequacy and accuracy of the RHM probably should be seen as somewhat inappropriate.

It seems that the RHM has not received consistent empirical support across all contexts. Given this, it would be helpful for us to step back and consider the necessary and sufficient conditions under which the RHM works so as to advance our understanding of the dynamics of L2 lexical access processing in spoken production. It comes as no surprise that many post-RHM attempts have been made to propose alternative models or to modify the RHM by emphasizing the influence arising from language variables, learner variables, and the learning process.

2.2.2.3 Third Phase: Post-RHM

Post-RHM examples include the *Spreading Activation Model* (SAM, Hermans, 2000;

Poulisse & Bongaerts, 1994), the *Inhibition-Control Model* (ICM, Green, 1986; Green, 1998), the *Bilingual Interactive Activation Model* (BIAM, Dijkstra & Van Heuven, 1998), the *Distributed Feature Model* (DFM, Van Hell & De Groot, 1998), the *Three-stage Vocabulary Acquisition Model* (TSVAM, Jiang, 2004), the Concept Selection Model (CSM, La Heij, 2005), *Schematic Representation of the Bilingual System* (SRBS, Costa, 2005), and the *Modified Hierarchical Model* (MHM, Pavlenko, 2009). Drawing on different theoretical frameworks and empirical evidence, these models vary widely in their scope, complexity, and purpose. In what follows, those models that are relevant to the RHM or the present study context will be reviewed in further detail, including the DFM, the MHM and the TSVAM. Note that those models that are relevant to language production, the SAM, the SRBS, and the CSM, will be reviewed in Section 2.3.2. The BIAM is beyond the scope of the present review since it addresses word recognition.

THE DISTRIBUTED FEATURE MODEL

To account for *concreteness effects* (i.e. concrete words are processed faster than abstract words), De Groot (1992a, 1992b, 1993; 1998) proposed the DFM (see Figure 2.7 below). This model distinguishes between the conceptual and lexical levels, much the same as in the RHM. The lexical representations of the two languages of a bilingual are relatively independent of each other, whereas their conceptual representations are shared to varying degrees, largely depending on the activation level of common features or nodes between the L1 and the L2. Concrete words, which

possess exactly or approximately the same meaning across languages and cultural boundaries, are assumed to have a high level of conceptual overlap irrespective of language-specific factors. In contrast, abstract words, which are relatively sensitive to cultural contexts and a complex set of variables that determines these contexts, may be endowed with divergence or zero equivalence in semantic meaning or conceptual underpinnings between any pair of languages. Such a word-type difference is illustrated in Figure 2.7 where concrete words have more conceptual representations (depicted as dark circles) in common than do abstract words.

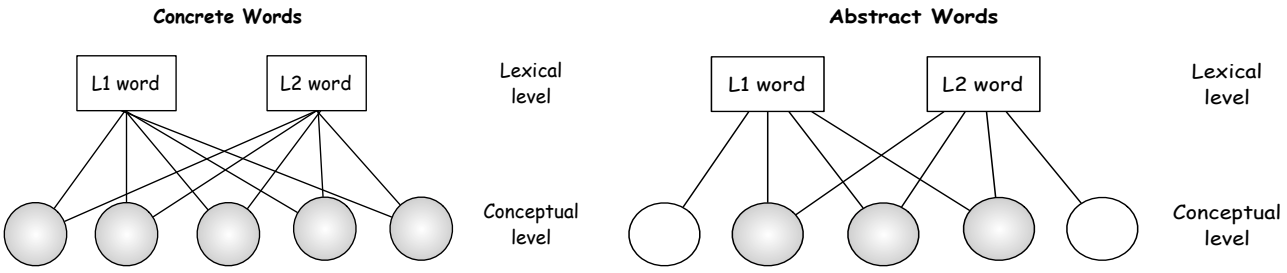


Figure 2. 7 Distributed Feature Model (adapted from De Groot, 1992b)

Empirical evidence supporting the DFM comes from experiments in which bilingual learners' performance on concrete words was compared against that on abstract words. For instance, in a study on unbalanced Dutch-English bilinguals, high-imageability nouns were translated faster and more often correctly compared to low-imageability nouns (De Groot, 1992b). Furthermore, concreteness effects were also observed in a study by Van Hell and De Groot (1998) who found that it took 681ms less time to come up with an associate for concrete relative to abstract words, and fewer association omissions occurred in concrete relative to abstract words. Such

a word-type difference may arise from the way in which conceptual nodes of concrete and abstract words overlap: a concrete word bears much resemblance to its semantically neighboring words, thus its associates can be produced with a relatively higher speed, whereas an abstract word and its semantic neighbors share less common meanings, thus accessing its associates takes more time and effort.

Drawing on the distributed notion of semantic representations, the DFM has made general claims about the effects of word concreteness on bilinguals' lexical processing. Moreover, the DFM holds explanatory power regarding the ways in which language interacts with cognition through emphasizing the profound influence that cultural contexts have on words' connotations and denotations (particularly, abstract words). In this sense, the DFM offers a way to interpret the similarities and differences that may exist between words across different languages.

However, it seems that the DFM fails to rule out certain possible lexical variables that may mediate the concreteness effect. A typical variable is the number of translation equivalents that a word has (i.e. translation ambiguity). It has been found that the concreteness effect might disappear when both concrete and abstract words had only a single or dominant translation equivalent, indicating that the ambiguity of semantic representations likely affects the extent to which word concreteness comes into play in L2 lexical processing (e.g. Tokowicz & Kroll, 2007). On the basis of these findings, it would be well advised to simultaneously include and assess a wide range of lexical variables so as to capture both accuracy and completeness in

ascertaining their effects on L2 lexical processing.

It should be acknowledged that the DFM has considered the role of semantics in bilingual lexical processing but it does not have much to say about the mechanisms underlying the conceptual-lexical mapping in L2 comprehension or production. As such, further work is needed to examine whether the DFM has equal explanatory power across different types of language tasks. Also, it remains to be specified whether or not the DFM is always valid, regardless of the way in which words are learned and the age at which these words are learned.

THE MODIFIED HIERARCHICAL MODEL

There has been an on-going debate as to whether thought affects language or vice versa. The first position is known as *linguistic universalism*, which holds that all languages are highly similar across cultures because they are underpinned by similar thoughts. The rival position is represented by *linguistic relativity*, popularly known as the *Sapir-Whorf Hypothesis*, which assumes that language not only serves as an instrument for expressing ideas but also as a shaper of mental activity. Due to the presence of such language-specific properties, some concepts encoded in one language are not completely commensurable with those encoded in another language. Based on this notion of linguistic relativity, the MHM, proposed by Pavlenko (2009), attempts to identify and capture the complexities and dynamics of conceptual relationships across languages.

The MHM acknowledges the existence of three types of between-language

conceptual relationships: equivalent or near equivalent, partial equivalent, and non-equivalent (see Figure 2.8). Although some words can be directly paired up with their translation equivalents, this is not always the case. Actually, under many circumstances, words are partially overlapping or even non-equivalent across languages. It is worth mentioning that the MHM, compared to other earlier models, represents an initial attempt to explicitly acknowledge the existence of conceptual non-equivalence. In other words, this model points out that some conceptual representations are completely language-specific, for instance, *privacy* and *personal space* do not have translation equivalents in Russian (Pavlenko, 2009).

Conceptual asymmetries, as noted by Pavlenko (2009), turn out to complicate L2 vocabulary learning, in that learners may find it difficult to map newly encountered L2 lexical items onto their already well-established conceptual categories and systems. When this occurs, learners are likely to exhibit non-target-like performance as a result of L1 conceptual transfer (i.e. using L2 words according to L1 concepts). By contrast, when L1 and L2 concepts are equivalent or nearly equivalent, positive transfer is most likely to occur, a process by which learners' L2 vocabulary learning can be facilitated to a great extent.

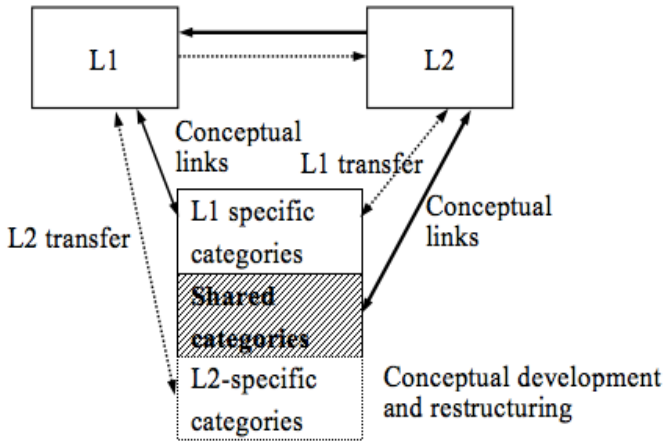


Figure 2. 8 Modified Hierarchical Model (Pavlenko, 2009, p. 147)

Evidence supporting the MHM comes from experiments showing cross-linguistic variations in naming tasks (e.g. Athanasopoulos, 2009; Malt, Sloman, Gennari, Shi, & Wang, 1999; Sachs & Coley, 2006), in categorization and sorting tasks (e.g. Cook, Bassetti, Kasai, Sasaki, & Takahashi, 2006; Gennari, Sloman, Malt, & Fitch, 2002), and in narrative elicitation tasks (e.g. Pavlenko, 2008; Pavlenko & Driagina, 2007). Using these task types allows researchers to identify subtle cross-linguistic differences as well as to infer word usage in contexts (Pavlenko, 2009). The Malt et al. (1999) study, for example, examined how 60 common containers were named by speakers of American English, Mandarin Chinese, and Argentinean Spanish in their L1s, the results of which demonstrated the presence of different naming patterns across languages. As an example, 15 objects all named 'container' in American English corresponded to four different linguistic items in Mandarin Chinese.

The MHM takes a 'contrastive' approach to modify the RHM by specifying different degrees of conceptual overlap across languages. The MHM views L2

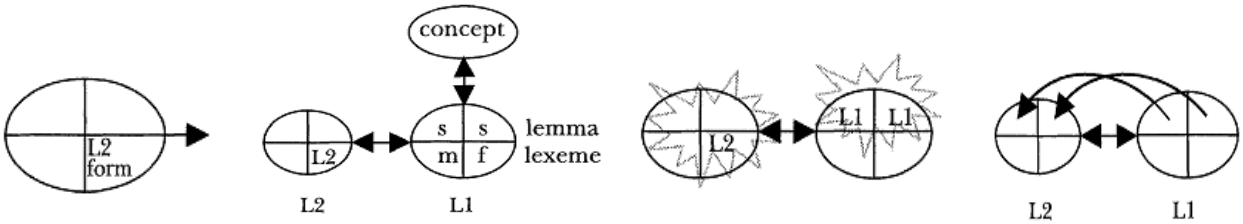
learning as a process of conceptual restructuring and development, and attributes L2 learners' non-target-like performance to negative semantic and conceptual transfers. It should be noted that the extent to which conceptual asymmetries are generalizable to all linguistic categories remains to be specified, as previous research has focused exclusively or predominantly on container names, emotional expressions, color names, and motion events. Additionally, future attempts are needed to examine the effects of cross-language conceptual symmetries and asymmetries on the rate, route and end-state of L2 vocabulary learning.

THREE-STAGE VOCABULARY ACQUISITION MODEL

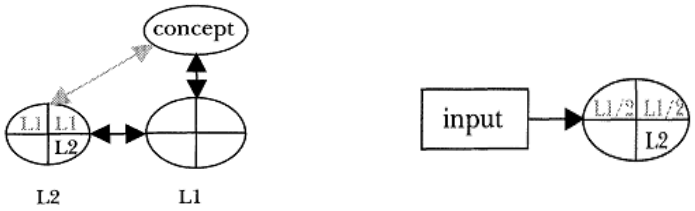
Different from the models described above, the TSVAM proposed by Jiang (2000, 2004) specifies the developmental route of L2 lexical representations in classroom settings. According to this model, the route involves three stages, including a word association stage, an L2 lemma mediation stage, and a full integration stage (see Figure 2.9).

Specifically, the first stage attempts to map an L2 word's phonological and orthographic specifications onto its translation equivalents' *lemma* (i.e. semantic and syntactic) and *lexeme* (i.e. morphological and phonological) information. L1 lemma information will be consequently transferred to an L2 lexical entry due to the repeated activation of the L1 in using the L2 word. The second stage is that the already transferred L1 lemma contents act as a bridge between the L2 word and its concept. As the contextualized input exposure continuously increases, the newly developed L2

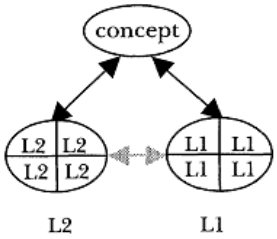
specific meaning will be added to the L2 word, and then the L2 lemma will contain both L1 and L2 specifications. In the third stage, as the strength of the L2 lexical link becomes increasingly weak, the L2 conceptual link will be established to the extent that L2-specific knowledge dominates the L2 entry.



First stage: word association



Second stage: L2 lemma mediation



Third stage: full integration

Figure 2. 9 Three-stage Vocabulary Acquisition Model (Jiang, 2004, p. 418)

It should be noted, as Jiang (2000) argued, that it might take longer than expected to complete these three stages. Even worse, the dominance of L1 lemma information blocks out specific L2 meaning representations, which makes it difficult for learners to undergo the transition period and likely leads to lexical fossilizations. However, in

principle, learners are able to reach the full integration stage provided that sufficient and highly contextualized L2 input is available for them and has been processed by them.

The most direct evidence supporting the TSVAM comes from the observed phenomenon of L1 semantic transfer. In Jiang (2004), Korean-English speakers and native speakers of English were tested with a semantic judgment task. Participants were asked to decide whether or not 128 English word pairs were related in meaning. Half of the 64 word pairs shared the same translation equivalent in Korean, for example, *chance* and *opportunity*, both translates as 기회 in Korean. The results showed that Korean-English speakers responded to the same-translation pairs faster than to different translation pairs; however, the *same-translation effect* was not observed with native speakers of English. This research confirms the presence of L1 lemma mediation in processing L2 words, even among highly proficient bilinguals. But it should be pointed out that the reported same-translation effect may be due to teachers' use of L1 translation in providing L2 lexical information, suggesting that if Korean-English speakers had learned the L2 words in a more natural and contextualized way, reasonably, the research results would be different.

Other supporting evidence, as summarized by Jiang (2000), comes from two main areas: studies on bilingual lexicons in lexical decision, naming, masked priming, and translation recognition, and studies on word choice, usage errors, and morphological errors in L2 production. For example, in a translation recognition task

(Kroll & De Groot, 1997), less fluent bilingual learners showed a large form interference from the L1 but its meaning interference was relatively small. A converse interference pattern, that is, a large meaning interference effect but a small form interference effect was found with more fluent bilinguals. This result supports the existence of a transition from the word association stage to the L2 lemma mediation stage as learners' proficiency levels increase. Additionally, it has been shown that some of the L2 learners' production errors were caused by L1 negative transfer (e.g. Dalgish, 1991; Zughoul, 1991).

To sum up, this three-stage model is consistent with the RHM, at least in the sense that both confirm the role that the L1 likely plays in L2 lexical processing, although its effect may vary according to L2 learners' language experience. The TSVAM is the first to focus on the process rather than the end state of bilingual lexicons by specifying the developmental route of L2 vocabulary learning within instructional contexts. The generalizability of the TSVAM remains to be studied to see whether or not every word follows the same developmental pattern and whether or not teaching strategies have a significant impact upon the developmental route.

2.2.3 Interim summary

As mentioned above, the development of bilingual models appears to follow the Kuhnian theory in which the history of scientific development involves the *pre-paradigm* stage, the *normal science* stage and the *revolution science* stage. The pre-RHM research studies operate in a pre-paradigm stage, in which preliminary

attempts are made to explore all possible structures of bilingual lexicons, as represented by Weinreich's (1953) work. The proposal of the RHM indicates entering into the normal science stage given that the RHM has provided a solid framework for a large body of research to be carried out. The post-RHM period seems to stand between 'normal science' and 'revolutionary science' (Kuhn, 1962), in which progress has been cumulatively made to address 'anomalies' that cannot be convincingly explained by the RHM and simultaneously to propose and try out new methods, theories, and models.

It is hoped that future models would explain more types of L2 lexical-processing-related phenomena, provide more information about underlying mechanisms, and are able to unify apparently disparate empirical evidence. Additionally, most previous models have been framed in a functional architecture, with very little detail paid to the processing algorithm or the conditions under which predicted outcomes would occur. In this light, the model that the researcher will propose attempts to accommodate previous research findings, as far as possible, and to clearly indicate under what conditions the model is valid.

The following section will discuss lexical processing with particular regard to spoken production and describe the characteristics specific to bilingual learners' production of L2 words.

2.3 Lexical processing in spoken production

Languages are used to communicate both through the process of comprehension

(listening and reading) and production (speaking and writing). For most people, these language activities permeate their lives; however, listening and speaking tend to occur more frequently than the literacy-based activities. Another major difference between these four types of language activities lie in how various sources of information are integrated in each process. It is often the case that listeners and writers map acoustic signals or written codes onto the entries in the mental lexicon, whereas speakers and writers face the converse order, a primarily top-down process that starts with conceptual intentions and ends with the production of corresponding language output (Treiman, Clifton, Meyer, & Wurm, 2003). Obviously, every language activity makes its own contribution to human communication but it seems implausible to cover all these language activities within the limited space available. Therefore, the focus of the present study is on spoken word production only.

The following sections will review the most important theories and models regarding monolingual speakers' production of single words since these studies have provided a conceptual framework for research on bilingual learners.

2.3.1 Monolingual speaker's production of single words: Components, stages, and models

2.3.1.1 Components and stages

It is widely held that spoken production proceeds through conceptualization, formulation, articulation, and self-monitoring (see Levelt, 1989, for a review). Each stage has its own output, namely, 'lexical concepts, lemmas, morphemes,

phonological words, and phonetic gestural scores' (Levelt, 1999, p. 3). A descriptive summary of what occurs and what can be achieved at each stage is presented below. It should be made clear that L1 and L2 spoken production are basically sustained by similar psycholinguistic mechanisms although there exist some differences between them (for a discussion on this issue, see Kormos, 2006).

The very first stage of spoken production is called *conceptualization*, which serves as an 'illocutionary act' (Searle, 1969) to produce utterances in compliance with conventional procedures and individual intentions. This stage demands a speaker's *macroplanning* and *microplanning* (Levelt, 1999). Macroplanning is the process by which a speaker decides what to say in order to meet the intended communicative needs. Microplanning, by contrast, is the process by which specific contents are incorporated into a message by following over-arching principles such as Grice's conversational implicatures (i.e. the maxim of quantity, the maxim of quality, the maxim of relevance, and the maxim of manner, see Grice, 1975).

Of particular interest in the present study context is the issue of whether or not the conceptual level is identical to the semantic level. In some models and theories, these two levels seem inseparable from each other (e.g. De Groot, 1993; Kroll, 1993; Kroll & Stewart, 1994) at least because both word meanings and conceptual representations are in a network of interrelated memory traces. However, other researchers (e.g. Paradis, 1997, 2000; Pavlenko, 1999) argue for the necessity of making a clear distinction between the conceptual level and the semantic level. The

supporting evidence can be found in the examination of patients with aphasia, who are known to suffer from language loss but have their conceptual ability largely retained. As such, Paradis (1997) argued that a semantic component, which contains explicit linguistic information stored in the declarative memory, is vulnerable to aphasia, whereas a conceptual component, which incorporates non-linguistic multi-sensory information stored in the implicit memory, is not vulnerable to aphasia.

One could argue that the semantic and conceptual levels overlap to some extent but they are different in the sense that the former is shaped by and for social intentions whereas the latter hinges on individuals' physical or psychological interactions with the referent or the entity. To put it simply, the semantic level is more amodal than the lexical level but such a difference may be inherently subtle in experimental practice, or possibly not empirically observed (Francis, 2005). Taking both sides of the debate into account, it would be difficult to draw any confident conclusions from what has been known thus far. As such, this study tentatively labeled all the conceptual and semantic variables under the category named the Meaning level or the conceptual-semantic level.

The first stage of spoken production leads to *grammatical encoding*, in the course of which the selection of a lemma and its syntactic environment occurs. The mechanisms underlying this stage have been addressed in Roelofs' (1992) model in which lemma nodes are linked to the conceptual stratum, and that lemma selection is essentially a statistic process by which only the highest activated lemma is selected

for production. Upon completion of lemma selection, syntactic information becomes available for further grammatical encoding, and the syntactic environment for the target item is then constructed.

After having completed grammatical encoding, spoken production proceeds to the surface/form level, that is, *morpho-phonological encoding and syllabification*. This stage begins with the processing of three kinds of information to generate phonological scores, including '*morphological make-up*' (e.g. -dog, -ing, and, re-), '*metrical shape*' (e.g. stress, pitch, syllables), and '*segmental shape*' (e.g. consonants, vowels and diphthongs) (Levelt, Roelofs, & Meyer, 1999). It should be acknowledged that morphological make-up is typically involved in continuous speech production rather than in single word utterances. The next step is to realize phonological scores with articulatory gestures, which are usually performed at different tiers in the oral system (Levelt et al., 1999). After this, a large number of different muscles are activated to convert the articulatory score into overt speech, followed by self-monitoring, with which speakers are able to modify or repair what they have said or how they have said it (Levelt, 1989).

The existence of a lemma-lexeme distinction can be found in the study of those individuals whose grammatical knowledge remains relatively intact but phonological and orthographical knowledge are impaired. For instance, Badecker, Miozzo, and Zanuttini (1995) reported the case of an Italian anomic patient named Dante, who was able to identify the grammatical gender (i.e. feminine/masculine) of target words with

an accuracy rate of 98%, but was unable to name them accurately and make form-related judgments on them. Another source of supporting evidence comes from examination of the tip-of-tongue (TOT) state in which a speaker is unable to recall the word that he certainly knows (R. Brown & McNeill, 1966). As Vigliocco, Antonini, and Garrett (1997) explained, a speaker's TOT can be attributed to the failure of retrieving a lexeme during successful retrieval of its lemma.

However, a lemma/lexeme distinction has not been accepted by every researcher, and one exception involves Caramazza (1997) who argued against the necessity of separating a lemma from its lexeme, and instead supported an interconnected network incorporating different aspects of word knowledge. Caramazza's argument, nevertheless, has not received wide acceptance, in part because it lacks strong explanations to account for empirical observations of a lemma/lexeme distinction (see Roelofs, Meyer, & Levelt, 1998 for details).

Taken together, the present study assumes the existence of a distinction between the lemma and the lexeme, and this distinction is manifested in the way lexical variables are layered. The components and stages that spoken production includes have been reviewed above. In what follows, how these components and stages incorporate into a model will be examined.

2.3.1.2 Models

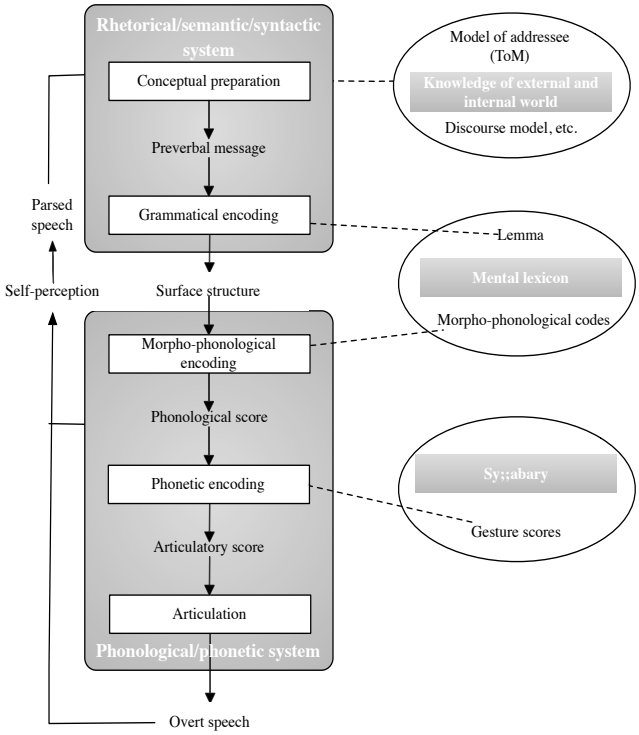
The models that describe the process of word production can be characterized as either serial or parallel. Serial models draw on Fodor's modularity of mind (1983),

which argues for information encapsulation (i.e. the operation of a module is unaffected by the information not contained in that module) and domain specificity (i.e. a module only processes certain kinds of information specific to that module itself). Examples of this type of module are the Fromkin-Garrett model (Fromkin, 1971; Garrett, 1980), the Blueprint of the Speaker (Levelt, 1989, 1995, 1999), and the WEAVER model (i.e. Word-form Encoding by Activation and VERification, Levelt et al., 1999; Roelofs, 1997). These models consider each processing component as 'a relatively autonomous specialist' (Levelt, 1989, p. 16) that flows top-down from one level to another and comes into play separately and independently of each processing component.

Drawing on an interactive view of languages, parallel models assume that multiple levels of processing take place simultaneously, and that different sources of information influence each other (D. W. Carroll, 2008). This is to say, the activation of a node can spread to reach other processing levels through between-node connections. Examples include those proposed by Dell (1986, 1988), MacKay (1982, 1987), Stemberger (1985), and Vigliocco and Hartsuiker (2005), but it should be acknowledged that models of this type have not been cited as widely as serial models in the field of language production (Kormos, 2006). It is beyond the scope of this chapter to describe all these models mentioned above; therefore, only the model that is most relevant to the present study, Levelt's model (1999), is reviewed below.

As shown in Figure 2.10, the model consists of two main components: the

rhetorical/semantic/syntactic system and the phonological and phonetic system. Word production proceeds through conceptual preparation, grammatical encoding, morpho-phonological encoding, phonetic encoding, articulation, and self-monitoring. Moreover, the model acknowledges the existence of three knowledge stores—knowledge of external and internal world, mental lexicon, and syllabary. A unique characteristic of this model is its incremental processing, which involves a large degree of relative independence and a small degree of simultaneity between/across processing components. As a consequence, in the words of Levelt (1989, p. 28), each stage of processing can be 'trigged into activity by only a minimal fragment of characteristic input' (i.e. *Wundt's principle*), and speakers are therefore able to produce 'fluent speech without unintended disruption'.



Notes:
Model of addressee: the context in which the speech event occurs;
ToM: theory of mind;
Discourse model: 'a speaker's record of what he believes to be shared knowledge about the content of the discourse as it evolved' (Levelt, 1989, p. 114);
Syllabary: automatized gestural scores used to produce syllables of a given language

Figure 2. 10 Levelt's model of speech production: A blueprint of the speaker (Levelt, 1999, p. 87)

How long it takes for each stage to complete in word production has been addressed by comparing chronometric data obtained in reaction time studies and in electrophysiological (EP) studies (Indefrey, 2011). Most of the EP studies used the evoked cortical responses such as N200 and lateralized readiness potential (LRP) to estimate the duration of different processing components. A large-scale meta-analysis was conducted by Indefrey and Levelt (2004) and recently updated by Indefrey (2011), with the results showing the time windows of word production components as

follows: ‘conceptual preparation 175 ms, lemma retrieval 75 ms, phonological code retrieval 80 ms, syllabification 125 ms, phonetic encoding 145 ms’ (p.3). These numbers give further support for the serial nature of word production; however, as rightly pointed out by Indefrey and Levelt (2004), particular caution should be taken when we interpret them in a rigid manner.

This section has discussed the components, stages and models of monolingual spoken production. The following section will pinpoint the differences between L1 and L2 production and then review those models that specifically address L2 lexical processing in language production.

2.3.2. Bilingual speakers’ production of single words

As mentioned earlier, the processing mechanisms underlying L1 and L2 production are thought to be similar in some sense but there are also some differences between them at least in the following respects. First, learners’ L2 knowledge is not as complete as their L1 knowledge; therefore, in some cases, L2 speakers have to use a variety of strategies, such as omission, code switching, and restructuring, to overcome communication problems that they may run into (for a review of these strategies, see Dörnyei & Scott, 1997). Second, the automaticity level in L2 production is not as high as that of L1 production (Poulisse, 1999), as evidenced by observations of slower speech and articulation rate, longer pauses, and shorter runs in L2 production (Wiese, 1984). Third, the two languages of a bilingual are competitively active in parallel even when speaking is restricted to one language alone.

Drawing on the monolingual language production process (Levelt, 1989; Levelt et al., 1999), Poulisse and Bongaerts (1994) and Hermans (2000) proposed the Spreading Activation Model of bilingual spoken production (i.e. SAM). An illustrative example of a Spanish-English bilingual naming the object *car* in English is shown in Figure 2.11. The SAM consists of three levels of representations: the conceptual level, the lemma level, and the phonological level. These levels are activated in a cascaded manner, and both L1 and L2 lexicons compete for selection at each level. To be specific, at the conceptual level, both the conceptual cues related to the picture object and the language cue work together to activate corresponding semantic features. At the lemma level, lexical representations in both languages, car in English and coche in Spanish, become active, along with their semantically or grammatically neighboring words, such as truck in English and libro in Spanish. At the phonological level, phonological forms of both languages are activated but only the sound of the intended language is specified. At issue theoretically is how bilinguals opt for the appropriate language that maps onto the target concept when both the lexical and conceptual links are activated equally. This issue has been regarded as a 'hard problem' in the bilingualism literature in the eyes of Finkbeiner, Gollan, and Caramazza (2006). As will be reviewed shortly, several proposals have been put forward to solve this hard problem.

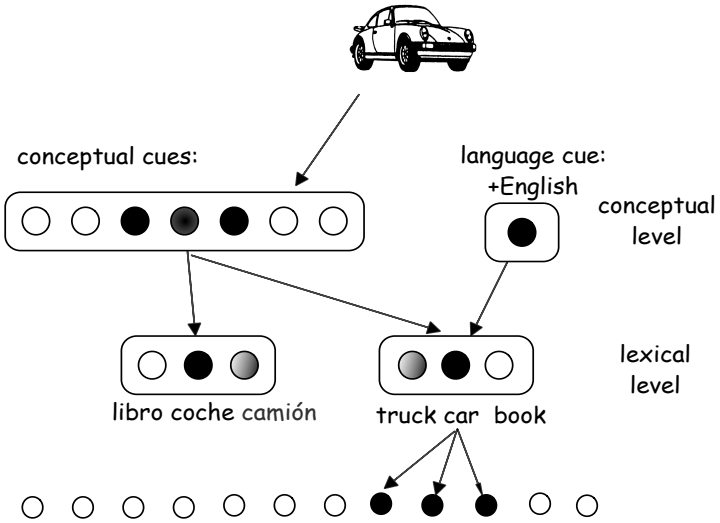


Figure 2. 11 Bilingual spoken word production: Spread Activation Model (adapted from Hermans, 2000; Poulisse & Bongaerts, 1994)

The first proposal pertains to the language-specific model advanced by Costa, Caramazza, and colleagues (e.g. Costa, 2005; Costa & Caramazza, 1999; Costa, Miozzo, & Caramazza, 1999), in which alternative candidates from both languages are active but bilinguals are able to direct their attention only to the selection of the target language item while ignoring the non-target one. Using the picture-word interference paradigm, Costa et al. (1999) showed that Catalan-Spanish bilinguals named pictures with translation distractors (e.g. the Catalan picture name taula 'table' with the Spanish distractor mesa 'table') faster than with semantically related or unrelated distractors (in either language), which is interpreted as evidence for language specific selection. This is because the common semantic representations shared by translation pairs are closely related. In this case, competition for selection should have resulted in significantly longer response times when the distractor is a translation equivalent relative to semantically related or unrelated distractors.

An illustrative model of language-specificity in selection is presented in Figure 2.12 (Costa, 2005). In this SRBS model, the rectangle represents the selection mechanism, whereby only the lexicon nodes of the target language are activated, rendering any activation of the non-target language nodes unrelated to the spoken production process. However, as pointed out by Finkbeiner, Gollan, et al. (2006), the language-specific selection mechanism fails to account for the 'language effect', that is, it took longer for participants to name pictures with unrelated target-language distractors relative to unrelated non-target-language distractors in the study by Costa and Caramazza (1999). One possible explanation is that lexical nodes of translation equivalents are activated to a lesser degree compared to same-language lexical nodes (Finkbeiner, Almeida, Janssen, & Caramazza, 2006).

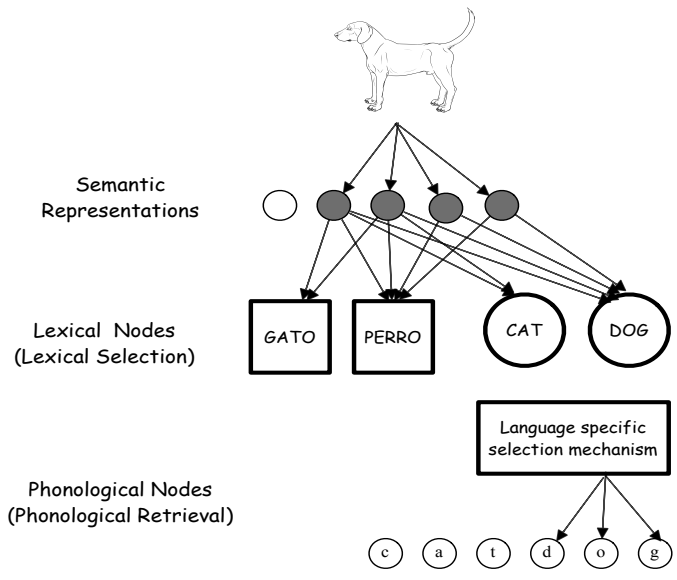


Figure 2. 12 Schematic Representation of Bilingual System (Costa, 2005, p. 314)

The Concept Selection Model (CSM) proposed by La Heij (2005) provides an alternative explanation as to how bilingual speakers end up selecting words from the

target language. According to this model, bilinguals are able to use a set of linguistic cues, particularly those cues that specify the target language. The activation of a language cue at the conceptual level or during the pre-verbalization suffices to produce words in the intended language, making lexical selection easy without any additional activation or inhibition needed at the lexical level.

Figure 2.13 provides an example of the mechanisms underlying the production of *dog* by English learners of Spanish. First, the preverbal message corresponding to the picture specifies the target concept and language. After this, the lexical nodes of the target language receive more activation compared to those of the non-target language. In other words, only the English cues are considered in the selection process although Spanish cues are activated. Once the activated lexical nodes have extended to the phonological level, corresponding phonemes are activated and the correct candidate is produced. To put it simply, lexical selection occurs from the very start of spoken production (i.e. the conceptual level). When the CSM is specifically used to accommodate the performance of proficient bilingual speakers, the CSM seems to be consistent with the RHM in the sense that both advocate the presence of conceptual mediation in processing L2 mental lexicons.

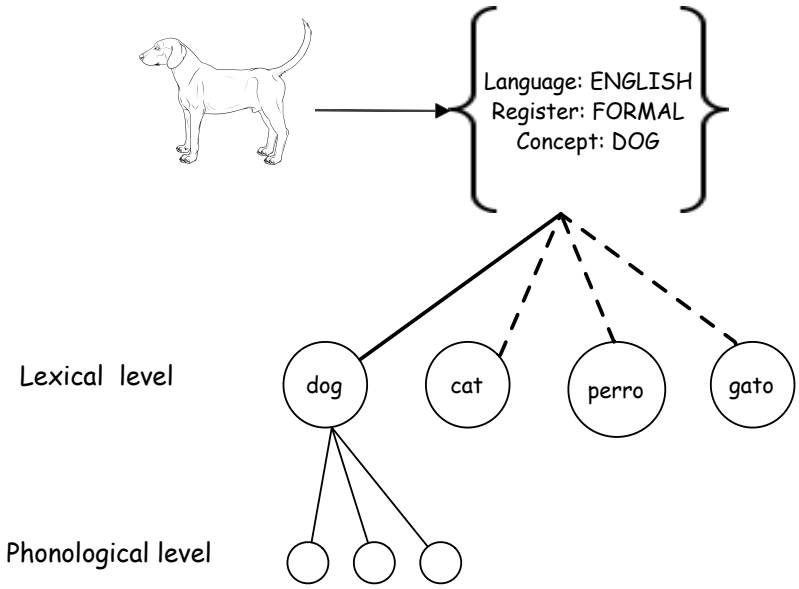


Figure 2. 13 Concept selection model (adapted from La Heij, 2005)

2.3.3 Interim summary

The models reviewed above imply that spoken production proceeds in different stages, and each stage involves different aspects of lexical characteristics. Moreover, both languages of a bilingual are switched on in L2 word production although there are disagreements regarding to which level of linguistic representations the non-target language can extend and regarding what mechanisms underlie the selection of correct words in the target language. Consistent with most previous models, the present study also assumes the presence of L1 activation in accessing L2 lexical items. Additionally, the studies reported in this thesis attempt to extend previous ones by modeling the roles of lexical characteristics and age in determining the degree to which bilingual speakers can activate the L1, since not all words are processed in an identical way and not all learners produce L2 words based on the same mechanism. The following sections will review the research literature on the effects of lexical characteristics on

the monolingual and bilingual performance of language activities and present age-related differences in relation to lexical processing.

2.4 What is in a word?

Words have their own conceptual representations, semantic properties, and physical forms, with which we are able to verbalize the contents derived from the intrinsic intentionality of our mental states. Three aspects of lexical characteristics—meaning, form, and usage—will be reviewed below. The following lexical variables were selected because they correspond to different stages of spoken production and have been found to affect the processing efficiency in language comprehension and /or production in the research literature.

2.4.1 Meaning level2.4.1.1 Word concreteness

Most words can be labeled as concrete or abstract, depending on whether or not their referents can exist independently of the physical world. Many studies have supported that concrete words exhibit preferential processing relative to abstract words (i.e. *concreteness effect*). Two prominent explanations have been advanced to account for this effect: one is Paivio's (1986) *Dual-coding Theory* and the other is the *Context Availability Theory* (Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983).

To be specific, the dual-coding theory postulates that there are two structurally and functionally independent, though interconnected, verbal representational systems. In addition to the verbal systems, a third system exists—the 'image system', which can be accessed by concrete words but not by abstract words (see Figure 2.14). This

difference likely explains why the priming effect was more salient for concrete than for abstract L2 words in a study by Jin (1990). However, as Jin pointed out, 'it may make more sense to think of the present concreteness effect as more conceptually rather than as visually based' (p. 1150). Moreover, in a recent study by Connell and Lynott (2012), perceptual strength norms in five separate modalities (i.e. visual, auditory, gustatory, olfactory, and tactile) contributed significantly to explaining the variance in word concreteness. This result indicates that word imageability alone may not sufficiently reflect a word's degree of concreteness.

The context availability theory postulates that concrete words benefit from the easier and richer availability of contextual information. This is because the conceptual representations of concrete words are more unitary as compared to those of abstract words. Empirical support for this proposal comes from a series of research by Schwanenflugel and colleagues. For instance, in Schwanenflugel and Shoben's (1983) study, participants took longer to read abstract words than concrete words without a paragraph context, but reading times was not significantly different between these two types of words when the context was provided. Additionally, the superiority of concrete over abstract words was retained in making lexical decisions without context. Regression analyses showed that context availability estimates served as a significant predictor of reading times and lexical decision times. Based on these findings, the authors argued that the context availability theory is superior to the dual-representation model because the latter has not adequately addressed the

interaction between word concreteness and its contextual representations.

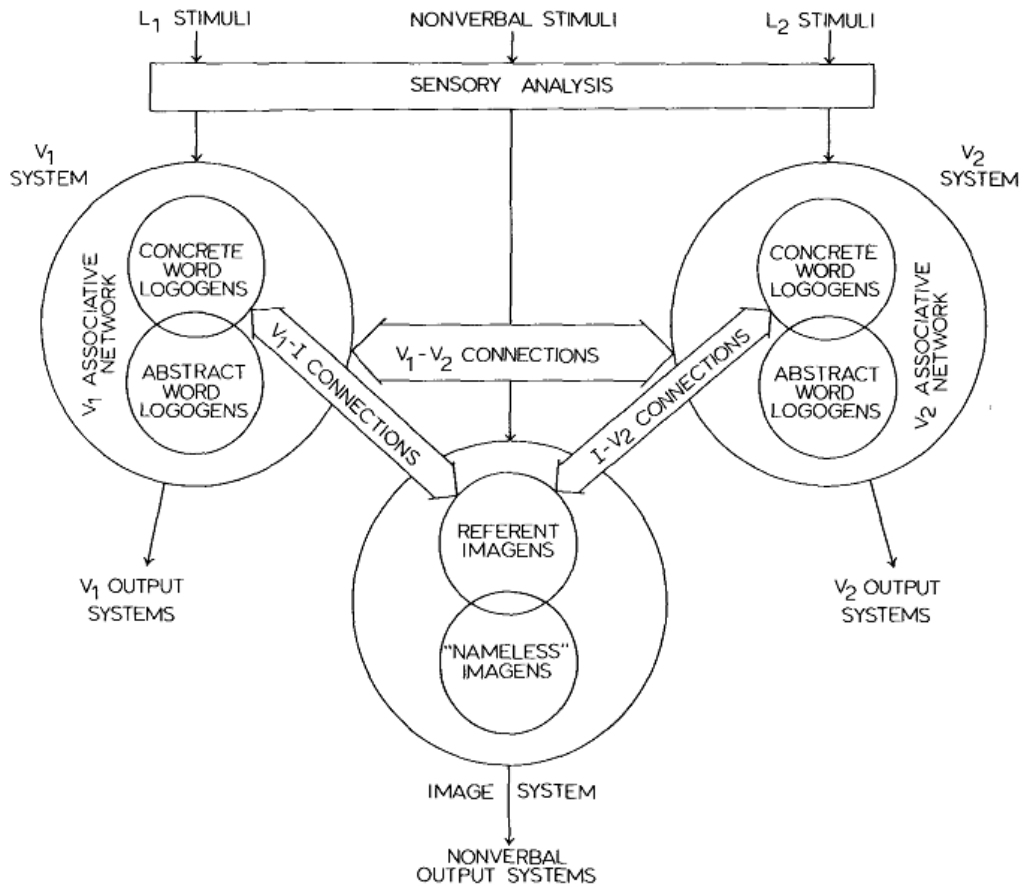


Figure 2. 14 Bilingual version of the dual-coding system (Paivio & Desrochers, 1980, p. 391)

Children learn concrete words usually earlier than abstract words, in large part because the referents that concrete words represent are easier to arouse sensory experience than ones which involve internal and subjective experience (Gardner, 1974). The superiority of concrete over abstract words for lexical decision was observed with third and fifth-grade monolingual children (Schwanenflugel & Akin, 1994). In a study by McFalls, Schwanenflugel, and Stahl (1996), second-grade monolingual children read concrete words with more accuracy relative to abstract

words in lexical decision and word naming but in neither task did word concreteness influence response times. Apart from these results, to the best of my knowledge, what role this lexical variable plays in children's L2 lexical processing remains to be specified. As for adults, concrete words were preferentially processed in a variety of tasks, for example, in normal translation, cued translation, and translation recognition (e.g. De Groot, 1992b), in lexical decision (e.g. Jin, 1990), and in serial recall (e.g. Tse & Altarriba, 2009).

A brief description of selected studies examining word concreteness is presented in Table 2.2. Additionally, the magnitude of concreteness effects was found to be larger for high-frequency words than for low-frequency ones (De Groot, 1992b), indicating that conceptual processing is more likely to occur among high-frequency words.

Table 2. 2 Selected literature examining word concreteness effects

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Schwanenflugel et al. (1988)	To investigate whether imageability had an effect on the recall of concrete and abstract words, independent of the accessibility of contextual information	Orienting-task condition Free recall	16 lists of 14 words	80 university students	An effect of imagery, which was independent of prior contextual knowledge, was observed, because reading times and lexical decision times were longer for abstract word than for concrete words without a context.
Jin (1990)	To examine whether the effect of inter-lingual semantic priming varied with concreteness levels	Lexical decision	Korean and English words	Korean-English bilinguals	Inter-lingual semantic priming effects were greater for concrete than abstract pairs but this effect was not observed within languages.
De Groot (1992b)	To identify determinants of word translation	Normal and cued translation Translation recognition	96 Dutch nouns for translation 192 word pairs for recognition	48 English-Dutch bilinguals	Four variables accounted for unique variance in translation: the stimulus-word frequency, the response-word frequency, cognate status, and context availability.
De Groot and Keijzer (2000)	To examine the effects of word concreteness on learning and retention of L2 words	Receptive and productive tests	60 words	40 Dutch university students	Concrete words were easier to learn and less susceptible to forgetting than abstract words. Receptive testing showed better recall results than productive testing.
Binder, Westbury, McKiernan, Possing, and Medler (2005)	To understand how the brain processed conceptual knowledge of concrete and abstract words using fMRI	Word/non-word lexical decision	200 stimuli English words and non-words (4-6 letters)	24 literate adults	A bilateral network of association and posterior multimodal cortices was activated during the processing of concrete concepts, and a strongly left-lateralized network was activated during the processing of abstract concepts.
Tse and Altarriba (2009)	To examine how word concreteness effects varied with the modulation of word valence (positive vs. negative) in immediate serial recall for emotion words	Immediate serial recall	5 lists of concrete/abstract words (positive and negative)	20 English-speaking participants	Concrete words were better recalled than abstract words. Positive words were better recalled than negative words.
Connell and Lynott (2012)	To examine whether concreteness effects were better predicted by concreteness/imageability or by strength of perceptual experience	Rating task Secondary data analysis	592 words	34 native English university students	Concreteness effects in lexical decision and naming were better predicted by perceptual strength ratings as compared to concreteness and imageability ratings.

2.4.1.2 Word typicality

The concepts that words represent may fall into the same category but can vary in terms of how typical they are of the category (Evans, 2010). This is to say, not all category members are considered to be equally 'good' examples or prototypes. For instance, as shown in Figure 2.15, a robin is a more typical bird than a penguin (Rosch & Mervis, 1975). Certain rules explain the presence of typicality observed in a category's internal structure (Rosch, 1973). A classic one has to do with the prototype theory, that is, the characteristics of a category are clustered around its exemplars, as the formula illustrates, $S(x, C) = \sum_i w(x, i)$ — the extent to which an instance x falls into the category C is defined in terms of the weight of attributes possessed by x relative to those by prototypes (Hampton, 1993).

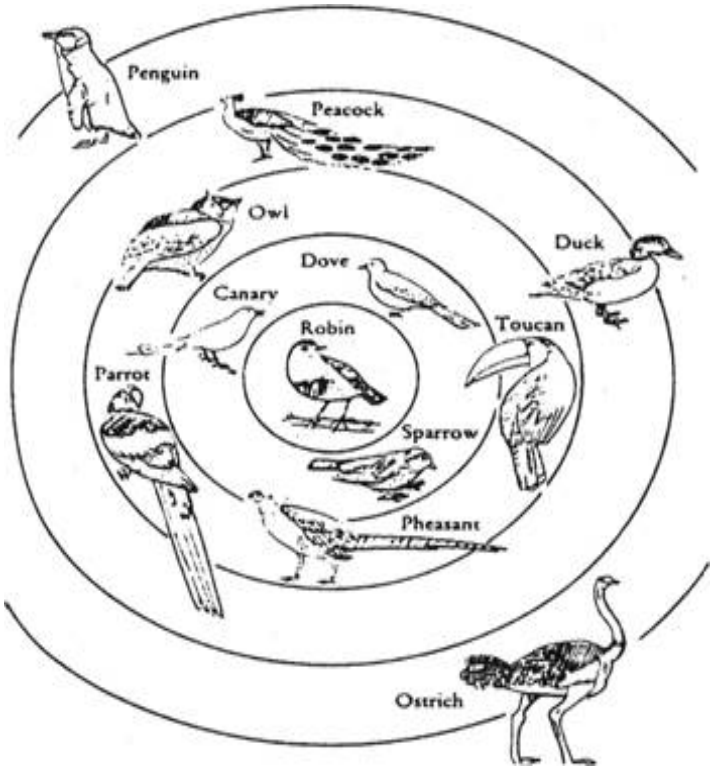


Figure 2. 15 The BIRD category (Aitchison, 2012, p.69)

The prototype theory stands in contrast to the classical componential view of semantic features, denying the possibility of defining a word by a single set of attributes, and instead arguing that meaning should be studied considering individuals' encyclopedic knowledge (Geeraerts, 2006). There is a large body of research that has provided evidence for the psychological validity of this theory (e.g. Mervis, Catlin, & Rosch, 1976; Rosch, 1973, 1975; Rosch & Mervis, 1975; Rosch, Simpson, & Miller, 1976). For instance, Rosch (1973) examined the internal structure of perceptual and semantic categories by asking participants to rate how typical a word's concept represented its category on scales. The estimates demonstrated that membership in a category is graded rather than all-or-none, indicating the boundary between category members is not completely clear-cut. In another study (Rosch & Mervis, 1975), one experiment on artificial categories showed that concepts with a higher degree of within-category family resemblance were learned more rapidly and identified more rapidly after practice than those with a lower degree of resemblance.

Typicality judgments are premised on the ability to taxonomically categorize objects and entities. This ability becomes increasingly prevalent and reliable as children's age goes up to preteen years even though its development lags behind another type of categorization, that is, thematic grouping (Bjorklund, 2000; Fivush, 1989; K. Nelson, 1996). Typical items are usually processed faster and more accurately relative to atypical items in a range of tasks performed by monolingual children, for example, in speeded category verification (e.g. Jerger & Damian, 2005),

in within-category verification (e.g. Southgate & Meints, 2000), and in cued recall (e.g. Bjorklund & Thompson, 1983). As regards adults, mixed results, however, have been obtained, for instance, priming effects for typical targets were observed with older adults but not with younger adults (Mäntyla, 1993), and word typicality significantly predicted within-category variances in positive reaction times but not in negative reaction times (Larochelle & Pineau, 1994). What has yet to be examined is how this lexical variable impacts L2 lexical processing in child and adult learners, the results of which are critically important for us to further our understanding of its role in L2 activities. A brief description of the existing literature on word typicality and its effects on language processing is provided in Table 2.3.

Table 2. 3 Selected literature examining word typicality effects

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
McCloskey and Glucksberg (1978)	To examine whether category boundaries were fuzzy or well-defined	Category membership decision task (10-point rating scale)	540 English name pairs	64 college undergraduates (native English speakers)	Substantial between-subjects disagreements and within-subjects inconsistency were observed, supporting the presence of fuzzy boundary between categories.
Caramazza and Brones (1980)	To examine whether or not bilinguals shared semantic memory representations across languages	Category verification task	3 category concepts and 6 instances for each category	Spanish-English bilinguals aged 18 years old	No effects of language condition were observed. Strong typicality and semantic-distance effects were found but there were no interactions between these two factors.
Malt and Smith (1982)	To examine the relationship between word familiarity and word typicality	Rating task Property list	English words from eight categories	260 English-speaking undergraduates	Familiarity alone could not account for variations in word typicality. The estimates of word typicality might change depending on which category concepts belonged to.
Bjorklund and Thompson (1983)	To examine how the child- and adult-generated typicality norms impacted upon the L2 lexical processing by children	Cued recall	Two lists of English L2 words	192 English-speaking children (64 kindergarten, 64 3rd, and 64 6th grade)	Recall was greater with the typical than with the atypical items for the child-norm lists but not for the adult-norm lists. Children's understanding of how concepts represent categories is more cognitively meaningful.
Mäntyla (1993)	To examine whether an increased activation of knowledge structures impacted memory for future actions	Category fluency task	2 lists of 8 target items and 112 non-target items	16 old adults 16 young adults (native Swedish speakers)	Old adults showed priming effects for typical targets only but younger adults showed priming effects for atypical targets only.
Larochelle and Pineau (1994)	To identify significant predictors of within and cross-category variations in semantic verification latencies	Semantic verification	72 instances, paired with correct/ inappropriate superordinate words	22 participants (native French speakers, university students)	Typicality significantly impacted upon within-category positive reaction times but no significant effect on negative reaction times.
Southgate and Meints (2000)	To examine whether word typicality influenced within-category verification	Within-category verification	24 image pairs	41 18-month-old infants 13 24-month-old infants (native English speakers)	18 and 24-month infants looked at typicality images significantly longer after hearing the item named than atypical images.

Table 2. 3 Selected literature examining word typicality effects (Continued)

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Malt, Sloman, and Gennari (2003)	To examine whether learners of different languages named containers in the same way	Picture naming task	60 common container names	28 native speakers of English 50 native speakers of Chinese 51 native speakers of Spanish	Prototypes were understood universally but there were substantial boundary variations.
Jerger and Damian (2005)	To examine whether the degree of word typicality and relatedness impacted upon the speed and correctness of children's classification	Speeded category verification	Colored pictures of 32 English concrete nouns	120 English-speaking children (aged about 3 years to about 14 years) 12 English-speaking adults	Children reacted more quickly to typical objects and made fewer errors as compared to atypical objects. Strong out-of-category objects were responded to more rapidly and correctly as compared to related out-of-category objects.

2.4.1.3 Semantic neighborhood density

Words do not live in isolation but are surrounded by and connected to significantly different numbers of semantic neighbors. Some words likely elicit a large amount of semantic information whereas other words may elicit much less semantic information. By definition, semantic neighbors (SN) refer to those words that are semantically associated with the target word.

There are different ways of measuring the size of SN, for instance, it can be measured by having participants produce the first word that comes to mind under indirect test instructions and then summing the associates across participants (D. L. Nelson, Schreiber, & McEvoy, 1992). The size of a word's SN is usually classified as large, medium, and small according to the total number of associates generated by participants. Another way is to count the number of all words that occur within a specified radius in high dimensional semantic space and then to average the distances between the word and its closet 10 neighbors, a computational model which is constructed on the basis of global co-occurrences (Buchanan, Westbury, & Burgess, 2001). Words with a dense semantic neighborhood are said to have a great deal of semantic neighbors and low semantic distance; conversely, words with a sparse semantic neighborhood are said to have few semantic neighbors, and high semantic distance. In the present study, the neighborhood measure refers specifically to this co-occurrence approach based on the work of Buchanan et al.

The SN effects have been investigated predominantly on word recognition in monolingual adults. In most of these studies, words with large semantic neighborhoods have been found to be recognized faster than those with sparse semantic neighborhoods in lexical decision, word naming, and semantic categorization (e.g. Buchanan et al., 2001; Siakaluk, Buchanan, & Westbury, 2003; Yates, Locker, & Simpson, 2003), and in the semantic categorization and progressive de-masking task (e.g. Pexman, Hargreaves, Edwards, Henry, & Goodyear, 2007; Yates, 2012). Such facilitative effects, however, have not been noted in every piece of research, for instance, the study by Bormann (2011) showed that the latencies recorded in picture naming and in *picture-word interference* (PWI) naming were not affected by the size of semantic neighbors. A similar result was also obtained in a PWI study by Caramazza and Costa (2001).

Are children capable of organizing words into neighborhoods in a way similar to adults? The research by Storkel and Adlof (2009a) suggested that the discrete associates generated by monolingual preschool children were significantly correlated with those generated by monolingual adults, with quantitative differences being limited to 3 neighbors. This research result is in line with Murphy's (2004) view that the mechanisms underlying concept learning and representations appear to be similar across age cohorts. There is a paucity of studies investigating the SN effects in children. In the only relevant study, Storkel and Adlof (2009b) showed that there were no differences in naming and identifying the referents of the nonobject-nonword pairs

with large versus small numbers of SN before or during training but an accuracy superiority was observed with small set sizes one week after training. The authors pointed out that the criteria set against the overall accuracy might dilute the SN effect, but on the other side this result might indicate that it did take some time for known neighbors to affect the processing of newly encountered words.

What can account for the facilitative SN effect reported in the literature? One possibility is to consider the course of semantic processing, in which semantically similar words likely recruit similar nodes and connections between them, and a word's semantic representations are then strengthened (Siakaluk et al., 2003). As a consequence, a word with higher semantic richness tends to be produced more rapidly than those with lower semantic richness. As mentioned above, such facilitative SN effects have not been consistently observed across studies. This suggests another possibility that the activation of competitive neighbors makes target word selection more difficult and may result in a delay in naming (Bormann, 2011). A brief summary of the existing literature examining SN effects is provided in Table 2.4.

Table 2. 4 Selected literature examining semantic neighborhood effects

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Buchanan et al. (2001)	To determine the impact of semantic size and semantic distance on the participants' performance of lexical decision and word naming	Lexical decision Word naming	Experiments 1: 750 words (600 four-letter and 150 five-letter words) 300 nonwords Experiment 2: 135 words and 135 nonwords	Experiment 1: 150 first-year undergraduates (native speakers of English) Experiment 2: 51 first-year undergraduates (native speakers of English)	Semantic size and distance predicted the participants' performance of both lexical decision and word naming.
Siakaluk et al. (2003)	To identify the effect of semantic distance in three semantic categorization experiments.	Experiment 1: Yes/No semantic categorization Experiment 2 and 3: Go/No-go semantic categorization	Experiment 1 and 2: 40 words in each group (high semantic distance and low semantic distance) Experiment 3: 50 words in each group	Experiment 1, 2 and 3: 36 first-year undergraduates (native speakers of English)	Semantic distance exerted a large effect in the go/no-go semantic categorization task but not in yes/no version of the task.
Yates et al. (2003)	To determine the impact of neighborhood density on lexical decision latencies	Lexical decision	60 four-letter stimuli by crossing two variables of letter string type and semantic neighborhood size	58 undergraduates (native speakers of English)	Lexical decisions on words with sparse semantic neighborhoods were slower than those with high semantic neighborhoods. The reverse pattern held for pseudohomophones (e.g. nale).
Pexman et al. (2007) Siakaluk et al. (2003)	To examine the neural consequences using event-related functional magnetic resonance imaging	Semantic categorization	36 words with low numbers of associates 36 words with high numbers of associates	20 native speakers of English	Words with low numbers of associates showed greater activation than words with high numbers of associates.
Storkel and Adlof (2009a)	To determine whether semantic set size affected word learning	Naming and referent identification task	10 non-object words paired with 10 nonwords	36 pre-school children	No differences were found in naming and identifying the referents of the nonobject-nonword pairs with large versus small neighborhood size before or during training but accuracy superiority was observed with small set sizes one week after training.

Table 2. 4 Selected literature examining semantic neighborhood effects (Continued)

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Bormann (2011)	To determine whether semantic neighborhood size influences lexical selection	Experiment 2: Picture naming Experiment 3: Picture word interference	Experiment 2 and 3: 27 stimuli in each group (many neighbors and few neighbors)	Experiment 2: 18 native adult speakers of English Experiment 3: 24 native adult speakers of English	There was no main effect of a target item's semantic neighborhood size on naming latencies.
Yates (2012)	To identify the effect of semantic neighborhood on the performance of word identification	Experiment 1: Semantic categorization Experiment 2: progressive demasking task	Experiment 1 and 2: 40 experimental stimuli (large and small)	Experiment 1 and 2: 20 undergraduates (native speakers of English)	Semantic neighborhood facilitates response latencies in both semantic categorization and progressive demasking tasks.

2.4.1.4 Number of related senses

Many words have different but related senses derived through extensions of a core meaning by following certain rules (i.e. polysemy), for example, ‘cloud’ means both ‘a visible body of droplets’ and ‘something that obscures’. The degree of polysemy of a lexical item can be defined with reference to the number of lexical entries recorded in the Wordsmyth dictionary (Parks, Ray, & Bland, 1998). It should be noted that polysemy has to be distinguished from another type of semantic ambiguity, that is, homonymy, which carries two (or more) distinct and unrelated meanings, such as ‘bark1’ meaning ‘harsh sound made by a dog’ and ‘bark2’ meaning ‘the covering of a trunk’. Two criteria have been proposed and used to differentiate between polysemy and homonymy: one concerns the etymological derivation of words and the other relates to the degree of relatedness or unrelatedness of meaning (Lyons, 1977). When the lexical entries of a word are derived from distinct historical contexts and/or highly unrelated to each other in meaning, the word should be seen as homonymous.

In comparison to monosemous words, polysemous words exhibit preferential processing among monolingual adults in lexical decision (e.g. Beretta, Fiorentino, & Poeppel, 2005; Klepousniotou & Baum, 2007; Rodd, Gaskell, & Marslen-Wilson, 2002), in priming (e.g. Finkbeiner et al., 2004), and in word naming (e.g. Hino & Lupker, 1996; Lichacz, Herdman, Lefevre, & Baird, 1999). Most of these studies examining the NoRS (number of related senses) effects have focused on word identification and recognition. Most relevant to the present study is the work by

Lichacz et al. (1999), in which participants named 80 words in an on-line naming condition followed by a delayed naming condition, and polysemous words were named faster than monosemous words.

There are several plausible accounts for such an ambiguity advantage: (1) the *Recurrent Network Model* (Kawamoto, Farrar, & Kello, 1994) attributes the differential effects of NoRS to the strength of between orthographic and semantic components; (2) according to the attractor-basin model (Rodd et al., 2002), the shallow and broad basin formed by polysemous words seems to speed up access to a correct attractor but the opposite pattern holds for the deeper and narrower basin formed by monosemous words; and (3) drawing on the context availability theory (Schwanenflugel et al., 1988), the ambiguity advantage observed in polysemous words comes from the availability of its rich semantic and contextual information. Most research examining the NoRS effects has been targeted at monolingual adults, and not much in the literature has investigated this lexical variable in relation to bilingual lexical processing in children and adults. A brief summary of the existing literature examining NoRS effects is presented in Table 2.5.

Table 2. 5 Selected literature examining number-of-related-senses effects

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Klepousniotou (2002)	To examine whether three types of ambiguous words were accessed differently and stored differently in the mental lexicon	Lexical decision (real word or not)	40 ambiguous words 40 control words matched for type of ambiguity 40 words matched for frequency occurrence	45 adults aged 25-30 years old (native speakers of English)	Words with multiple senses showed shorter fixation times than words with multiple meanings. Ambiguous words with multiple related meanings were accessed faster than words with multiple unrelated meanings.
Rodd et al. (2002)	To examine the differential impacts of word senses and meanings on word recognition	Lexical decision (word or non-word)	124 ambiguous words 60 unambiguous words	25 members of the MRC Cognition and Brain Sciences Unit subject panel	Competition between multiple meanings of ambiguous words was found to delay recognition but words with many senses were found to facilitate recognition.
Finkbeiner et al. (2004)	To examine the priming effects in two directions (from many senses to fewer senses and from fewer senses to many senses)	Masked translation priming Lexical decision (real or not)	40 word pairs (synonymy, one with one sense and the other with multiple senses) 40 control words	44 undergraduates at University of Arizona (native speakers of English)	Robust priming was found in the many-to-few direction but no priming was observed in the few-to-many direction, a result giving support to the Sense model.
Rodd (2004)	To investigate how word ambiguity and consistency affected word naming	Reading word aloud	4 groups of 33 monosyllabic words	22 Cambridge students	There was no ambiguity effect for the consistent words but the inconsistent words showed a sense benefit.
Beretta et al. (2005)	To examine the effects of semantic ambiguity on the response latency and accuracy, and on the peak latency	Lexical decision (real word or not)	128 lexical items	19 right-handed English-speaking volunteers aged 18–31 years	Words with multiple senses were responded to more quickly and accurately than with few senses. Words with more than one meaning were responded to more slowly and with approximately equal accuracy than with words with only one meaning. The behavioral results were correlated with the neural results.

Table 2. 5 Selected literature examining number-of-related-senses effects (Continued)

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Klepousniotou and Baum (2007)	To investigate whether sense-relatedness facilitated lexical processing and to clarify whether this advantage could be found for both types of polysemy (metaphor and metonymy)	Lexical word decision (word or non-word)	18 unbalanced homonymous words 18 balanced homonymous words 18 metaphorical words 18 metonymous words 18 unambiguous words 45 filler words	20 native speakers of English	Reaction times were faster for metaphorical and metonymous ambiguous words than unambiguous control words. No significant difference was found between homonymous words and unambiguous words.
Rodd et al. (2012)	To examine the impact of semantic relatedness on people's ability to learn new meanings	Cued recall task	36 target words	22 participants	The relatedness between old and new meanings facilitated the meaning-form mapping.

2.4.2 Form level

Word length can be measured orthographically (i.e. NoL: number of letters) or phonologically (i.e. NoP: number of phonemes and NoS: number of syllables). It is generally assumed that the more letters, phonemes, and syllables a word has, the longer it takes for language speakers to retrieve and process it. The presence of this effect has been reported in previous studies examining adults (e.g. Alario et al., 2004; Cuetos, Ellis, & Alvarez, 1999; De Groot et al., 2002, see Table 2.6 for details), although the predictive ability of each specific measure of word length was not consistent across research contexts possibly due to their examination of different languages (Bates et al., 2003). For instance, Severens et al. (2005) found that the NoS of a Bulgarian word, but not its NoL and NoP, significantly predicted naming latencies. The study by Alario et al. (2004), on the contrary, showed that the NoP of a French word contributed significantly to naming times in the familiarization stage of picture naming, but such a predictive effect was not observed when word length was determined by the NoS. No matter which specific measures have predictive ability beyond others, the presence of a length effect leads to conclude that the segments involved in articulation planning are serially processed; this is to say, successive letters, phonemes, and syllables are generated in sequence rather than in parallel.

As for bilingual adults, the presence of word length effects were observed in a study by De Groot et al. (2002), in which 40 Dutch-English bilinguals performed lexical decision and another 40 bilinguals performed standard word naming in English.

The results suggested that word length affected the recognition components of English, and this length effect occurred only for low-frequency words. As the authors commented, these language effects might actually arise from proficiency effects.

The absence of a word length effect has been reported in some studies (e.g. Bonin et al., 2002; Hirsh et al., 2003) or the effect takes place in the opposite direction than expected (e.g. Alario et al., 2004). What these results suggest is that word length does not affect naming latencies to the same degree all the time, but only when some specific conditions are met. One possibility has to do with word frequency, which has been found to interact with the length effect. In the work by Weekes (1997), the effects of NoL on word naming was restricted to words of low frequency. Another possibility concerns the pattern of blocks, as Meyer, Roelofs, and Levelt (2003) reported that the length effect can only be found in an experiment with a blocked design but not with a mixed design, possibly due to the influence of strategic response deadlines. Clearly, it is hard to draw confident conclusions based on these inconsistent findings, and exactly for this reason the present study is intended to better understand which lexical variables may modify the length effect in what manner by performing a simultaneous test of a wide range of lexical predictors.

Most of the studies examining the length effect have involved adults only; children, in contrast, are a comparatively under-researched population. With such limited knowledge, it is difficult to make comparisons across age cohorts in terms of the role of word length in spoken production. To the best of my knowledge, the only

relevant study (D'Amico et al., 2001) compared L1 picture naming performed by 50 Italian undergraduates with that performed by 34 Italian children, the results of which showed no significant effects of the number of letters and syllables on naming times. Moreover, little research has thus far focused on bilingual learners, so the present study set out to address this gap by conducting a comparative study of L2 picture naming in children and in adults so as to investigate how word length behaves in each age cohort.

Different studies in the literature have reached different explanations as to the presence or absence of length effects. A widely accepted view is that if the retrieval of phonological segments and articulatory programs are essentially sequential, producing long words is assumed to be slower than producing shorter words. By contrast, if such processes are parallel in nature, latencies for producing long and short words should not be significantly different (D'Amico et al., 2001; Meyer et al., 2003). Given inconsistent results obtained in the literature, replications of these studies using different kinds of tasks and types of learners are needed before any definite conclusions are drawn. A brief summary of the existing literature examining word length effects is provided in Table 2.7.

Table 2. 6 Summary of multiple regression studies examining the contribution of form variables to the fluency of lexical processing

Studies (author/year)	Task	Participant	NoL	NoP	NoS
Alario et al. (2004)	Spoken picture naming (familiarization session)	46 university students Native speakers of French	–	√	n.s.
	Spoken picture naming (immediate naming)		–	n.s.	√[1]
Bonin et al. (2002)	Spoken picture naming	36 undergraduate students Native speakers of French	n.s.	n.s.	–
Cuetos et al. (1999)	Spoken picture naming	64 undergraduate students Native speakers of Spanish	–	n.s	√
Hirsh et al. (2003)	Spoken picture naming	40 university students Spanish-English bilinguals	–	n.s	–
	Spoken picture naming[2]	23 undergraduate students Native speakers of English	–	n.s	–
C. M. Morrison et al. (1992)	Object picture naming	20 undergraduate students Native speakers of English	–	√	–
	Semantic categorization		–	n.s	–
De Groot et al. (2002)	Lexical decision[2]	20 university students Dutch-English bilinguals	n.s	√	√
	Word naming (standard) [2]		n.s	n.s	√
	Word naming (delayed) [2]		n.s	n.s	n.s
Severens et al. (2005)	Timed spoken picture naming	40 native Belgian Dutch speakers	n.s	n.s	√
De Groot (1992b)	Normal translation[2]	47 university students Dutch-English bilinguals	n.s	–	–
D'Amico et al. (2001)	Spoken picture naming	34 Italian children (ages 5-6)	n.s	–	n.s
	Spoken picture naming	50 university students Native speakers of Italian	n.s	–	n.s
Weekes (1997)	Word naming	20 undergraduate students Native speakers of English	√[3]	–	–

Notes: √: significant effect; –: not available; n.s.: effect but not significant; NoL: number of letters; NoS: number of syllables; NoP: number of phonemes; [1]: this effect was in the unexpected direction (negative β coefficient); [2]: in this study, L2 was examined rather than L1; [3]: significant effects observed for low-frequency words and non-words but not for high-frequency words

Table 2. 7 Selected literature examining word length effects

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Baddeley, Thomson, and Buchanan (1975)	To compare the subjects' memory span for sets of long and short words of comparable frequency	Short-time serial recall	8 monosyllabic words 8 five-syllable words	8 university students (native speakers of English)	Lists containing short words were better recalled than lists containing long words.
De Groot (1992b)	To examine the effect of word frequency, imageability, context availability, cognate status, definition accuracy, length of stimulus words and translation words, and familiarity on normal translation.	Normal translation	458 Dutch nouns	47 undergraduate students	All the variables were correlated with the performance measures, but only four variables accounted for unique translation variance: frequency of stimulus words, the frequency of response words, cognate status, and context availability.
Cuetos et al. (1999)	To examine how the number of syllables and phonemes impacted upon Spanish undergraduates' object naming	Picture naming	140 line drawings of objects	64 undergraduates of Spanish	Number of syllables and number of phonemes were highly correlated with the speed of picture naming. The number of syllables made independent contribution to predicting the speed of picture naming.
Santiago, MacKay, Palma, and Rho (2000)	To examine onset complexity and the number of syllables on picture naming latencies	Object naming	112 pictures	30 participants	In the standard naming task, there was a main effect of onset complexity and of syllable numbers on the picture naming latencies. In the name-picture matching task, there was a main effect of onset complexity but no main effect of syllable numbers on the picture naming latencies.
Meyer et al. (2003)	To investigate whether naming latencies would be longer for objects with disyllabic names than for objects with monosyllabic names	Experiment 1: Object naming	Experiment 1: 15 pictures with disyllabic names 15 pictures with monosyllabic names	Dutch-speaking undergraduates	The effects of the number of syllables were significant in pure conditions but not in mixed conditions.

Table 2. 7 Selected literature examining word length effects (Continued)

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Alario and Ferrand (1999)	To assess the possible effects of nine lexical variables on picture naming latencies	Picture naming	388 pictures	46 French-speaking students	The results showed that visual complexity, image agreement, and name agreement were major determinants of naming speed. Frequency and age of acquisition made independent contribution to naming times. Number of phonemes did not contribute significantly to predicting naming latencies. Number of syllables contributed marginally to naming latencies.
Graves, Grabowski, Mehta, and Gordon (2007)	To examine the effect of word length on picture naming latencies	Picture naming	95 pictures of animals 101 pictures of tools	14 subjects	Word length did not significantly predict naming latencies.
Tehan and Tolan (2007)	To examine word length effects in long term serial and free recall	Serial recall Free recall	120 long words 120 short words	English-speaking undergraduates	Word length effects were found in the long-term serial and free recall tasks.

2.4.3 Usage level

2.4.3.1 Subjective word frequency

Among studies examining the role of lexical variables in comprehension and production, as pointed out by Balota, Pilotti, and Cortese (2001), no single variable has been studied so intensely as usage frequency. A large body of research showed that adults take less effort to access high-frequency words relative to low-frequency ones in performing different types of tasks such as lexical decision (e.g. Balota & Chumbley, 1984; Schilling, Rayner, & Chumbley, 1998), picture naming (e.g. Barry, Morrison, & Ellis, 1997; Oldfield & Wingfield, 1965), standard word naming (Forster & Chambers, 1973), and speeded word naming (Gerhand & Barry, 1999b). Despite the consistency with which frequency effects have been observed in these studies, some researchers have raised concerns regarding their use of objective frequency counts, partly because a text-based corpus may not provide a sufficiently accurate estimate of word usage in listening and speaking (Brysbaert & New, 2009), and because the actual frequency of a rare word is likely to be underestimated, and the actual frequency of a common word is likely to be overestimated (Desrochers & Thompson, 2009). Most importantly, objective estimates may not be representative of the frequency with which learners use L2 words in input-limited instructional contexts, since most corpora are established with specific reference to monolingual resources.

An alternative way of measuring word usage is based on subjective frequency, which is usually generated by having participants rate a large set of words on a Likert

scale, typically of 7 points with the end points labeled ‘rarely’ and ‘commonly used’. Subjective and objective word frequency have been found to be correlated in the study of L1 words, with Pearson correlation coefficients ranging between .46 and .81 (for a review, see Desrochers, Liceras, Fernández-Fuertes, & Thompson, 2010). Such a less-than-perfect correlation indicates the existence of discrepancies between these two variables. More importantly, Balota et al. (2004) found that subjective word frequency turned out to be more powerful in predicting both lexical decision and naming performance in adults above and beyond objective frequency of occurrences. Also, subjective estimates have been shown to capture more accurately the frequency difference among low-frequency words relative to objective counts (Gernsbacher, 1984; Gordon, 1985).

It should be noted that subjective frequency appears to overlap with the variable of word familiarity, as both essentially reflect learners’ experiences encountering and using a particular word. However, they are different from each other in the sense that making a judgment on word familiarity is likely to be under the influence of other lexical variables such as semantic and/or orthographic and phonological components. When this occurs, word familiarity may not actually measure what the researchers intend to measure (Balota et al., 2001). Taking all the above into consideration, the present study used subjective frequency in the present study, a variable which has been found to explain more variance in the adults’ lexical processing. Note that very few studies have investigated the effects of subjective word frequency in relation to

L2 children's productive performance; in this regard, the present study allows for gaining an understanding of how this lexical variable behaves in young learners.

Frequency effects, as extensively observed in the literature, can be partly explained by the *Three-stage Model* proposed by Kirsner (1994), that is, separate stages of production are all governed by the power law of practice whereby the time spent in producing words decreases as practice increases. More specifically, lexical knowledge is believed to be embodied in three procedures—input processes, output productions, and decision productions, each of which is subject to amounts of practice (Kirsner, Lalor, & Hird, 1993). The combination of these practice-based components leads to a composite value that largely determines the overall latency with which a word can be accessed (Kirsner et al., 1993). Another explanation relates to the *Episodic Model* (Goldinger, 1998; Hintzman, 1986), in which it assumes that memory is partly formed based on input experiences, and high-frequency words have left many traces in the mind as compared to low frequency words, inspiring 'abstract' echoes (p. 255). When we have experienced a word in many contexts, we will be able to appreciate its high-frequency status, syntactic functions, and connections with other words. The various contexts in which a word has been encountered and stored will make it possible for us to amass and use myriad routes to access that word. Details of the selected literature examining frequency effects are summarized in Table 2.8.

Table 2. 8 Selected literature examining word frequency effects

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
Forster and Chambers (1973)	To compare naming times and decision times between words and non-words, and between high frequency and low frequency words	Pronounce the letter sequence Word-nonword classification	15 low frequency words 15 high frequency words 15 unfamiliar words 30 nonwords	10 English-speaking subject in each group	The result showed that real words were named faster than nonwords. High frequency words were named faster than low frequency words. Naming times were positively correlated with lexical decision times, and such a correlation was not observed with nonwords.
Segui, Mehler, Frauenfelder, and Morton (1982)	To investigate the effect of objective word frequency on lexical decision latencies for open and closed class items	Lexical decision	55 words 44 nonwords	French university students	Significant frequency effects were identified for both open and closed class items. A negative correlation was found between the frequency of open and closed class items in French and lexical decision latencies.
Balota and Chumbley (1984)	To examine the impact of five lexical variables (instance dominance, category dominance, objective word frequency, word length in letters and word length in syllables) upon the performance of three types of word recognition tasks	Category verification Lexical decision Pronunciation	72 target words	20 undergraduates involved in each task	The effect of word frequency was insignificant in category verification task but significant in lexical decision and pronunciation task.
Gordon (1985)	To determine the relationship between frequency and lexical decision times across a wide range of frequencies	Word rating Lexical decision	High frequency words Medium frequency words Low frequency words	80 undergraduates (native speakers of English)	There was a good correlation between subjective and objective word frequencies. The subjective word frequency estimate accounted for more of the decision time variance for the low frequency words as compared to high frequency words.
Mutter and Hashtroudi (1987)	To examine the cognitive requirement for high and low frequency words in word recognition and lexical decision	Word recognition Lexical decision	High frequency words Low frequency words Distractors	English-speaking university undergraduates	The result suggested that low frequency words required more effort during presentation and were recognized better than high frequency words in both tasks.

Table 2. 8 Selected literature examining word frequency effects (Continued)

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
McRae, Jared, and Seidenberg (1990)	To examine the frequency effects in naming by using stimuli that varied in terms of frequency	Immediate naming Short delay naming Long delay naming	72 homophone pairs 72 rhyme pairs 72 word/pseudo homophone pairs	English-speaking students	This study showed significant frequency effect in immediate naming, smaller effects in the short delay condition, and no effects in the long delay condition.
Balota et al. (2001)	To estimate the relationship between objective word frequency and subjective word frequency	Rating task	2,938 monosyllabic words	574 undergraduates 90 healthy older adults 1,590 individuals (All are English native speakers)	The result showed that the log frequency and meaningfulness accounted for most of the variance in subjective frequency estimates.
Desrochers and Thompson (2009)	To examine the reliability and validity of subjective word frequency with regard to Spanish words.	Word rating task	330 Spanish words	102 native speakers of Spanish	An analysis of simple words and compounds showed the reliability of subjective word frequency. Subjective word frequency was highly correlated with objective word frequency.
Desrochers et al. (2010)	To examine the relationship between subjective word frequency and imageability	Word rating task	3,600 French nouns	Two groups of 72 young adults	Subjective frequency estimates were significantly correlated with imageability. Correlations with ratings drawn from subjective and objective word frequency were found to be positive and significant.

2.4.3.2 Age of acquisition

Usage is defined not only by word frequency but also by the age at which a particular word is acquired (i.e. AoA). AoA norms reflect a trajectory on which a distribution of the occurrence of a word can be identified (Zevin & Seidenberg, 2004), a lexical characteristic that cannot be captured in word frequency alone. Moreover, AoA is believed to influence many stages of lexical processing, whereas the effect of word frequency is largely restricted to the lexical level of name retrieval (see Dent, Johnston, & Humphreys, 2008 for a review). Meanwhile, there exists a high correlation between AoA and word frequency, as high-frequency words are usually acquired earlier than low-frequency words (e.g. Barry et al., 1997; C. M. Morrison et al., 1992; Zevin & Seidenberg, 2002). Beyond that, AoA interacts with word frequency in such a way that a more pronounced frequency effect was observed on items introduced later than on those introduced earlier (Barry et al., 1997; but see Cuetos et al., 2006).

AoA effects, that is, early-acquired words are processed more accurately and rapidly, have been observed with monolingual adults in picture naming (e.g. Bonin et al., 2002; J. B. Carroll & White, 1973; Cuetos et al., 1999; C. M. Morrison et al., 1992), in lexical decision (e.g. Brysbaert, Lange, & Wijnendaele, 2000; Cortese & Khanna, 2007; Gerhand & Barry, 1999a; Turner, Valentine, & Ellis, 1998), and in word naming (e.g. G. D. A. Brown & Watson, 1987; Gerhand & Barry, 1999b; C. M. Morrison & Ellis, 1995, 2000). The relative magnitude of these effects has been

compared by Brysbaert and Ghyselinck (2006) based on multi-task investigations, the result of which showed that AoA effects were numerically larger for word meaning than for lexical decision, with the largest effect noted in picture naming. These observations provide evidence for the involvement of semantic components when AoA effect comes into being (Brysbaert, Wijnendaele, & Deyne, 2000). As for bilingual adults, Hirsh et al. (2003) found a significant effect of L2 AoA on error rates and reaction times of L2 picture naming performed by a group of unbalanced Spanish-English bilinguals. In a study by Izura and Ellis (2004), 20 late Spanish-English bilinguals performed a translation judgment task in which participants were asked to decide whether or not pairs of words in the two languages were translation equivalents, with the result suggesting that L2 AoA has significant impacts upon decision latencies. Not much research has been conducted on children and one of the few that exist is D'Amico et al. (2001) who found that AoA rated by adults accounted for 16% of the variance in children's picture naming latency in the L1, and when adult-rated AoA excluded from the regression model, an objective AoA measure also made a unique contribution to it (1.3%, $p < .01$).

Most studies collected AoA norms by asking participants to rate when they first learned a given word; the validity of this measure, however, has been challenged because the rating process, in and of itself, appears to be influenced by other lexical variables such as word frequency, familiarity, and imageability (C. M. Morrison, Chappell, & Ellis, 1997). In contrast, an objective measure of AoA, based on an

extensive examination of children's vocabulary in development, is considered preferable to rated AoA estimates (C. M. Morrison et al., 1997). This is especially true when research participants are children since their performance of word naming and lexical decision was not sensitive to subjective AoA ratings (C. M. Morrison & Ellis, 2000). As such, the present study used an objective measure to generate AoA norms with specific reference to the course books the participants had used at school, given that their L2 exposure was almost exclusively limited to instructional settings.

The relationship between frequency effects and AoA effects has been discussed for years in the literature. Although some researchers failed to observe frequency effects once AoA was controlled for (C. M. Morrison & Ellis, 1995), it is now more widely accepted that both frequency and AoA affect lexical processing (for a review, see Johnston & Barry, 2006), and that an AoA effect consists of frequency-dependent and frequency-independent components (Brysbaert & Ghyselinck, 2006). Thus, these two variables were used simultaneously to represent the latent variable of Usage in the current study.

Finally, AoA effects are subject to those changes in phonological and semantic representations as learners' vocabulary develops. Early learned words are usually stored more phonologically holistic and more imageable than late learned words (G. D. A. Brown & Watson, 1987; Van Loon-Vervoorn, Van Ham, & Van der Koppen, 1988). In this regard, word usage is not assumed to exist independently of word meanings and forms but rather comes into effect partly through them. In addition,

Zevin and Seidenberg (2002, 2004) considered AoA to be a behavioral outcome variable, which is affected by many semantic and phonological variables that are related to learners' performance. The present study adopts this view so that AoA serves as a mediator between meaning-and-form variables and the performance of L2 lexical processing in the hypothesized models for analysis. The selected literature on AoA effects is summarized in Table 2.9.

Table 2. 9 Selected literature examining age-of-acquisition effects

Author (year)	Aim	Task(s)	Stimuli	Participants	Conclusion(s)
J. B. Carroll and White (1973)	To investigate how age of acquisition and word frequency affected picture naming latencies	Picture naming	103 line drawings of objects	50 subjects	Age of acquisition was found to be the major determinant of naming latencies, and word frequency was only incidentally associated with naming latencies.
C. M. Morrison et al. (1992)	To investigate the differential effects of age of acquisition, frequency and other lexical variables on the time taken to name objects	Object naming	58 line drawings of objects	20 undergraduates or postgraduates	Word length and age of acquisition exerted significant effects on the speed of object naming after the word frequency was controlled.
Cuetos et al. (1999)	To examine how age of acquisition affected object naming latencies	Picture naming	140 line drawings of objects	64 undergraduates of Spanish	The highest raw correlation with naming reaction times was age of acquisition. Age of acquisition also affected the speed of retrieving phonological forms of words.
C. M. Morrison and Ellis (2000)	To investigate the effects of an objective measure of age of acquisition and other lexical variables on picture naming latencies	Picture naming	220 object names	27 undergraduate students	Age of acquisition and word length had significant impacts upon the naming speed but word frequency and orthographic neighborhood size had less impact on naming latencies.
Hirsh et al. (2003)	To examine whether age of acquisition impacted upon spoken naming performance in a second language	Picture naming	87 line drawings	40 Spanish-English bilinguals	Regression analysis showed that L2 age of acquisition and the degree of imageability significantly predicted the time spent on picture naming. This study indicated that the age-of-acquisition effect might not be restricted in a critical-period-type fashion.
Cortese and Khanna (2007)	To assess the effect of age of acquisition and other lexical variables on the performance of word naming and lexical decision	Word naming Lexical decision	2,870 single syllable words	32 undergraduates	Age of acquisition significantly predicted naming and decision latencies above and beyond other predictors. Age of acquisition was found to be more related to lexical decision performance than to word naming performance.

2.4.4 Interim Summary

This section has elaborated on the lexical variables that are pertinent to the present study and summarized previous research examining their effects on L1 and/or L2 lexical processing. Altogether, preferential processing was found in words with higher levels of concreteness and typicality, with more semantic neighbors and related senses, with shorter lengths, and with higher usage. However, it should be noted that mixed results have been observed across almost all lexical variables. Thus, one has to be careful in making a direct comparison between studies since the contexts in which they are embedded are different particularly in terms of the lexical variables included. Under this scenario, the present study has therefore included lexical variables, as many and varied as possible. The following section will focus on identifying age-related differences in relation to lexical processing.

2.5 Lexical Processing in Child and Adult L2 Learners

The efficiency of L2 word production appears to be influenced by a variety of learner variables, which can primarily be broken down into two clusters: internal and external (Littlewood, 1984). Variables of the former cluster are directly related to how a learner acts and behaves in L2 learning such as proficiency level, cognitive ability, motivation, and language aptitude. The latter cluster involves those variables that are relevant to the overall learning environment such as the quantity and quality of L2 input, the types of teaching strategies employed, and teachers' language proficiency.

For example, the size of lexical knowledge has been found to affect monolingual children's production accuracy and fluency in non-word repetition in a study by Edwards, Beckman, and Munson (2004). According to Cummins' *Interdependence Hypothesis* (1991), the development of L1 literacy is likely to be in parallel with the acquisition of comparable skills in the L2 due to the common learner characteristics that underpin the acquisition of both languages. Among all the learner variables, age is another factor that we need to consider in this study.

The research literature abounds with studies that examine the sources of age-related effects and of end-state differences between early and late L2 acquisition. These factors that directly relate to lexical processing can be broken down into three respects: memory system, L1 entrenchment, and the process of acquisition.

2.5.1 Memory system

The role that the procedural memory system plays in language learning becomes increasingly diminished as learners' age increases; in other words, compared to the declarative memory, the procedural memory tends to be more affected by age (Ullman, 2001, 2005). Additionally, according to Paradis (2004), and Ullman (2001, 2005), the procedural memory serves to sustain the processing of linguistic structures (e.g. phonology, morphology, and syntax), whereas the declarative memory is implicated in the processing of explicit knowledge (e.g. vocabulary). As introduced in Section 2.2.1, there are two types of declarative memory: semantic and episodic memories. With increasing age, the episodic memory becomes less accessible, whereas access to the semantic memory is preserved or facilitated (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002) even though bilingual experience likley attenudates the decline in episodic memory to some extent (Schroeder & Marian,

2012). Because of such an age-related difference in memory, presumably, the mechanisms underlying children and adults' L2 lexical processing may not be the same.

2.5.2 L1 entrenchment

An interpretation of age-related differences relates to the competition between the L1 and the L2, or more precisely, the varying degree of L1 consolidation and entrenchment at different time points of learning (e.g. Birdsong & Molis, 2001; Flege et al., 1999; Johnson & Newport, 1989). Adults learn an L2 with a background of more entrenched L1, thus their neural system for embracing and processing new input is not as plastic as those of children (Hernandez, Li, & MacWhinney, 2005). Moreover, the two languages of early learners, particularly simultaneous balanced bilinguals, are likely to be more separated relative to those of late learners, with particular regard to local cortical processing maps for articulation, lexical forms, sensory mappings, and grammatical processes (Hernandez et al., 2005).

Following the logic of the arguments stated above, the likelihood of relying on cross-linguistic routes rather than conceptual-mediated routes to process L2 words increases with age, but this assumption does not seem to be consistent with what the *Developmental Hypothesis* postulates, that is, there exists a shift from lexical to conceptual processing of L2 words with increasing fluency (Kroll & De Groot, 1997). Moreover, some literature shows that the conceptual route can be accessible to emergent bilingual learners (e.g. Altarriba & Mathis, 1997; Chen & Leung, 1989;

Comesaña, Perea, Piñeiro, & Fraga, 2009; Comesaña, Soares, Sánchez‐Casas, & Lima, 2012). For instance, Chen and Leung (1989) investigated the mechanisms of L2 lexical processing among 2nd grade children who had studied English for 2 years. The result showed that the speed of translation was not significantly faster than that of picture naming, indicating that the L1 does not mediate child learners' L2 spoken production. Given the mixed conclusions reached in this line of research, a perspective to adjudicate between these alternative positions is not to address whether the lexical route or concept mediation becomes more valid with increasing age or fluency, but, rather, to what extent and in what circumstances these two processing mechanisms come into play and at what level(s) of processing they can reach.

It should be noted that, compared to lexical domains, the more risky areas for L1 entrenchment are phonology (e.g. Major, 1987) and syntax-lexicon interactions (e.g. R. M. DeKeyser, 2000). However, as MacWhinney (2012) pointed out, the downsides of entrenchment can be attenuated by the supportive resonance, which serves to 'reconfigure old neural territory' through new encoding dimensions (p. 214). In practice, refreshing and repeating previously covered information at different time points of learning is considered important, which makes 'graduated interval recall' (Pimsleur, 1967, p. 75) achievable.

Other often-cited consequences of L1 entrenchment include positive/negative transfer and the parasitic use of L1 word meanings for L2 words. As a matter of fact, negative transfer does not very often occur with regard to lexical items' meanings,

because most of the words have globally consistent meanings across different contexts. According to the RHM model (see Figure 2.6 above), L2 lexicons are apparently parasitic on L1 equivalents at the initial stage of L2 learning. After completing this stage, as MacWhinney (2012) pointed out, learners are able to move away from the parasitic use of the L1 through the strengthened route between L2 forms and concepts. This 'internalization' process likely weakens the forces of parasitism. Assuming that the L1 is deeply entrenched in adult learners, then one issue that deserves our serious consideration is whether or not and when the strength of conceptual mediation can catch up with that of word association. Answering this question would hold implications for the developmental trend of L2 lexical processing.

2.5.3 Acquisition process

Another difference pertains to what child and adult L2 learners have experienced in L2 classes. As rightly pointed out by Muñoz (2006), the learning process, in and of itself, seems to be more important than the age of onset. Likewise, R. DeKeyser and Larson-Hall (2005) questioned the generalizability of critical period effects to formal educational contexts due to the limited availability of class time and a heavy reliance on form-focused instructions. Importantly, the role that learning and teaching strategies play in L2 lexical representations should not be deemphasized, as they likely influence the way in which learners process words (Comesaña et al., 2009).

In the context of the present study, child L2 learning and teaching predominantly

rely on the use of extra-lingual strategies, that is to say, new words are taught through non-verbal aids such as pictures, objects, and physical contexts. This can be evidenced by the insertion of pictures or cartoon images in introducing new words in the participants' course book called *PEP Primary English Studies*. In addition, according to the *National English Curriculum Standards* (Ministry of Education, 2011), the aim of primary-level English teaching is to develop the students' communicative competence by engaging them in a variety of speaking and listening activities.

By contrast, adult L2 learners learn vocabulary mostly through the use of inter-lingual (i.e. translation equivalents) strategy. As Yang and Dai (2011) pointed out, many Chinese university students learn and recycle words in favor of memorizing books in which a target word is presented together with its translation equivalent. Also, new words in the adult participants' course book *New Standard College English* (Foreign Language Teaching and Research Press, 2008) are usually introduced through L1 equivalents. Presumably, the age-related difference in how words are learned in instructional contexts may influence the representations of the bilingual lexicon in the mind.

Despite acknowledging the existence of several age-related differences, it is important to note that there may be some similarities between child and adult L2 lexical processing due to the existence of universal characteristics that mental lexicons have across age groups, for instance, lexical memory is not information

encapsulated but accessible to more than one mental system (Ullman, 2001). Although idiosyncratic word knowledge is initially subserved by media temporal lobe structures, in particular the hippocampus, eventually they will spread to other neocortical areas (Ullman, 2004). What this suggests is that lexical memory is distributed per se, in which case child and adult L2 learners may not use the same route to access and process every word. In light of this, item analyses are considered necessary for examining whether or not and to what extent lexical items are processed differently in the current research.

To sum up, this section has reviewed differences and similarities with regard to lexical processing between child and adult L2 learners from several perspectives. One thing to note is that it is not the exact age of an L2 learner that determines child-adult differences but a complex set of cognitive, cross-linguistic, social and acquisition factors, and constant interactions between these factors, which are accompanied by a certain age, characterize age-related differences and have a persistent influence on L2 learners' end-state performance.

2.6 Summary: Chapter 2

Figure 2.16 below outlines the key issues that have been addressed so far. First, the characteristics of long-term memories were discussed, followed by a review of the structure of bilingual memories according to a chronological order of the development of relevant models: Pre-RHM period, RHM period, and post-RHM period. Next, the research literature that addresses the components, stages, and models of monolingual

spoken production was reviewed, and then the models of bilingual spoken production were presented. After this, the lexical variables of interest were each described, and what their effects on language processing were specified. This chapter concluded with a discussion of age-related differences and similarities in lexical processing between child and adult L2 learners.

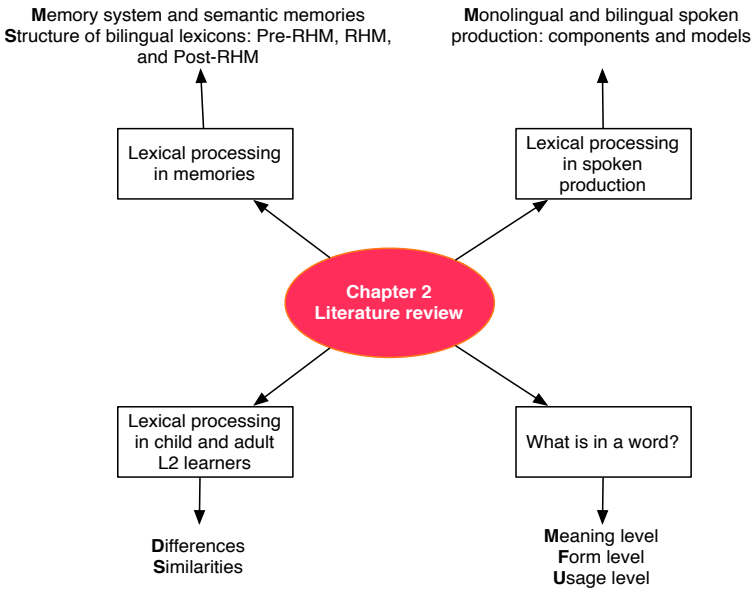


Figure 2. 16 Key issues reviewed in this chapter

Chapter 3 Research Methodology

This third chapter begins with an introduction of the research paradigm on which the present study hinges, followed by a description of the experimental and non-experimental methods used, and of the research design adopted. After this, characteristics of the sample will be described, and the research procedures detailed. The following sections will discuss the analytical strategies and statistical techniques that are employed to achieve the specific aims of this thesis. Finally, issues relevant to the ethical considerations, and validity, generalizability and reliability of this research will be addressed one by one.

3.1 Methodological rationale

There are two main research traditions in the field of applied linguistics—explanation and understanding, with each having their own ontological, epistemological and methodological 'dos' and 'don'ts' for research. The tradition of explanation, following the logic of natural sciences, claims that humans and society belong to the natural order (Hollis, 1994), and methods borrowed from natural sciences suffice for uncovering reality. In contrast, the tradition of understanding holds that the social world must be understood 'from within' rather than from the outside (Hollis, 1994), partly because human practice inevitably involves some mixture of people's 'subjective sense' (Winch, 2003) and 'collective intentionality' (Searle, 2010).

Within each tradition, different but interrelated approaches have been devised to explain and to understand language-related activities and practices. Each approach has

its own characteristics, advantages and disadvantages, and can be used to address different types of questions, and accordingly to generate different kinds of knowledge. In light of this, which approach to take is believed to hinge on the research topic in hand; in other words, the approach that best accommodates the posed research questions should be adopted. Following this logic, since the present research intends to compare conditions, to ascertain correlations, and to make predictions, positivism is deemed to be the most appropriate one that guides its investigations. To be more precise, this research is particularly situated within the stance of *modern positivism* developed by Popper (1959), who argues that science is 'a system of statements' (p.12) that can only be falsified based upon empirical evidence. This is to say, a theory cannot be proven to be correct, even for the time being. On the contrary, attempts should be made to disprove a theory by checking whether or not observations go against predictions. Provided that a theory withstands multiple severe tests, and is not replaced by another theory in the course of the accumulation of human knowledge, we may say that 'it has proved its mettle' (Popper, 1959, p. 10). Within this philosophical perspective, the present study has therefore drawn on *hypothesis-testing deductive* models, which usually start with testable hypotheses that are formulated based upon the literature reviewed, and end with a comparison of the predictions against the data collected.

Other features common to positivism, as summarized by Hacking (1981), can also be detected from the present study in the following ways: (1) the researcher's

values and interests had little influence on the specific research results (i.e. the separation of facts and values); (2) the generated numeric data were used to ascertain the strength of the relationship between the variables concerned, and to test predictions against observations (i.e. statements are measurable); and (3) experiments were conducted in identical ways according to a predetermined scheme (i.e. the use of standardized procedure).

Without doubt, positivism has long dominated science fields; however, this paradigm, as compared to interpretivism, has been criticized for oversimplifying the complex natural and social worlds. Nevertheless, as Kuhn (1962) argued, competing paradigms are inherently '*incommensurable*', because 'practicing in different worlds, the two groups of scientists see different things when they look from the same point in the same direction' (p. 150). It should be pointed out, however, that such a notion of incommensurability has not been favored by every philosopher of science, and one exception has to do with Laudan (1978), who argued that the superiority of a theory over another could actually be assessed by comparing their '*overall problem-solving effectiveness*' (p.5).

3.2 Research methods

Over the past few decades, an extensive collection of experimental paradigms and behavioral techniques has provided the foundation upon which contemporary theories and models of bilingual lexical processing are built. A survey of the literature suggests that response-latency (RL) or reaction-time (RT) based experiments,

comparatively speaking, are by far one of the most commonly-used research paradigms used to infer how lexical items are accessed, processed and structured in the bilingual mind, a process and a state that are themselves not directly observable and measureable. In what follows, specific research methods used in the present study are described in detail.

3.2.1 RL-based experiments

By definition, RL-based experiments refer to any investigations in which research questions can be answered by measuring and analyzing the time individuals take to perform a task or to respond to a stimulus (Jiang, 2012). There are at least four characteristics common to all different types of RL-based experiments, as summarized by Jiang (2012), that is, 'accurate timing, rigorous variable manipulation and control, time-sensitive assessment of behavior, and step-by-step progression' (p. 3). As to the present study, the first three characteristics are included. The reason why the fourth a step-by-step approach was not adopted here is that the main aim of this research was to carry out a simultaneous test of complex relationships between different characteristics of lexical items, as well as their predictive effects on the processing of L2 words.

Moreover, there are different types of tasks that have been used in RL-based research such as lexical decision, priming, and semantic categorization. Of particular interest to the present study is picture naming and forward translation.

3.2.1.1 Picture naming

Picture naming (PN) has been widely used in the field of psycholinguistics since Cattell's (1886) work, a task which serves as an elementary means to investigate the mechanisms underlying language in use (Glaser, 1992). PN is typically conducted in such a way that participants are presented with a series of pictures one by one, and as a picture appears on a computer screen, participants need to name it as rapidly and accurately as possible.

It is widely held that there are three main stages involved in the process of PN performed by monolingual speakers: (1) *object recognition and conceptual activation*; (2) *lexical selection and activation*; and (3) *response execution and production* (C. J. Johnson, Paivio, & Clark, 1996). To be specific, the first stage refers to the process by which participants need to identify the properties of an object, and then to match it with a corresponding representation in their long-term memory. The degree of difficulty in recognizing an object largely depends on the following factors: the speed of extracting relevant visual information, prior mental representations of the object, the nature of matching processes, and the degree of overlap between stimulus and mental representations. All these influences can be disentangled more in a conceptual sense but less at an empirical level (C. J. Johnson et al., 1996). It is estimated, in most cases, that it takes 175 ms for monolingual adults to complete the stage of visual processing and conceptual activation (Indefrey & Levelt, 2004, see Figure 3.1 below).

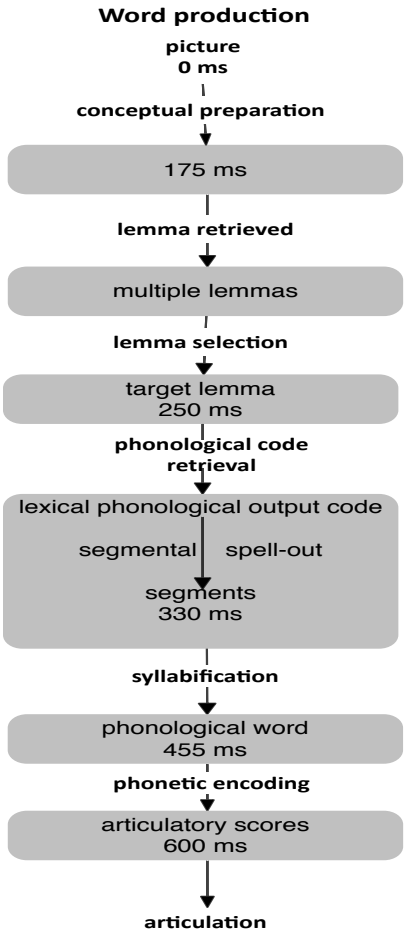


Figure 3. 1 Processing components involved in picture naming (Indefrey & Levelt, 2004, p. 109)

In the second stage where lexical selection and activation occur, a potentially best lexical candidate that matches with the activated concept or exceeds a threshold level of activation is about to be switched on. It is estimated that it averagely takes 175 ms for monolingual speakers to select and to activate a lexical item corresponding to the concept that a picture represents (see Figure 3.1 above, Indefrey & Levelt, 2004). In the last stage, response execution and production involve phonological code retrieval, syllabification, phonetic encoding, and articulation, all of which averagely take 350

ms to complete by monolingual speakers (see Figure 3.1 above, Indefrey & Levelt, 2004).

Added together, the whole process of PN averagely takes approximately 600 ms to complete. It should be noted that such an estimate has been derived from the performance of monolingual speakers; therefore, its implications for the performance of bilingual speakers are limited, especially in terms of L2 word production. This is because monolingual speakers almost entirely hinge on conceptual routes to reach lexical items. By contrast, as an alternative to the conceptual mediation (Potter et al., 1984, see Figure 3.2a), bilingual speakers are also able to rely on lexical routes between the two languages to produce L2 words (Potter et al., 1984, see Figure 3.2b). In comparison, more processing time is required to go through the route of word association relative to that of conceptual mediation.

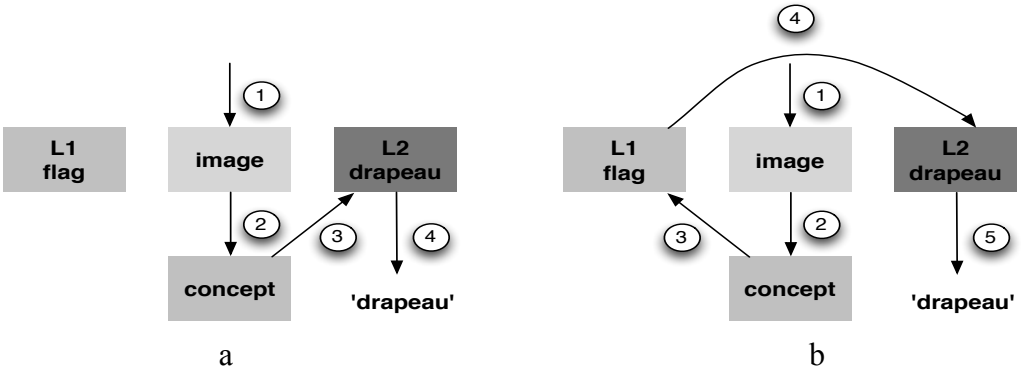


Figure 3. 2 Pathways of conceptual mediation and word association in picture naming (Potter et al., 1984, p. 25)

As regards the field of L2 studies, PN has been widely used to infer the structure of bilingual lexicons (e.g. Chen, 1990; Kroll & Stewart, 1994; Potter et al., 1984), to examine the directional asymmetry of switching costs between the L1 and the L2 (e.g.

Costa & Santesteban, 2004; Meuter & Allport, 1999), to investigate the effect of L2 learning on L1 processing (e.g. Ivanova & Costa, 2008), and to assess the role of certain lexical variables that might play in bilingual lexical processing (e.g. Hirsh et al., 2003).

3.2.1.2 Forward translation

Except for the early work of Cattell (1886) and that of Lambert, Havelka, and Gardner (1959), a rather recent addition to the pool of RL-based experiments is translation (Jiang, 2012). In performing a translation task, participants are asked to translate a visually presented stimulus word into another language as rapidly and accurately as possible. There exist two directions of translation: one is *forward translation* (FT) from the L1 to the L2; and the other is *backward translation* (BT) from the L2 to the L1. On average, it takes longer to complete FT than BT (Kroll & Stewart, 1994). FT is the only translation task used in the present study so a brief description will be given to it.

In general, FT proceeds through similar stages as PN, except that FT starts with visual word recognition, whereas PN begins with object recognition. Such a qualitative difference, as Snodgrass (1993) noted, was considered to exist between the very early stage of PN and that of FT. Quantitatively speaking, however, the amount of time spent on visual word recognition and object recognition is estimated to be largely similar. This is because it averagely takes monolingual speakers 175 ms to complete the stage of conceptual preparation in PN (see Figure 3.1 above). Likewise,

it averagely takes them 175-250 ms to reach the conceptual level in performing FT (Indefrey & Levelt, 2004, see Figure 3.3 below). As to the present study, whether or not there existed a statistically significant difference between the encoding speed of pictures and that of Chinese characters was examined by conducting an RL-based experiment using the semantic categorization task (see Section 3.6.2 for details).

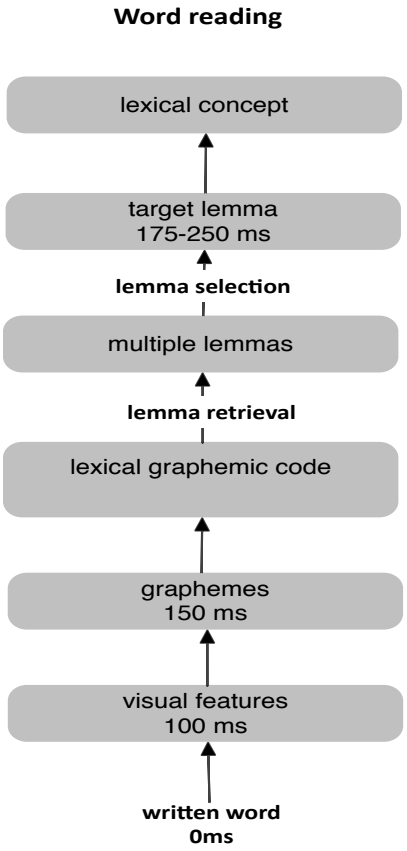


Figure 3. 3 Processing components involved in visual word recognition (Indefrey & Levelt, 2004, p. 109)

Similar to PN, there are two options available for bilingual speakers to perform FT: one is the conceptual mediation (Potter et al., 1984, see Figure 3.4a), and the other is the word association (Potter et al., 1984, see Figure 3.4b). Comparatively

speaking, more processing time is needed to go through the route of conceptual mediation relative to that of word association.

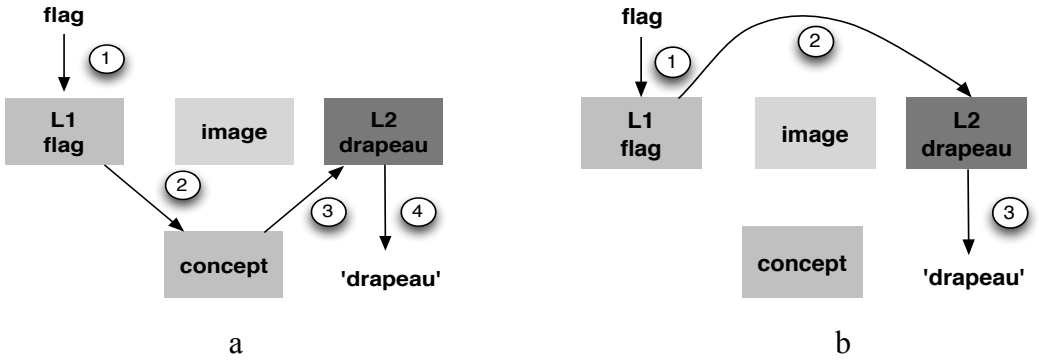


Figure 3. 4 Pathways of conceptual mediation and word association in forward translation (Potter et al., 1984, p. 25)

As regards the field of L2 studies, FT has been widely used as a means to investigate the nature of bilingual lexicons, and to examine the process of bilingual lexical processing. For instance, Potter et al. (1984) studied the nature of bilingual lexicons through comparing bilinguals' performance between PN and FT, De Groot (1992b) assessed the relative contributions each lexical variable made to the RL in both FT and BT, and Sholl et al. (1995) used translation tasks and PN to reveal the different routes that forward and backward translation might take.

3.2.2 Questionnaires and assessments

Demographic and language background questionnaires were used in the present study to collect learner-related information such as their age of starting L2 learning, the average amount of L2 exposure per week, and their academic achievements. Such

information served to profile participant characteristics as well as to assess their predictive effects on the fluency of L2 lexical processing. Moreover, participant-rated questionnaires were employed to collect judgments on a variety of lexical features, including word concreteness, word typicality, and subjective word frequency, the results of which could be used to assess and to compare the specific roles that each lexical variable likely played in the processing of L2 lexicons as well as in predicting the tendency toward a certain pattern of L2 lexical processing. Together with the delivery of RL-based experiments, and with the administration of questionnaires, several assessments were given to the participants (see Section 3.6.3 for details).

In what follows, the specific research design adopted in the present study is described in detail.

3.3 Research questions

The present study intends to address the following specific questions. It should be pointed out that the first research question, although not directly related to the focus of this study, paved the way for the second one. It was therefore considered necessary to investigate the relationships between four meaning-related variables (i.e. word concreteness, word typicality, semantic neighborhood density, and the number of related senses), three form-related variables (i.e. the number of letters, phonemes, and syllables), and two usage-related variables (i.e. subjective usage frequency and age of acquisition) of L2 words before their predictive effects on processing efficiency was examined.

RQ1. Do the meaning-related and form-related variables predict the reported usage of L2 words?

RQ2. Do learner-related and word feature-related variables predict the speed of L2 lexical processing in picture naming and forward translation?

a. Do learner-related variables predict the speed of picture naming and forward translation?
b. Do word feature-related variables predict the speed of picture naming and forward translation?

RQ3. What is the general pattern of L2 lexical processing among child and adult L2 learners?

a. What is the general pattern of L2 lexical processing among adult learners with very limited exposure to L2 (less than 20 hours per week)?
b. What is the general pattern of L2 lexical processing among adult learners with a certain amount of exposure to L2 (more than 20 hours per week)?
c. What is the general pattern of L2 lexical processing among child learners?

RQ4. Are there any variables that may predict the tendency toward a certain pattern of L2 lexical processing among child and adult L2 learners?

a. Do learner-related variables predict the tendency toward a certain pattern of L2 lexical processing among child and adult learners?
b. Do word feature-related variables predict the tendency toward a certain pattern of L2 lexical processing among child and adult learners?

How each research question was answered is summarized in Table 3.1.

Table 3. 1 Overview of research questions, data sources and analyses

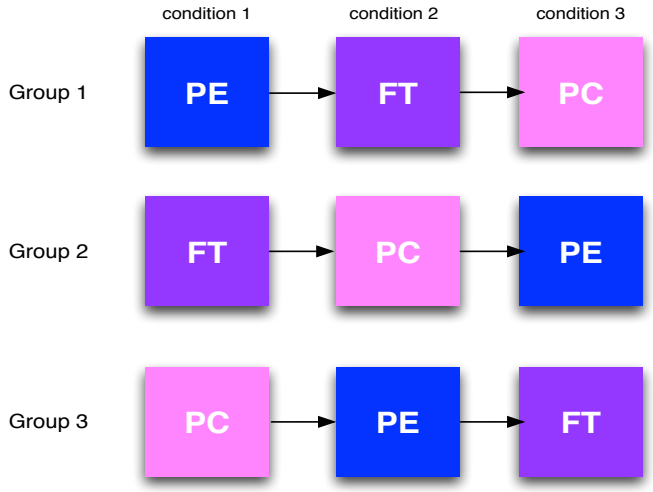
Research questions	IVs or Predictors and DVs or Outcomes		Methods of data collection	Methods of analysis
RQ1. *Do the meaning-related and form-related variables predicted the reported usage of L2 words?*	Predictors	Meaning and Form variables	Participant-rated questionnaires	CFA
	Outcomes	Usage-related variables	Psycholinguistic databases	SEM
RQ2. *Do learner-related and word feature-related variables predict the speed of L2 lexical processing in picture naming and forward translation?*	Predictors	Learner-related variables (e.g. age, gender, English proficiency level)	Demographic and language background questionnaires Student progress files	MR
		Word feature-related variables (e.g. WC, NoL, SWF)	Participant-rated questionnaires Psycholinguistic databases	SEM
	Outcomes	RLs and RAs	Computer-based experiments	
RQ3. *What is the general pattern of L2 lexical processing among child and adult L2 learners?*	IVs	Stimulus conditions Sequence of stimulus presentation Blocks of stimulus	Not applicable	ANOVA (F1 and F2)
	DVs	RLs and RAs	Computer-based experiments	
RQ4. *Are there any variables that may predict the tendency toward a certain pattern of L2 lexical processing among child and adult L2 learners?*	Predictors	Learner-related variables (e.g. age, gender, English proficiency level)	Demographic and language background questionnaires Student progress files	MR
		Word feature-related variables (e.g. WC, NoL, SWF)	Participant-rated questionnaires Psycholinguistic databases	MR
	Outcomes	RL differences between PE and FT	Computer-based experiments	

Notes: ANOVA: analysis of variance; CFA: confirmatory factor analysis; F1: by participant analysis; F2: by item analysis; MR: multiple regression; SEM: structural equation modeling.

3.4 Research design

Experiments, as a typical form of positivist investigations, were used in the present study to examine differences and relationships between the variables concerned. Experiments can be conducted in different ways depending on how participants are allocated to conditions (randomized or non-randomized), on how independent variables are manipulated (within-subjects or between-subjects), as well as on how many types of intervention a participant receives (counterbalanced or non-counterbalanced).

The present study adopted a *within-subjects counterbalanced Latin square experimental design* (see Figure 3.5), in which the participants were randomly assigned to one of the three groups; and each participant received three different types of stimuli that were presented in a given order. To illustrate, the participants in Group 1 named pictures in the L2 first, followed by performing forward translation and then by naming pictures in the L1. The participants in Group 2 started with the FT task, and the participants in Group 3 started with L1 picture naming.



Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

Figure 3. 5 Latin square design adopted in the present study

In such a design, the randomization of participants into different conditions allowed for an explicit control of potentially confounding variables so as to ensure that any observed significant differences in the participants' performance could be attributed to the stimulus conditions rather than to participant-related variables. Moreover, a within-subjects design served to reduce the chance of error variance that stemmed from individual variability, as the performance of the same participant was more likely to be consistent across a series of different conditions relative to the performance of different participants (Ray, 2012). Lastly, each stimulus condition (PE, FT, or PC) appeared equally often in each ordinal position so as to control for *order effects*. It should be acknowledged that *the effects of sequence* were only partially controlled for due to the limited availability of samples, but these effects tended to be attenuated, if not eliminated, by the insertion of a *washout period* (i.e. practice trials

and break time) between stimulus conditions. Moreover, statistical tests were performed to identify whether there existed any significant interactions between the sequence of stimulus presentation and stimulus conditions, and the results suggested no significant sequence effects observed in the present study (see Sections 4.4 and 4.5 for details).

In order to attenuate the disadvantages associated the use of experiments, for example, a high degree of *artificiality*, a lack of *generalizability*, and low *construct validity* (Myers & Hansen, 2011), the following efforts were made in the current study: (1) the tasks used in the present study such as PN and FT closely resembled what learners actually did in L2 classrooms; (2) the sampled participants were considered to be a representative part of the population of interest as far as possible; and (3) independent variables (IVs) were selected carefully to match the construct being investigated.

As mentioned above, together with experiments, other instruments like demographic questionnaires, language background questionnaires, participant-rated questionnaires, and assessments were administered to the participants in the present study. An overview of the research design is provided overleaf, consisting of a pilot study phase, a preparation study phase, and a main study phase.

Table 3. 2 Overview of research design: pilot, preparation and main study phases

Pilot study phase

CHILD

Main Aims	Validated procedures and instruments
Participants	5 Chinese children at Oxford
Tasks	1. Named 20 pictures in English 2. Translated 20 Chinese words to English 3. Named 20 pictures in Chinese 4. Rated 60 English words on SWF, WC, and WT 5. Completed a language background questionnaire

ADULT

Main Aims	Validated procedures and instruments
Participants	9 Chinese postgraduates at Oxford
Tasks	1. Named 30 pictures in English 2. Translated 30 Chinese words to English 3. Named 30 pictures in Chinese 4. Rated 90 English words on SWF, WC, and WT 5. Completed a language background questionnaire

Preparation study phase

CHILD

Main Aims	1. Examined name agreement, image agreement, and image variability 2. Examined the speed and accuracy of picture encoding and L1 word encoding
Participants	28 Chinese children in China
Computer-based Tasks	1. Named 123 pictures in English 2. Semantic categorization judgment on 30 pictures and on 30 Chinese words

ADULT

Main Aims	1. Examined name agreement, image agreement, and image variability 2. Examined the speed and accuracy of picture encoding and L1 word encoding
Participants	28 Chinese undergraduates in China
Computer-based Tasks	1. Named 220 pictures in English 2. Semantic categorization judgment on 30 pictures and on 30 Chinese words

Table 3. 2 Overview of research design: pilot, preparation and main study phases (Continued)

Main study phase

CHILD

Participants	39 Chinese children in China	
Variables	Predictors or IVs	Meaning: WC, WT, SND, NoRS Form: NoL, NoP, NoS Usage: SWF, AoA Stimulus conditions, sequences, and blocks Learner-related variables
	Outcomes or DVs	Response latencies Response accuracies
Step 1: Assessed participants' general intellectual ability and vocabulary depth		
Task:	WASI, WAT	
Step 2: Compared performance across stimulus conditions Assessed participants' vocabulary breadth		
Tasks	Picture naming in English; Translating Chinese words into English; Picture naming in Chinese X_Lex	
Stimuli	Three blocks of pictures selected from 10 semantic categories	
Step 3: Rated some features of target words		
Rating tasks	WC, WT, SWF	
Step 4: Assessed participants' listening and reading Collected participants' experience learning English		
Tasks	YLE	

ADULT

Participants	94 Chinese undergraduates in China	
Variables	Predictors or IVs	Meaning: WC, WT, SND, NoRS Form: NoL, NoP, NoS Usage: SWF, AoA Stimulus conditions, sequences, and blocks Learner-related variables
	Outcomes or DVs	Response latencies Response accuracies
Step 1: Compared performance across stimulus conditions		
Tasks	Picture naming in English; Translating Chinese words in to English; Picture naming in Chinese	
Stimuli	Three blocks of pictures selected from 10 semantic categories	
Step 2: Rated some features of target words		
Rating tasks	WC, WT, SWF	
Step 3: Assessed participants' vocabulary breadth and depth Recorded participants' experience learning English		
Tasks	X_Lex, WAT	
	Demographic and language background questionnaires	

Notes: AoA: age of acquisition; DV: dependent variables; IV: independent variables; NoL: number of letters; NoP: number of phonemes; NoRS: number of related senses; NoS: number of syllables; SND: semantic neighborhood density; SWF: subjective word frequency; WASI: Wechsler abbreviated scale of intelligence; WAT: word associates test; WC: word concreteness; WT: word typicality; YLE: Cambridge Young Learners English (Starters level).

3.5 Population and sample

3.5.1 Population

A *population* refers to all of the people or cases that share the defined characteristics of interest. What defines a population, as Hartas (2010) pointed out, is not the size but the presence of those characteristics common to all its members. As to the present study, it involves two populations: a population of child L2 learners (aged 10-11) studying English at primary schools in Mainland China, and a population of adult L2 learners (aged 17-20) studying English at public 'Project 211' universities in Beijing in Mainland China. They were believed to be representative of learners studying an L2 in input-limited contexts. Both populations had Chinese as their native language and English as their second.

The child population consists of 5th grade pupils (aged 10-11 years) who have learned English for two and a half years. Since early 2001, English has become a compulsory course at the primary level in Mainland China (Ministry of Education Document, 2001). In most cases, English is introduced in about the 3rd grade, and the time allocated to it should not be less than 80 minutes per week. Child learners are expected to have an average size of 500 receptive words by the end of the 6th grade, according to the *National English Curriculum Standards* (Ministry of Education, 2011) .The child population under study came from Datong, a northern city in Shanxi province. The 5th grade population was targeted at, particularly because they had been learning English for 2.5 years and were table to perform the present experiments. Also,

they have not yet entered adolescence.

Another target population, adult L2 learners (aged 17-20 years), consists of two sub-groups in the present study: non-English majors and English majors. Non-English majors were undergraduates reading for a degree of Bachelor in natural sciences, humanities or social sciences. Before embarking on university studies, the adult population received high-school education in different geographical areas in Mainland China (see Table 5.2 below for details). In general, non-English majors make up the majority of university-level EFL learners in China, as compared to English majors. On average, non-English majors have learned English in instructional settings for at least 10 years, amounting to an estimated 1440 hours of classroom exposure to English (about 4 hours per week on average). This population is required to take compulsory English courses in the first two years at university. They are believed to have acquired at least 3500 words since they have successfully passed the university entrance examination (Ministry of Education, 2012).

As regards English majors, they were reading for a Bachelor degree in English language and literature. Before embarking on university studies, their amount of classroom exposure to English was deemed to be similar to that of non-English majors. But, once they start to study at university, this population is required to receive rigorous training in four language skills as well as to take courses on English literature, translation, and linguistics. The medium of instruction is considered to be English on most courses although non-native English speaking teachers may

occasionally switch to the L1. A survey of the degree programs offered at different universities showed that English majors are generally scheduled to receive at least 25 hours of English-mediated instructions per week. On average, their average proficiency levels are thought to be about 60 on Test for English Major Level-4 (TEM-4), suggesting that they have an average vocabulary size of approximately 5000 words according to the Syllabus for TEM-4 (2005).

In a word, the child and adult populations under current investigation had characteristics unique to each age group at least in terms of L2 proficiency, L2 learning experience, and the degree of L1 consolidation.

3.5.2 Sample, sample size, and sampling strategies used

A piece of rigorous research is premised not only on the appropriate methods used and on the well-planned design adopted but also on the adequacy of sample size and the suitability of sampling strategies employed.

3.5.2.1 Sample and sample size

By definition, a *sample* refers to a smaller group or a subset of the target population. In the real world, a sample instead of the entire population is used due to resource limitations in terms of time and space, as well as due to other practical considerations (Cohen, Manion, & Morrison, 2007). According to Cohen et al. (2007), the following five issues need to be thoroughly considered: '(1) the *sample size*; (2) the *representativeness and parameters of the sample*; (3) *access to the sample*; (4) the *sampling strategy* to be used; and (5) the *kind of research* that is being undertaken'

(p.143).

Sample size is important in terms of the confidence level we have in making inferences based upon the sampled data, as well as the power of a statistical test performed. The commonly-used strategies for determining the sample size is to conduct statistical power analyses before collecting the data (*a priori*) or/and after the data collected (*a posteriori*) (Dattalo, 2008).

The adequacy of the sample size, in the present study, was ensured by running a priori power analyses with the G-Power software (Faul, Erdfelder, Lang, & Buchner, 2007). For instance, in examining the picture codability and variability, the power analysis results suggested that a minimum of 34 participants (17 children and 17 adults) were needed to detect an estimated effect size of .60 with a power of .80 for a correlation analysis set at the .05 level. Another study comparing the RL of PE, FT and PC needed a minimum sample of 24 child/adult participants to detect an estimated power of .80 for mixed-model ANOVAs with three groups and with three levels of measurement, with a priori alpha of .05 and with an estimated effect size of .60. In addition to a priori analyses, whether or not the size of the sample met the specific assumptions that each statistical test requires was also checked and further confirmed in an a posteriori manner (see Chapter 4 and Chapter 5).

3.5.2.2 Sampling strategies used

By definition, *sampling* refers to the process by which a subset of a population of interest is selected and then included according to some criteria in a study (Daniel,

2011). Ideally, the sample should be representative of the distribution characteristics possessed by the whole population, as far as possible, because such a high degree of representativeness would most likely enable researchers to make general inferences from a given data set, and also to apply the findings to other cases beyond the sample studied.

There exist two basic sampling schemes—*probability sampling* and *non-probability sampling*. As to the former type, in principle, it is regarded by many researchers as gold standard methodological procedures, since it allows for minimizing selection bias and maximizing both internal and external validity of a research study (Dattalo, 2010). There exist several subtypes of probability sampling such as *simple random sampling*, *systematic sampling*, *stratified sampling*, *cluster sampling*, *stage sampling*, and *multi-phase sampling* (Cohen et al., 2007). In practice, probability sampling, however, is likely to be constrained by a set of legal, ethical, and regulatory concerns. In such a case, non-probability sampling serves as its alternatives or supplements, a type of sampling that is more practically feasible although it has its own weaknesses such as the existence of case selection bias, an underrepresentation or overrepresentation of the target population, and an underestimate of variability in the population (Daniel, 2011). There are different types of non-probability sampling, among which the most commonly used are *convenience sampling*, *quota sampling*, and *purposive sampling*. All in all, both probability and non-probability sampling strategies have their own advantages and disadvantage.

(Examples can be found in Table 3.3.)

Table 3. 3 Sampling strategies: Advantages and disadvantages (based on Breakwell, Hammond, & Fife-Schaw, 2000; Cohen et al., 2007; Daniel, 2011)

Sampling strategies		Definition	Main Advantages	Main Disadvantages
Probability sampling	Random sampling	Each individual in the population has an equal chance of being selected.	Highly representative of the target population	a) High data collection costs b) Difficult to access the whole population
	Systematic sampling	Every n^{th} person in the population is selected.	a) Highly representative of the target population b) Samples spread across the whole population	a) Violating the principle of independence b) Estimating variance is more complex than random sampling
	Stratified sampling	The population is divided into strata and the researcher randomly chooses samples from the strata	a) Highly representative of the target population b) Inference can be made within a stratum and comparisons across strata	a) Difficult to select stratification variables b) More expensive, time-consuming, and complicated c) Proportionate allocation is required
	Cluster sampling	Samples are selected in naturally occurring groupings	a) Analysis of individual clusters is permitted b) Increase operational efficiency	Lower precision as compared to simple random sampling
Non probability sampling	Convenience sampling	Samples are selected based on their availability	a) Easy to access the participants b) Least resource requirements	Difficult to generalize to the entire population
	Quota sampling	Non-probability equivalent of stratified sampling	Including members of different subpopulations	a) Knowing characteristics of the population b) Difficult to control over fieldwork
	Purposive sampling	Samples are chosen in light of certain purposes	a) Control over who is selected and unselected b) Less selection bias	a) Requiring knowledge about the population b) Underestimating hidden populations

Taking into account both theoretical and practical considerations, the present study employed a mixed use of cluster, convenience, and stratified sampling strategies for recruiting the adult participants who joined the main study. The sampling took place between the end of March and the early of April in 2013. In all, there are 112 '211 Project' colleges and universities (Ministry of Education, 2008) in Mainland China, which are geographically dispersed. Given this, it might not be possible, or indeed be too costly and time-consuming to generate a simple random sample. As a result, a geographical cluster was sampled to be included in this study, a particular area which is believed to be the higher education hub in China, and to which the researcher had access.

After this, stratified and convenient sampling strategies were employed so as to ensure the quality of sampling procedures as well as the adequacy of sample size (see Figure 3.6). More specifically, four departments and five student associations that met the sampling plan criteria were randomly identified out of four universities, and then contacted by the researcher. A letter of introduction and invitation (see Appendix 3.1) that outlined the purposes, procedures, and ethical issues of the present research was sent electronically to the key persons asking for potential participating students. Only one university student association replied, and was accessed successfully by the researcher. In general, the response rate was not high, in part because the academic atmosphere in China is not felt to be sufficiently supportive, and the rapport between academics and the researched is not yet fully established. From this point onwards,

the student association forwarded introduction, invitation letters (see Appendix 3.2), and consent forms to all its members, and afterwards 15 first-year undergraduates replied and indicated their willingness to participate in the present study. The researcher then contacted each of them to set up a time that fitted into their schedules. A total of 14 adult participants were recruited on the basis of stratified sampling, with one participant lost due to an illness. Consent was obtained before they participated in this study. (An example can be found in Appendix 3.3.)

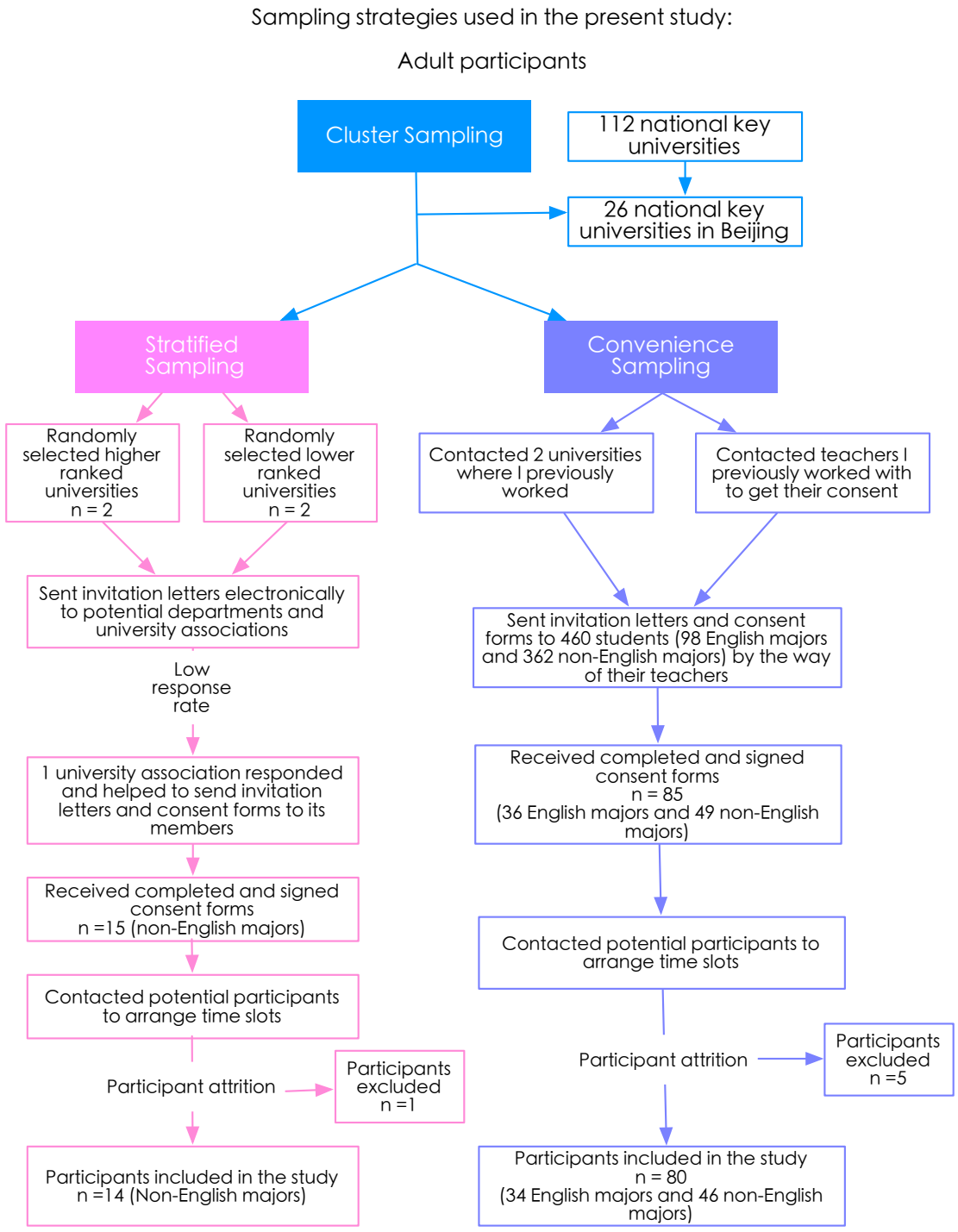


Figure 3.6 Sampling Strategies used in the present study: Adult participants

Along with the use of stratified sampling, a convenience sample was recruited in the following steps. The researcher first contacted the teachers with whom she had previously worked to get permission to include their students in the present study. An e-mail introduction and invitation letter (also see Appendix 3.1) was sent to six teachers, informing them of the general research plans. Four of them indicated their willingness to invite their students to be included in this study. Later on, they introduced the research project, and sent letters of invitation (also see Appendix 3.2) and consent forms to their students. In all, 85 students gave their consent, and eventually 80 of them participated in the present study, with 5 participants lost due to unexpected circumstances like suddenly conflicting schedules, an overload of assignment work, and a voluntary withdrawal from participation.

Taken together, 94 participants took part in the present study, with 14 of them selected based upon stratified sampling and 80 of them selected based upon convenience sampling. It should be noted that the sample might not be representative of all university-level EFL learners in the whole country but at least approximate those learners who learn EFL at '211 Project' universities in the geographical area being investigated or its characteristically similar areas.

As regards the child participants, convenience sampling was employed to recruit them. This is because, as compared to the adult part, there were more disruptions that might challenge the well-established educational routines, and particularly because pupils needed to be pulled out of the classroom one by one to complete

computer-based tasks. Considering such inconveniences, schools were likely to raise concerns about potential negative influences that might arise on pupils' educational progress and their final exam scores, as well as about the negative attitudes parents might hold towards their children's participation in a research project. As such, non-probability convenience sampling was conducted at the end of April 2013. The researcher contacted and paid visits to two primary schools that she had prior connections to. During those visits, the researcher explained the significance of and the procedure of this project to their head teachers and directors, and emphasized the ethical codes with which the study would comply. After hours of negotiations, one school finally allowed the researcher to conduct experiments on two classes of 5th-grade pupils, while the other school declined to participate in the present research.

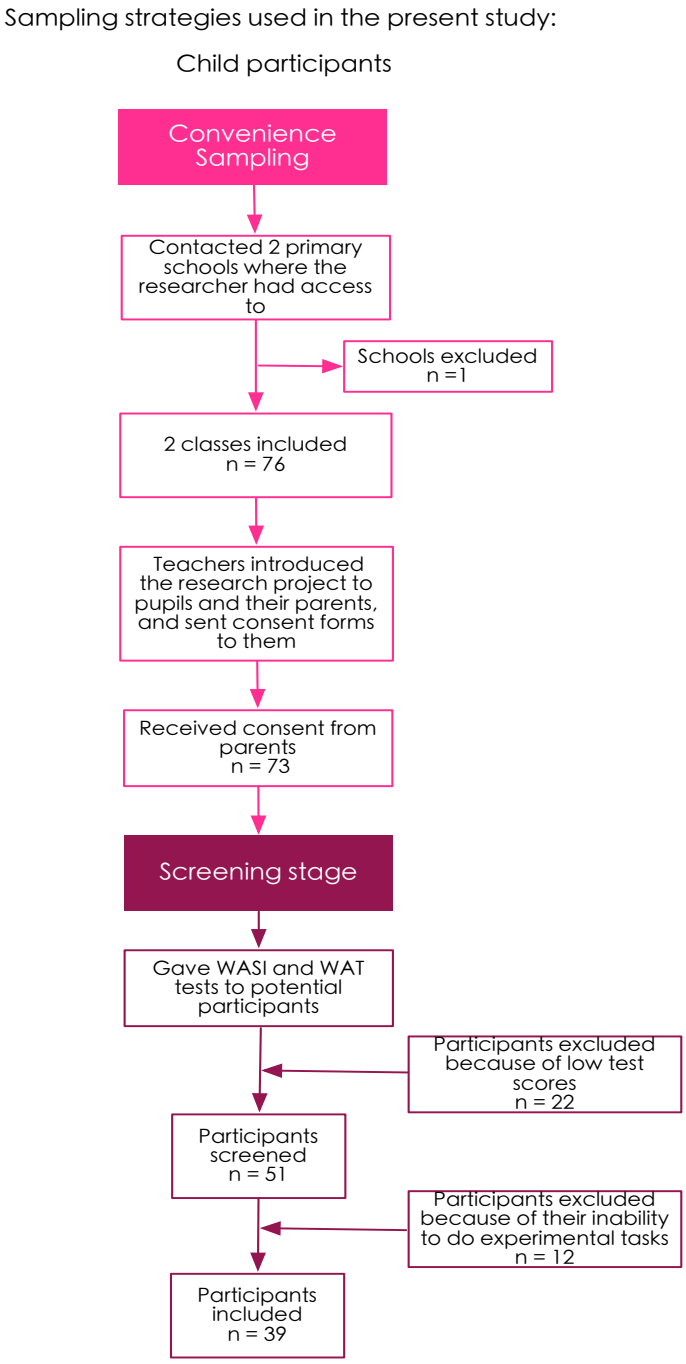


Figure 3. 7 Sampling Strategies used in the present study: Child participants

Having selected the sample school, the researcher contacted the homeroom teachers who were in charge of those two classes, and discussed on the time slots available for their pupils to perform the experiments. After this, the teachers

introduced the research project to their students in a class meeting, and to the students' parents in a parent's meeting. The invitation letters written in Chinese (see Appendix 3.4) and 'opt-in' consent forms were sent to parents by the way of the homeroom teachers, and the completed and signed consent forms (see Appendix 3.5 for an example) were returned from 73 out of 76 parents in the next few days. Before embarking on the experiments, the WASI and WAT tests were given to potential participants for the purpose of screening out the children with unusually low intellectual abilities and with very limited vocabulary knowledge. After the screening procedure was completed, 22 children, either with the standardized WASI scores below 50 and/or with the WAT scores below 35, were excluded from the present study, as these cut-offs were far below the average score. Moreover, the computer-based tasks were found to be beyond 12 participants' ability to complete at a later stage so that they were also excluded afterwards. In all, the child sample consisted of 39 pupils in the final analysis.

3.6 Research procedures

The research procedures are presented overleaf, and will be described in detail below.

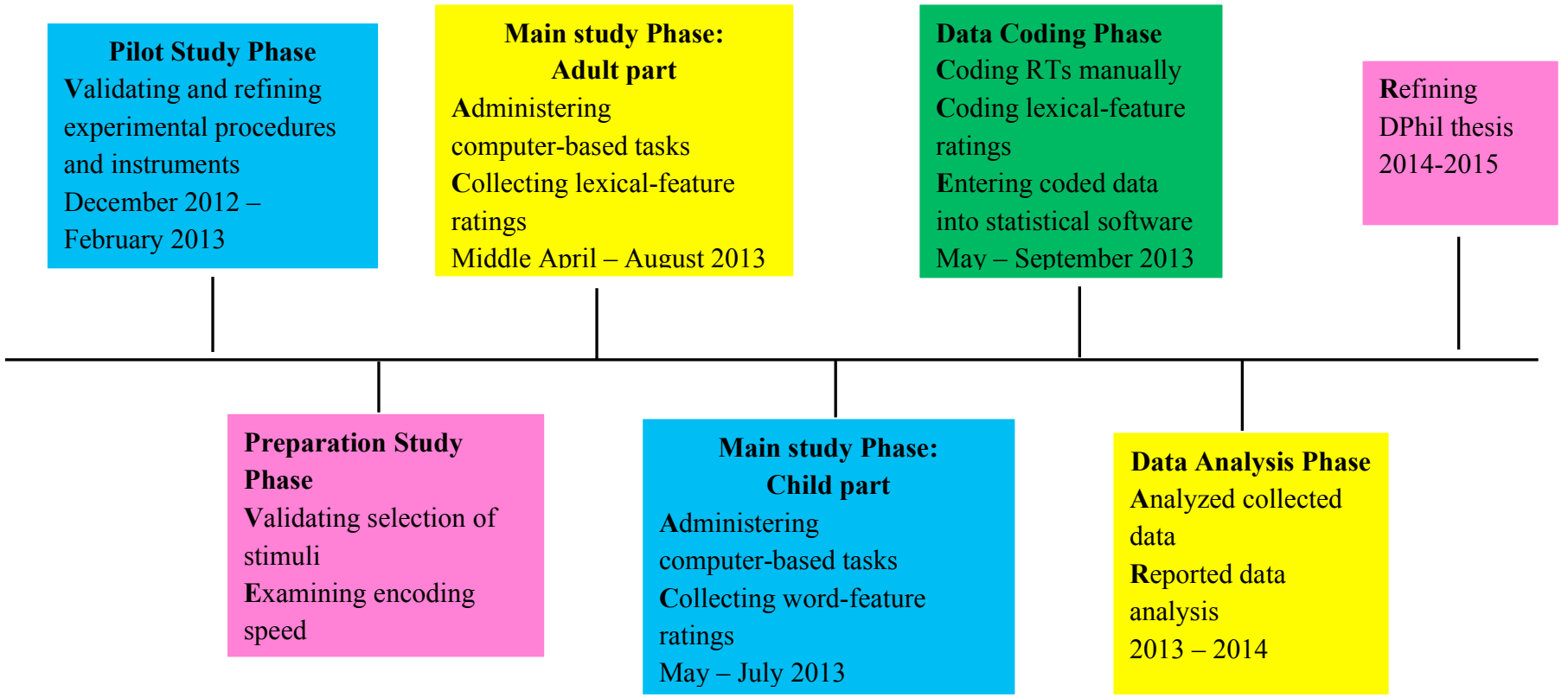


Figure 3. 1 Overview of research procedures

3.6.1 Pilot study phase

A *pilot study* is a smaller version of the actual study and aims to try out the procedures planned for the main study (Cohen et al., 2007). Such a trial enables researchers to assess the adequacy of a research design, to evaluate the appropriateness of research methods, and to test and modify research instruments (Robson, 2002). This is especially the case with experimental studies, as it is hard to get an experimental procedure right the first time. The purposes of the present pilot study are as follows:

- To check whether or not the written instructions were made clear to the participants;
- To validate, diagram and refine the experimental procedures;
- To become familiar with operations of the apparatus and the software;
- To estimate the minimum time needed for all experimental trials;
- To identify adverse effects (e.g. tiredness and distress) that participants might have;
- To conduct a preliminary analysis of the pilot data.

The pilot study was conducted with 5 Chinese pupils (Mean = 9.45; SD = 1.67), and with 9 Chinese postgraduates (Mean = 22.13; SD = 1.46) at the University of Oxford between December 2012 and February 2013. The participants were selected using convenience and snowball sampling. It should be acknowledged that although the 'pilot' participants were not identical to the population of interest, they were

highly similar to the main study participants in terms of several demographic characteristics as well as the experience learning foreign languages. For instance, the pilot study participants, much as the same as the main study participants, all had Chinese as their native language and English as their second language.

Moreover, in order to secure the consistency in experimental procedures between the pilot and the main study, what occurred in the pilot closely resembled what occurred in the main study at every step (cf. Section 3.6.3): first, the child and adult participants were presented with instructions explaining how to perform the experiments. After this, the participants were asked to respond to three blocks of 20 stimuli (in the child group) or three blocks of 30 stimuli (in the adult group) as rapidly and accurately as possible. This was followed by the administration of a set of word-feature rating tasks through which the necessity and appropriateness of those tasks were validated. Lastly, participants' advice and feedback were sought through informal interviews so as to improve, as far as possible, the experimental procedures and instruments to be used in the main study.

The pilot study served the set purposes although its results might hold limited implications for the main study as different populations were involved in both studies. Based on what happened in the pilot, several significant modifications were made to attenuate, if not overcome, the problems raised and identified during this phase of study. For example, some child participants produced invalid responses (coughs, hesitations, false starts, or prenominal verbalization like 'that's an apple). When this

occurred, response latencies could not be accurately coded. Therefore, it would be advisable to remind the main study participants to generate interpretable and usable responses as far as possible. Additionally, it was decided that not all features of words were rated on a 7-point scale because some adult participants reported that it was difficult to discriminate between seven points in the WT rating task. Rating word concreteness was beyond the child participants' cognitive operations to do so. Instead, they rated the strength of perceptual experience with which the meaning of each stimulus word could generate. The last modification worthy of mention was that the researcher found it useful to have a checklist containing all necessary information related to the experiment such as steps, duration, and apparatuses on hand (see Appendix 3.6).

As mentioned above, the pilot data did not hold significant implications for the population of interest; it was therefore felt unnecessary to present the statistical results within the limited space available. But even so, the present pilot study did help the researcher to validate the appropriateness and suitability of research procedures and instruments, as well as to identify and make specific modifications to the main study.

3.6.2 Preparation study phase

This phase consisted of two main sub-studies, the purposes, procedures, and results of which are described in detail below.

3.6.2.1 Sub-study 1: Encoding of pictures and characters

PURPOSES

The aim of this study was to compare the encoding speed of pictures with that of Chinese characters through a semantic categorization task. This step was deemed necessary because any comparisons between picture naming and forward translation could be validated only if their encoding speed was not significantly different from each other (also see Section 3.2.1 above). However, if this was not the case, it was difficult to interpret the results and subsequently to determine the general pattern of L2 lexical processing in spoken production.

EXPERIMENTAL PROCEDURES

A total of 28 children (Mean = 10.56 years, SD = .68) and 28 adults (Mean = 18.25 years, SD = .93) were recruited, a sample who was comparable to the main study participants in terms of demographic characteristics and the experience learning English as a foreign language. Note that the participants who took part in the preparation study did not join the main study.

This preparation study was run using the Superlab software (Cedrus Corporation, 1999) on a MacBook computer. There were two blocks of five practice trials each and two blocks of 30 test trials each. One block contained different pictures, and the other block contained different disyllabic Chinese characters. On both blocks, half of the stimuli represented objects, while the other half related to motion events. Both blocks

were matched in terms of the frequency of usage based on the Language Corpus System of Modern Chinese Study (LCSMCS, www.dwhyyjzx.com/cgi-bin/yuliao/). The stimuli were presented in a random order within either block, and the blocks were counterbalanced using a Latin square design. All pictures were kept in the same size (10 cm * 10 cm), and all Chinese characters in the same size and font (50-point Song font), and all displayed against a black background screen. The following instruction was given to the participants: Please decide whether a picture or a word depicts a motion event or an object as accurately and rapidly as possible. Please press the appropriate marked button on the keyboard.

Each trial was structured as follows (see Figure 3.9): a blank screen appeared lasting 1500 ms first, and then a red fixation point (+) was presented in the center of a computer screen for 800 ms. After this, a blank screen was presented for 200 ms, immediately followed by a picture or a word on the screen. The participants were asked to make a semantic judgment on the presented stimulus by pressing a marked button, and this was immediately followed by a next trial. The participants' RLs were recorded from the stimulus onset to a button press. The Superlab software recorded both the specific response (button press, D or W) and RL for each press.

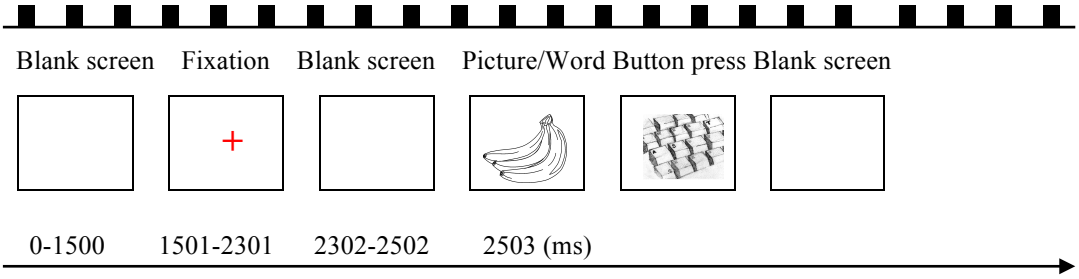


Figure 3. 9 Structure of a trial: Encoding pictures and Chinese characters

3.6.2.2 Sub-study 2: Stimulus pictures

PURPOSES

The aim of this study was to examine the degree to which participants agreed on the name of a picture (i.e. name agreement), the degree to which mental images generated by participants in response to a picture matched with its visual appearance (i.e. image agreement), and how many images the name of an object could elicit (i.e. image variability), as these three variables have been found to significantly impact upon PN latencies (Bonin et al., 2002).

EXPERIMENTAL PROCEDURES

A total of 123 candidate pictures (in the child participant group) and 220 candidate pictures (in the adult participant group) were assessed by the participants. The participants were the same as in the first sub-study. They were tested individually in a quiet room. They were instructed to name a picture aloud in the L2 using as many words as possible; in doing so, the number of different names assigned to a particular picture across participants could be measured. The participants were allowed to say 'I don't know' whenever they were not sure what a picture depicted. After this, image

agreement and image variability of each picture were rated by participants on a 5-point Likert scale (1 = strongly disagree/few images, 5 = strongly agree/many images).

The apparatuses were the same as in the first sub-study. The written instructions in Chinese, which explained the steps taken in performing this task, were presented on the screen first. The detailed instruction was as follows: You are about to see a set of pictures one by one. Please name each picture with English words as many as possible. After this, please judge how closely each picture resembles your mental image of the object or entity on a 5-point scale. Lastly, please rate how many different images each stimulus word could evoke on a 5-point scale. The child and adult participants performed 123 and 220 successive test trials respectively, with five practice trials preceding them. As illustrated in Figure 3.10, each trial began with a blank screen lasting 1500 ms, and then a central fixation point in red was on the screen for 800 ms to direct participants' attention to where a stimulus appeared. The fixation point disappeared 200 ms before a stimulus picture was presented. The presentation format was the same as in the above experiment. The participants were asked to name the picture in their L2 using as many words as possible, and the experimenter wrote down their answers manually on a record sheet. Then, the participants gave their ratings on the image agreement and image variability of the stimulus presented, the results of which were also recorded by the experimenter on a sheet. After this, the participants pressed the left mouse button, which initiated the next trial. A short break was given

to the participants when needed. The researcher monitored the whole experimental process. Together with manual recordings, participants' responses were also recorded digitally using Audacity software 2.0.2 (http://audacity.sourceforge.net) so as to make a backup copy of their answers just in case the manual ones needed to be crosschecked or further confirmed.

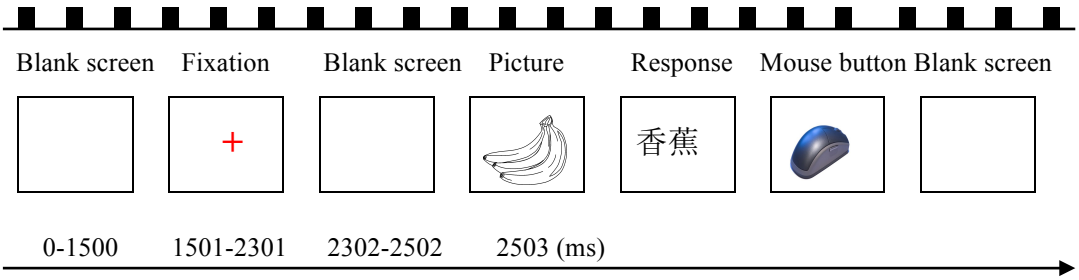


Figure 3. 10 Structure of a trial: Name agreement, and image agreement and variability

As mentioned above, only those pictures that met the following criteria were used as stimulus pictures in the present study: (1) high name agreement (the mean percentage higher than 95); (2) high image agreement (the mean ratings higher than 4.00); and (3) low image variability (the mean ratings lower than 2.00). A total of 117 pictures (in the child participant group) and 210 pictures (in the adult participant group) were finally selected for the PN task.

In addition to the above-mentioned three criteria, the visual complexity of pictures was measured by their file sizes, a measure which has been previously adopted in the study by Bates et al. (2003). The mean values of image complexity were compared across stimulus blocks, and its results are presented in Section 4.2.2 and Section 5.2.2.

In order to ensure that all word pairs were translation equivalents for the participants, another 15 adult participants and 15 child participants who were comparable to the main study participants were asked to back translate all the selected stimuli from English to Chinese. Only those words that were consistently regarded as translation equivalents across all participants were considered for this study. In all, a total of 111 words (in the child group) and 201 words (in the adult group) served as stimuli in the present study.

3.6.3 Main study phase

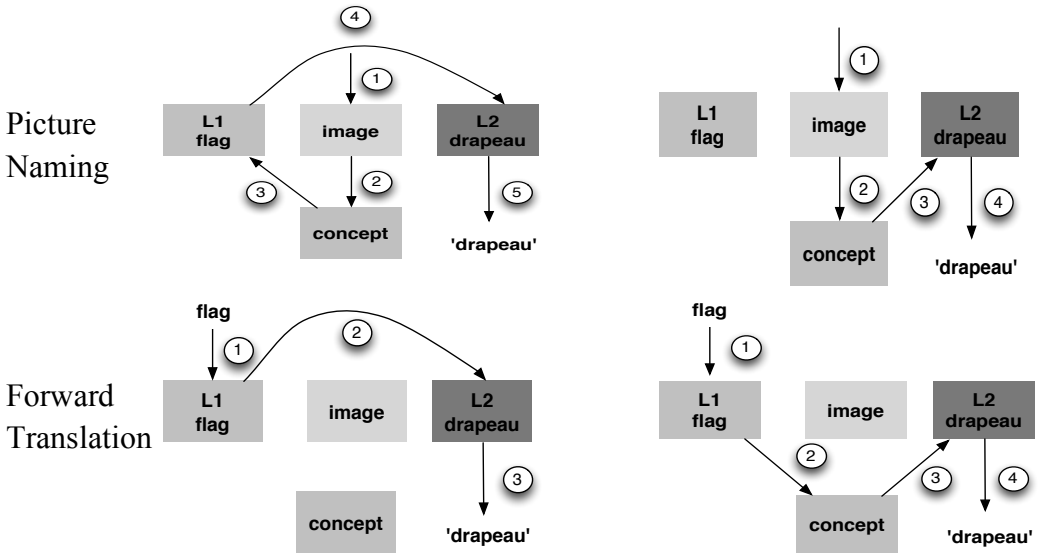
During the main study phase, RL-based experiments were performed by the participants. The values of relevant lexical features assigned by the participants were collected, and the participants' vocabulary knowledge in terms of both breadth and depth was assessed. The demographic and language background questionnaires were administered to the adult participants. How these tasks were ordered is outlined in Table 3.1 above. It should be noted that those tasks that demanded more effort and attention were arranged prior to those that required less effort and attention. The specific procedures taken and the instruments used throughout this phase are described in detail below.

3.6.3.1 RL-based experiments

LOGIC BEHIND EXPERIMENT: GENERAL PATTERN OF L2 LEXICAL PROCESSING

The fundamental logic behind the current experiment was the same as the logic behind Potter et al.'s study (1984, see Figure 3.11)—if L2 words are generally

processed through lexical routes (i.e. word association model), the RL of PE, on average, should be significantly longer than that of FT; by contrast, if L2 words are generally processed through conceptual routes (i.e. concept mediation model), the RL differences between PE and FT should not be statistically significant.



A. Word Association Model B. Concept Mediation Model

Figure 3. 11 Logic behind the experiment (Potter et al., 1984)

PARTICIPANTS

Thirty-nine native Chinese 5th grade pupils (Mean = 10.31 years, SD = .47) learning EFL in Mainland China, and 94 native Chinese undergraduates (Mean = 18.80 years, SD = .84) learning EFL in Mainland China participated in the present main study. A general description of their characteristics is presented in Section 3.5 above, and a more detailed profile of the sample presented in Sections 4.2.1 and 5.2.1.

APPARATUSES

The apparatuses were basically the same as those used in the preparation studies (see Section 3.6.2), that is, experiments were delivered using the Superlab software, and the responses given by the participants were recorded using the digital audio editor Audacity software.

STIMULI

A total of three blocks of 37 words (in the child participant group) and three blocks of 67 words (in the adult participant group) served as stimuli in the present study (see Appendix 4.1 and Appendix 5.1 for a full list of stimuli). Stimulus words mainly fell into the following semantic categories: animals, fruits and vegetables, clothes and accessories, parts of human body, colors and shapes, numbers, occupations, actions, nature, and vehicles. These stimulus words shared the following characteristics: 1) the stimulus pictures had high image agreement, high name agreement, as well as low image variability according to the participants' judgments (cf. Section 3.6.2 above); 2) within each block, the stimulus words varied in terms of certain lexical features such as word length, the frequency of usage, and age of acquisition; 3) the stimulus words could be matched in terms of the lexical variables concerned across blocks (cf. Section 4.2.2 and Section 5.2.2 below); and 4) the stimulus words or pictures were not beyond the participants' ability to translate or name.

As mentioned above, in order to ensure consistency in presentation and format, all stimulus words were presented in the same format over the main study period (Chinese: white 50-point Song font), and all displayed against a black background

screen. All stimulus pictures were kept in the same size (10 cm * 10 cm) and displayed at the center against a black background screen.

EXPERIMENTAL PROCEDURES

The participants were randomly assigned in almost equal numbers to one of the three sequences of stimulus conditions. Each participant was seated at a comfortable distance in front of the computer screen in a quiet area to perform a set of tasks individually over the whole experiment. The specific procedures and practices involved in this experiment are described in detail below.

First of all, the experimenter wrote down the participant's unique identification number on a record sheet. Before performing the experiments, the participant was asked to read step-by-step instructions presented bilingually on the computer screen. The instructions explained what occurred on each block, and how the participants should respond to each stimulus. Specifically, in picture naming, the instruction said 'you are going to see a set of pictures one by one, please name each picture in English (or Chinese in Chinese picture naming task) as accurately and rapidly as possible'. In forward translation, the instruction said 'you are going to see a set of Chinese words one by one, please translate each into English as rapidly and accurately as possible.' Once the participant finished reading the instructions, they were given a chance to ask any questions that they might have about the experiment. After this, the participant was tested with five practice trials before the start of each block, followed by verbal feedback, if needed, given by the experimenter. Practice trials might be repeated if the

participant did not know how to perform the experiment correctly.

In the test trials, the child participants received three blocks of 37 random trials each, and the adult participants received three blocks of 67 random trials each. Each trial was structured as follows (see Figure 3.12 below): a blank screen appeared for 1500 ms first. This was immediately followed by a 'beep' sound (11,025 Hz, 476 ms), which triggered the presentation of a stimulus word or picture. As the stimulus appeared on the screen, the participant needed to either name the picture or to perform forward translation as rapidly and accurately as possible. Once a response was given, the experimenter left clicked the mouse to trigger the next trial. The participants took a two-minute break between blocks to relieve their tiredness.

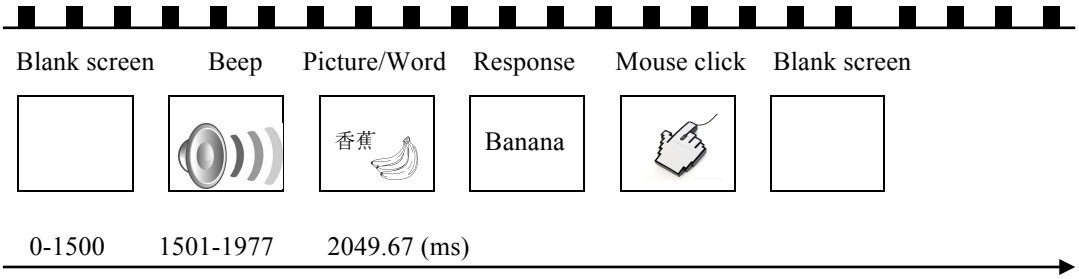


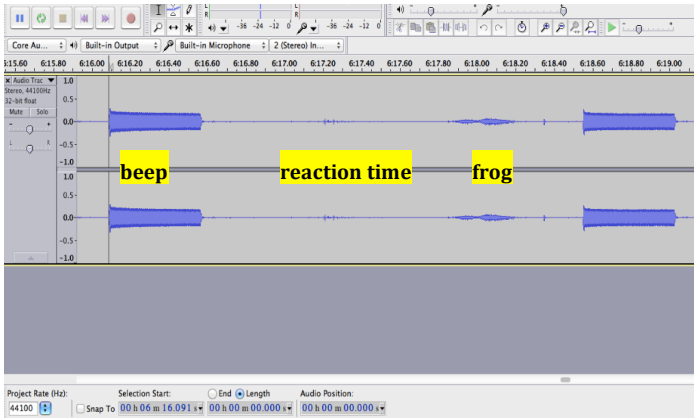
Figure 3. 12 Structure of a trial: Comparing picture naming and forward translation

3.6.3.2 Recording RAs and RLs

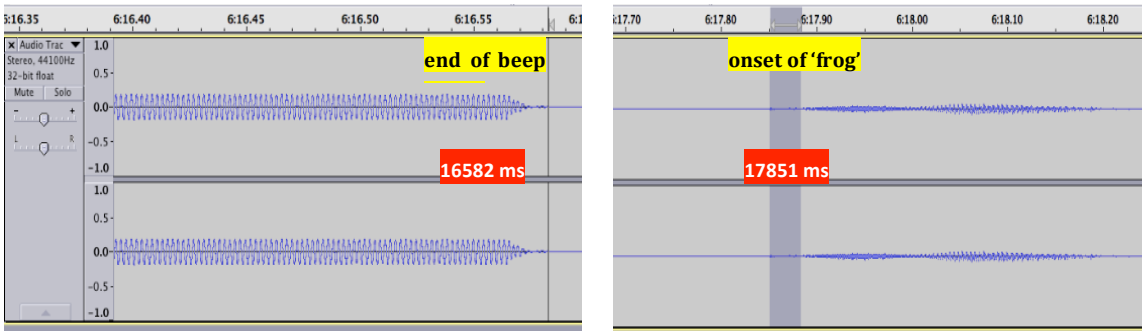
The accuracy of each response was recorded manually on a response sheet by the experimenter. The latency of each response was recorded using the Audacity software 2.0.2 without a voice key. This is because, as suggested by Duyck et al. (2008), a manual recording of the entire naming production and a manual measure of all RLs are enabled to overcome the differential sensitivity thresholds observed in using a

voice key, thus providing a more reliable measure of the RLs. In light of this, the Audacity was used in this experiment to record the sound waves of each response, upon which the RL could be measured.

In the present study, RLs were defined as the latencies between the presentation of a stimulus and the initiation of a vocal response. Before the presentation of each stimulus, a 'beep' sound was delivered, and the time difference between the end of a 'beep' sound and the onset of a stimulus was calculated to be 80 ms. To illustrate, the RL in naming the word 'frog', for example, was measured as follows (see Figure 3.13 below): 17851 ms (onset of the response wave) – 16582 ms (end of beep sound) – 80 ms (duration between end of beep sound and onset of stimulus) = 1189 ms. All the RLs were calculated manually in this way by the researcher, as well as by another student researcher. Their calculations were compared against each other, and any significant differences (> 10.00 ms) identified were double-checked and then corrected when needed by the researcher.



The original version



The enlarged version

Figure 3.13 Sample of an acoustic waveform for the production of 'frog'

MEASURES AND RATINGS OF WORD FEATURE-RELATED VARIABLES

The values of some lexical features were collected through participant-rated questionnaires. To be specific, the child participants completed the questionnaires independently in class for the purpose of minimizing the disruptions that might have on the established educational routines. The adult participants answered the questionnaires individually, immediately after their RL-based experiments were completed. It took approximately 80 minutes for each child participant, and approximately 60 minutes for each adult participant to fill out all the questionnaires. Together with their ratings, some other lexical variables were normed through psycholinguistic databases (see Table 3.4 below for a summary). These lexical variables were selected because they correspond to different stages of spoken production. Also, they have received much attention in the research literature, and have been found to influence the efficiency of lexical processing among child learners (e.g. Bjorklund & Thompson, 1983; D'Amico et al., 2001; Greenberg & Bjorklund, 1981; McFalls et al., 1996; Schwanenflugel & Akin, 1994) and among adult learners

(e.g. Alario et al., 2004; Balota & Chumbley, 1984; Cuetos et al., 1999; De Groot, 1992a; De Groot et al., 1994; De Groot & Keijzer, 2000; Hino & Lupker, 1996; Lichacz et al., 1999; Prior, MacWhinney, & Kroll, 2007). It should be noted that it seems implausible to cover every lexical aspect of L2 words, partly due to the sample size and data type required for using SEM in the present study. Nevertheless, the lexical variables of interest reasonably represent each lexical layer of words.

Moreover, word concreteness, word typicality, and subjective word frequency were estimated by the participants rather than generated by the psycholinguistic databases. This is because most external referents are constructed with reference to monolinguals' resources, which may not adequately represent bilinguals' estimates. Importantly, as shown in a study by Balota et al (2014), subjective estimates better predicted lexical decision and naming performance above and beyond objective norms. Arguably, the variance in the participants' performance of picture naming and translation may be better explained by the subjective estimates than by the existing databases.

Table 3. 3 Values of word feature-related variables: Ratings and databases

Level	*Variable*	*Method*	*Questionnaire format/Descriptions*	*Example/Source*
Meaning level	WC	Rating	Adult participants: A 7-point Likert scale (1 = abstract, 7 = concrete), adapted from Brysbaert, Warriner, and Kuperman (2013)	See Appendix 3.7 for an example
			Child participants: A 3-point Likert scale used to judge how target words can be experienced by five senses, adapted from Connell and Lynott (2012)	See Appendix 3.8 for an example
	WT	Rating	A 5-point categorization task (1 = atypical, 5 = typical) adapted from Schwanenflugel and Rey (1986)	See Appendix 3.9 (adult part) for an example See Appendix 3.10 (child part) for an example
	SND	Database	The counts for the semantic neighbors of the target items, obtained from the Wordmine2 database (Durda & Buchanan, 2006)	http://web2.uwindsor.ca/wordmine/
	NoRS	Database	The total number of word senses listed in the Wordsmyth dictionary (Parks et al., 1998)	http://www.wordsmyth.net
Form level	NoL/NoP/NoS	Database	The number of letters, phonemes, and syllables calculated with reference to the Wordsmyth dictionary (Parks et al., 1998)	http://www.wordsmyth.net
Usage level	SWF	Rating	A 7-point (for adult participants) / 5-point (for child participants) Likert scale, adapted from Balota et al. (2001)	See Appendix 3.11 (adults) for an example See Appendix 3.8 (children) for an example

To further elaborate, a more detailed description of these measures is presented below.

3.6.3.3 Participant-rated questionnaires

Participant-rated questionnaires were used to collect the values of WC, WT, and SWF. Before describing the specific measures of each variable, there are some points worthy of mention. First, the number of points assigned to each rating scale differed slightly from questionnaire to questionnaire but all guided by the following consideration, that is, enough points should be used to reflect the differentiation in the variables being rated, but at the same time those points should not go far beyond the participants' ability to discriminate them accurately and consistently. Second, all the questionnaires were presented in a booklet form. For the adult participants, instructions on how to fill in the questionnaires were stated clearly on the instruction page. For the child participants, instructions were given in class by the researcher. Once the completed questionnaire was returned and received, the researcher and the homeroom teachers checked it carefully to ensure that no answers were inadvertently missed out. Third, in order to check the reliability of participants' ratings, 10 (in the child group)/20 (in the adult group) stimulus words were repeated in each questionnaire.

Word concreteness (WC). The adult participants were asked to indicate how concrete the meaning of a word was by using a 7-point scale from abstract to concrete, and the child participants were asked to indicate how strong the meaning of a word

could be experienced through our five senses (seeing, hearing, touching, tasting, and smelling) by using a 3-point scale from weak to strong. Clearly, the questionnaire format designed for the child participants differed significantly from that designed for the adult participants. This is because the pilot study found that the child participants did not have a clear understanding of what was word concreteness but this was not the case with word typicality and subjective usage frequency. Instead of directly rating word concreteness, the main study participants were asked to estimate the strength of their perceptual experiences about the meaning of the stimulus words, a method that has been used in a study by Connell and Lynott (2012).

Word typicality (WT). There were five rating levels from atypical to typical under each semantic category, and the WT was rated category by category. The child and adult participants were asked to estimate the WT degree of each stimulus word by assigning the word into a column corresponding to its rating level. In the pilot study, such a questionnaire format was found to better capture the accuracy and to better secure the reliability of participants' estimates as compared to the format used for rating the WC.

Subjective word frequency (SWF). For the child participants, the SWF was measured on a 5-point scale (1= never, 2 = rarely, 3 = sometimes, 4 = often, and 5 = always); whereas, for the adult participants, this variable was measured on a 7-point scale (1= never, 2 = rarely, 3 = occasionally, 4 = sometimes, 5 = often, 6 = almost always, and 7 = always) The child and adult participants were asked to use a rating

point to indicate how often they had encountered or used the stimulus words in listening, speaking, reading, and writing.

3.6.3.4 Databases

Number of related senses (NoRS). The NoRS was measured by the total number of senses listed in the Wordsmyth dictionary (Parks et al., 1998), a measure that has been adopted in previous studies (e.g. Beretta et al., 2005; Rodd et al., 2002; Rodd, Gaskell, & Marslen-Wilson, 2004). Figure 3.14 is an example, showing that there are three Wordsmyth entries for the word 'apple'.

ENT: **apple**

SYL: ap-ple

PRO: ae pəl

POS: noun

DEF: 1. a firm round fruit with succulent white flesh and a red, green, or yellow skin.

DEF: 2. the tree on which this fruit grows.

DEF: 3. any of various fruits resembling the apple, such as the custard apple.

Figure 3. 14 Description of the Wordsmyth entry for the word 'apple'

Semantic neighborhood density (SND). Semantic neighbors refer to those words that co-occur or are used in similar contexts (Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008). The SND was defined by the number of global neighbors identified for each word within a specified radium of semantic space (Pexman et al., 2008). The values of SND were obtained from Wordmine2 (global semantic neighborhood size, Durda & Buchanan, 2006), which is constructed based on a corpus of English texts.

The Wordmine2 has been used in the pervious study by Zdrazilova and Pexman (2013) and by Schock, Cortese, Khanna, and Toppi (2012).

Number of letters (NoL)/Number of phonemes(NoL)/Number of syllable(s)(NoS). The values of NoL/NoP/NoS were calculated with reference to the Wordsmyth dictionary (Parks et al., 1998). A phoneme is the smallest sound unit, and a syllable is a relatively bigger unit of a spoken language. For instance, the NoL/NoP/NoS for the word *rainbow* /'reɪnbəʊ/ are 7, 5, and 2.

Age of acquisition (AoA). The estimates of AoA were determined with reference to a corpus of the participants' course books. This measure was deemed to be valid and reliable in that the participants' exposure to the L2 in non-native English speaking countries is largely limited to instructional settings.

3.6.3.5 Assessments, and demographic and language background questionnaires

A series of assessments and a questionnaire requesting demographic information and language backgrounds were administered to the participants. These tests and the questionnaire are described in detail below.

ASSESSMENTS

Two kinds of word knowledge tests were given to both the child and adult participants in the present study: one was used to assess their Vocabulary Breadth (VB, i.e. a size test), and the other used to assess their Vocabulary Depth (VD, i.e. a word associates test). The VB and VD, as two distinct but interconnected components

of vocabulary knowledge, are both believed to be crucial to L2 learning and development.

More specifically, how many words a participant knows is assumed to predict their performance in reading and writing, the overall proficiency level, and their academic achievements (Laufer & Goldstein, 2004). It should be noted that the VB alone captures only a dimension of vocabulary knowledge. This is because, as rightly pointed out by Laufer and Goldstein (2004), vocabulary knowledge does not necessarily appear to be an all-or-nothing phenomenon but actually develops cumulatively. In light of this, the VD, as an estimate of various aspects of an individual lexical item in a more comprehensive and precise manner, should also be considered (Read, 2004). Additionally, it has been reported that the VD might make a unique contribution in predicting L2 proficiency, over and above the predictions made by vocabulary size (Xiao, 2007). It should be acknowledged that both VB and VD tests have their own limitations, for instance, the former has been criticized for 'the superficial treatment' of lexical items, and the latter has been criticized for only allowing a test of a small number of words (Laufer & Goldstein, 2004). In the present study, the VB was assessed using the X_Lex test (Meara, 2005), and the VD assessed using the Word Associates Test (WAT) developed by Read (1993). The appropriateness of the VB test was also confirmed by the suggestion indicated in the Foreign Language Assessment Directory (available at http://www.cal.org/CALWebDB/FLAD/).

As for the child participants, together with the VB and VD tests, a non-verbal cognitive ability test, the Wechsler abbreviated scale of intelligence (WASI) test, was administered to them for a screening purpose. The WASI test was not given to the adult participants because it was assumed that university students were all operating within the normal range of cognitive ability. In addition to the WASI, the child participants also received the Cambridge Young Learners English test (YLE, starters level), which served to assess their listening, reading and writing proficiency levels. Each test is described below.

X_Lex. The X_Lex (version 2.05, Meara, 2005) is a computer-based 'Yes/No' test, providing an overall estimate of learners' receptive orthographic vocabulary size. The words included in this test are randomly sampled from five frequency bands (the 1000-word level to the 5000-word level), and the frequency bands are based on the work of Hindmarsh (1980) and that of Nation (1984). This well-designed test was considered to be appropriate for the adult participants; therefore, it was used without any further modifications to estimate their vocabulary size. As regards the child participants, a revised X_Lex test was administered to them because they had a limited size of vocabulary. The words included in this test ranged from the 100-word level to the 500-word level. These words were drawn from the suggested vocabulary list appended to each lesson in the participants' course books. They were arranged according to the grade at which the participants had learned them, this is to say, words learned in the first term of grade three were viewed as the first hundred words, words

learned in the second term of grade three were seen as the second hundred words, and so forth. An X_Lex test and a revised X_Lex test (child version) eventually contained 20 words selected from each 1000-/100-word frequency band as well as an addition of 20 pseudo-words that sound or look like English words but were not real words. The specific test procedures are described below.

In the X_Lex test and revised X_Lex test (child version), a total of 120 words appeared on a computer screen one by one. (An example can be found in Figure 3.15.)

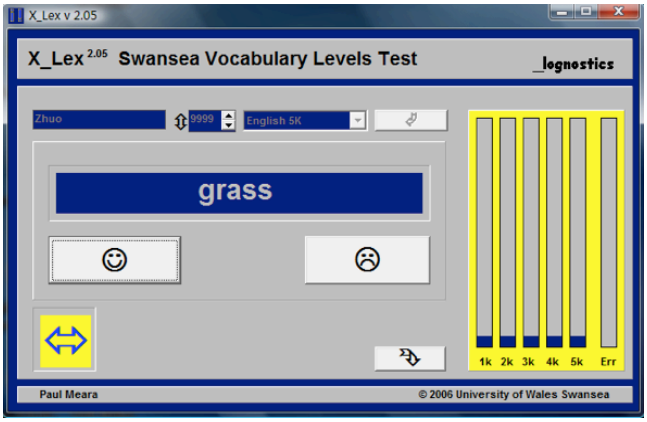


Figure 3. 15 Presentation of a lexical item in X_Lex

The participants were asked to indicate whether or not they knew their meanings by pressing Yes (i.e. a happy face) No (i.e. an unhappy face). The total test score was 5000 (in the adult group) / 500 (in the child group). The participants were awarded 50/5 points for knowing a real word, and lost 250/25 points for giving a 'Yes' response to a pseud-word. An example of the X_Lex result is shown in Figure 3.16. In this example, the adult participant's score of the test on 100 real words were 4550

but he gave a 'Yes' response to two pseud-words. In which case, his vocabulary size was eventually estimated to be 4050.

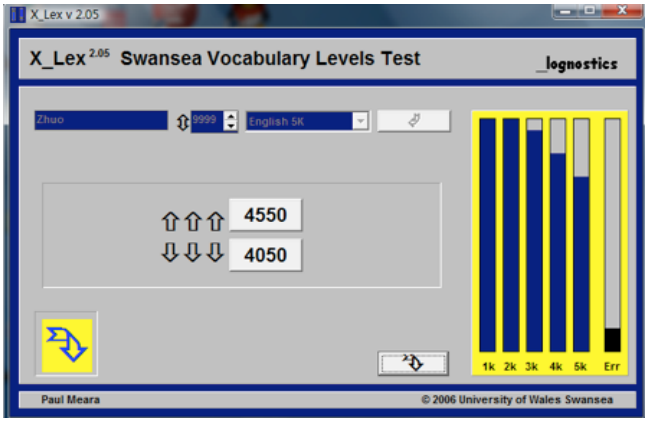


Figure 3. 16 Result of an X_Lex test

It should be pointed out that the inclusion of those pseudo-words helps to check the reliability of test takers' Yes responses to the real words (Meara, 1992). The concurrent validity of the X_Lex test has been demonstrated in the literature in which it correlates well with a multiple choice test used in Cambridge First Certificate Examination (Meara, 1992). Moreover, learners' performance on the Yes/No test format has been found to significantly predict their performance on the VLT (Vocabulary Levels Test) (M. Harrington, 2006). In short, the X_Lex test serves as an effective means of profiling the participants' vocabulary size. As shown in Table 4.7 below, the revised X_Lex (child version) correlated significantly with the English test scores (measured by the school's final exam), YLE (reading), YLE (listening), and the latencies of picture naming and forward translation. Such high correlations

corroborate the validity of this revised vocabulary-size assessment.

Revised word associates test. Another aspect of vocabulary knowledge, the depth, was assessed by a revised word associates test, which is based on the format of the WAT designed by Read (1993). This type of test has been viewed as a reliable way to assess how well a particular word is known by a learner (Greidanus & Nienhuis, 2002). The rationale behind the revised WAT is that the extent to which a learner is familiar with a test word can be estimated from how well they know its paradigmatically, syntagmatically, and analytically associated words. A paradigmatic relation holds when two words are synonyms, antonyms or have a hypernym-hyponym relationship. Moreover, a syntagmatic relation holds when two words are enabled to occur together (i.e. collocability). To take the test word *grass* as an example, since green grass can be viewed as an adjective + noun collocation, it can be reasonably assumed that the word grass has a syntagmatic relationship with the word green. Lastly, an analytical relation holds when the associate is likely to form part of the definition of a test word or to represent its core meanings.

The child version of the revised WAT test (see Appendix 3.12) contained 24 content words (8 concrete nouns, 8 adjectives, and 8 verbs) randomly drawn from the participants' course books. The adult version of the revised WAT test (see Appendix 3.13) contained 30 content words (10 concrete nouns, 10 adjectives, and 10 verbs) randomly selected from a CET-4 vocabulary list. These test words were considered most appropriate in light of the participants' proficiency levels. Each lexical item had

six choices: three word associates (i.e. one paradigmatic, one syntagmatic, and one analytic) and three distractors. The participants were asked to decide whether or not the presented word was an associate to the test word by tapping a marked icon on the iPad screen. A more detailed description of the characteristics of those associates and distractors are presented in Table 3.5 below.

In scoring, one point was awarded for each correctly chosen associate, with a potential full score of 3 for each target word, and 72 (in the child group) / 90 (in the adult group) for all the WAT words. Before administration of this test to the child and adult participants, three EFL teachers were invited to take the child and adult versions of the WAT, and the intra class correlation coefficient between the test takers' scores was high (child version: ICC = .98, and adult version: ICC = .97) and significant at 95% level (both versions: $p < .0001$). As shown in Table 4.3 below, the child participants' vocabulary depth scores were correlated significantly with the English test scores and YLE (Listening), which indicate the validity of this assessment. As shown in Table 5.8 below, the adult participants' vocabulary depth scores were correlated significantly with vocabulary size, English proficiency, and the latencies of picture naming and forward translation. Such high correlations suggest the validity of administering the revised WAT test to the adult participants.

It should be noted that different lexical items were used for the child and adult participants respectively. This is because the present study aimed at assessing the predictive effect of vocabulary knowledge on each age group's processing efficiency

rather than making a cross-age comparison on their vocabulary size and depth. If the same lexical items were used for both groups, it would pose a threat to the ecological validity of the present study. Moreover, the child participants had a limited size of vocabulary, so there were slight differences in the design of the revised WAT test, for example, the child participants did not respond to phonologically similar words since adopting this design would include too many new words for the child participants.

Table 3. 4 Descriptions of associates and distractors used in the revised WAT

Word class	**Target Words**	**Associates**	**Distractors**
Nouns	10 concrete (adult) 8 concrete (child)	1 P (sub- or super-ordinate) 1 S (agent-**action** or **adjective**-noun) 1 A (dictionary definition)	*Adult version:* 1 phonologically similar word (e.g. lamp-lamb) 1 word containing the test word (e.g. press-express)
Verbs	5 path and 5 manner (adult) 4 path and 4 manner (child)	1 P (synonym or antonym) 1 S (**agent**-action-**object** of action) 1 A (dictionary definition)	1 word syntagmatically unrelated to the target word (e.g. press-present) *Child version:*
Adjectives	5 positive and 5negaitve (adult) 4 positive and 4 negative (child)	1 P (synonym or antonym) 1 S (adjective-**noun**) 1 A (dictionary definition)	1 word paradigmatically unrelated to the target word (e.g. old-pink) 1 word syntagmatically unrelated to the target word (e.g. easy-jacket) 1 word analytically unrelated to the target word (e.g. swim-comic)

Notes: A: analytical; P: paradigmatical; S: syntagmatical.

Wechsler abbreviated scale of intelligence. The WASI test, designed by Wechsler (1999), is widely considered as a reliable and standardized measure of children's general intelligence (Sparrow & Davis, 2000). This test consists of four subtests: *Vocabulary, Block Design, Similarities, and Matrix Reasoning*. In the subset of

Vocabulary, examinees are asked to name pictures and then orally define the visually or orally presented words; in the subset of Block Design, examinees are asked to replicate a set of modeled or printed two-dimensional geometric patterns within a specified time limit using two-color cubes; in the subset of Similarities, examinees are asked to explain in what ways two items are alike through verbal reasoning; and in the subset of Matrix Reasoning, examinees are asked to opt for an item from the response choices that would complete the pattern shown in the grid (Wechsler, 1999). According to the WASI manual, the reliability of this test has been confirmed using several methods such as split-half reliability, test-retest reliability, and inter-rater agreement. With children, the WASI has shown good reliability ranging from .95 to .97 (Wechsler, 1999). The construct of WASI measures is similar to that of WICS-III measures (Wechsler, 1999).

As mentioned before, the WASI served a screening purpose in the present study; therefore, using only the subtest of Matrix Reasoning to obtain a quick assessment of the child participants' non-verbal intellectual status was considered acceptable, especially given the limited amount of time available. In performing the Matrix Reasoning subtest, the child participants were asked to complete a missing section of a matrix by selecting the correct response option. (An example can be found in Figure 3.17.) In order to save class time, this test was administered to all the child participants in class. Each participant was asked to complete 32 test items and to write down their choices on answer sheets. One point was awarded for each correct choice,

with a maximum score of 32. The participants' raw scores were then totaled and converted into standard scores following the WASI manual.

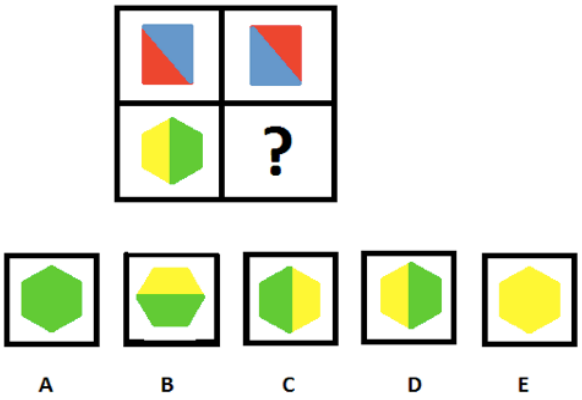


Figure 3. 17 Example of Matrix Reasoning test (Wechsler, 1999)

Cambridge young learners English test. The Cambridge YLE test is designed to measure the overall English proficiency of young learners in the EFL context. This test has three different levels: Starters (around the Pre-A1 level of the CEFR), Movers (around the A1 level of the CEFR), and the highest, Flyers (around the A2 level of the CEFR) (Cambridge ESOL, n.d.). The Starters level is believed to be suitable for those who have received around 100 hours of EFL training, the Movers level for those with 175 hours of training, and the Flyers level for those with 250 hours of training. The YLE test has been reported to have high validity and reliability (Bailey, 2005).

In the present study, in light of the child participants' proficiency levels and of the amount of their L2 exposure, a sample paper of the Starters test was used. The YEL test consisted of two components: listening, and reading and writing. More specifically, the listening component consisted of four parts, with each having 5

questions and with the total score being 20. The reading and writing component consisted of five parts, with each having 5 questions and with the total score being 25. The child participants all took the test in class, and the scores they achieved served to indicate their overall L2 proficiency.

DEMOGRAPHIC AND LANGUAGE BACKGROUND QUESTIONNAIRES

The background information about the adult participants was collected through questionnaires. The questionnaires were administered electronically using the app called QuickTap Survey after the participants completed assessments on the VB and VD. A brief description of the content of the questionnaire is presented below.

The questionnaire was divided into two parts. The first part contained demographic personal questions, including (a) age; (b) gender; (c) ethnicity (i.e. Han or Minority); (d) place of birth (North China, East China, South China, Southwest China, Northwest China, Northeast China, Southeast China, and Central China); and (e) living areas (Rural or Urban). The second part contained questions about the participants' experience learning the L2, including (a) the age of starting L2 learning; (b) the amount of L2 exposure per week; (c) main ways of English learning (through English courses, through cram courses, and through interactions with others); (d) academic achievements (measured by the overall score achieved on the university entrance exam); and (e) English proficiency levels (measured by the English score achieved on the university entrance exam).

Instead of using questionnaires, the background information about the child

participants was gathered through student progress files to which the researcher had access. The following information was collected in this way: (a) age; (b) gender; (c) the living area; and (d) the mid-term test scores on the Chinese, math, English, and science subjects.

To sum up, RLs were collected through computer-based experiments; values of word-feature related variables were either collected through the participants' ratings or directly obtained from psycholinguistic databases; learner-related variables such as their vocabulary knowledge, demographic information and L2 learning experience were gathered using assessments, questionnaires, and student progress files.

3.7 Analytical strategies used

3.7.1 Basic analysis strategies

3.7.1.1 Independent-measures ANOVA

In order to compare the mean values of lexical features across three blocks, one-way independent-measures ANOVAs were used. This inferential test is the appropriate one to be used as the means of more than two independent variables need to be compared (Tabachnick & Fidell, 2012). It should be pointed out that the independent ANOVA analyzes all the mean differences simultaneously; therefore, it is less susceptible to the family-wise error as compared to the use of a series of independent-samples t-tests (Field, 2013).

Independent-measures ANOVAs operate based on two estimates of variance, one is residual or error variance (i.e. differences among scores within each group, denoted

as MS_R), and the other comes from the variance between the groups (i.e. denoted as MS_M) (Tabachnick & Fidell, 2012). The major test statistic is the *F-ratio*, which can be used to determine whether or not there exist statistically significant differences among independent groups. F-ratio is equal to the mean squares of the model divided by the mean squares of the residuals (see the formula below).

$$F = \frac{MS_M}{MS_R} = \frac{\frac{SS_M}{df_M}}{\frac{SS_R}{df_R}}$$

The assumptions that need to be met for running independent-measures ANOVAs are as follows:

Data appropriateness. The DVs should be continuous, and the group variables should be dichotomous or categorical.

Independence. Each value of the outcome variable is independent of each other value.

Distribution. The continuous variable should be approximately normally distributed within each group. In the present study, the distribution of normality was checked with reference to skewness and kurtosis values. However, if skewness or/and kurtosis values were found to be greater than 2 or less than -2, an appropriate form of transformation, if any available, should be applied to improve the data distribution (Leech, Barrett, & Morgan, 2007).

Homogenity of variance. Homogenity of variance refers to the spread of scores is roughly equal in different groups of cases (Field, 2013). This assumption is important because any substantial differences among group variances may hardly reflect

between-group effects. Homogenity of variance was checked by the results of Levene's Test, and insignificance ($p > .05$) indicates that the variance is homogeneous. However, it should be pointed out that this assumption is less critical with equal or nearly equal sample sizes (Gravetter & Wallnau, 2007).

If the assumptions were violated, the *Kruskal-Wallis test*, which is a nonparametric equivalent of the independent-measures one-way ANOVA, was used.

3.7.1.2 Repeated-measures ANOVA and Mixed-model ANOVA

The one-way repeated-measures ANOVA was used in the present study to compare the encoding speed of pictures with that of Chinese characters. This test is considered appropriate since the same participants received both experimental conditions (Field, 2013). In such a case, their performance is likely to be related rather than independent across conditions. The overall structure of the repeated-measures ANOVA is presented in Figure 3.18 below. At the first stage, the total variance is partitioned into between-treatments variance and within-treatment variance (Gravetter & Wallnau, 2007), and then the second stage measures and subtracts out the between-subject variance within the same group, a process which makes it different from independent-measures ANOVAs (Gravetter & Wallnau, 2007). The assumptions of repeated-measures ANOVAs are as follows: data appropriateness, the distribution of normality, and sphericity (only applicable if there are more than 2 levels of the IV. If the assumptions were violated, the *Friedman's ANOVA* by rank, which is the nonparametric equivalent of a single-factor repeated-measures ANOVA was used.

The repeated measures ANOVA has the additional assumption of sphericity. Sphericity refers to the quality of variances of the differences between condition or treatment levels (Field, 2013), which is commonly checked using the Mauchy's test. If the assumption of sphericity is violated, either Greenhouse-Geisser correction (for epsilon less than or equal to .75) or Huynh-Feldt correction (for epsilon greater than .75) should be used.

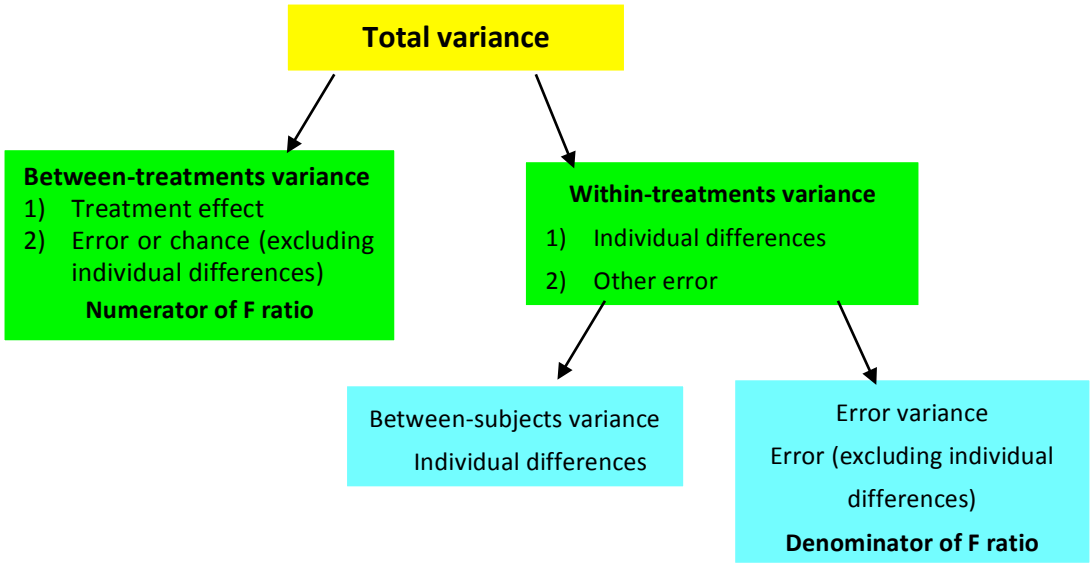


Figure 3. 18 The partitioning of variance for a repeated-measures experiment (Gravetter & Wallnau, 2007, p. 442)

In order to identify the general pattern of L2 lexical processing among non-English majors and among English majors, the mixed-model ANOVA F1 (by participants) and F2 (by lexical items), which combines the repeated measures and independent designs, was applied. The inclusion of an F2 test is because, as argued by Clark (1973), in addition to the variability due to participants, the variability due to lexical items should also be taken into account (Brysbaert, 2007). Given this,

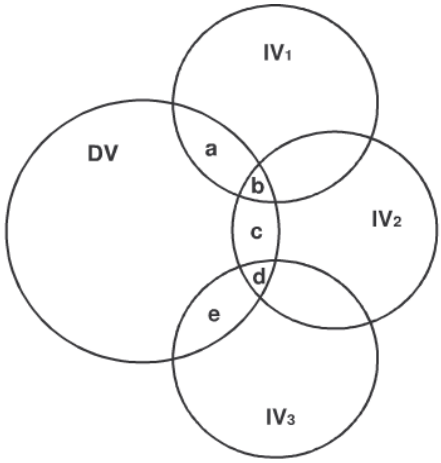
researchers are suggested to treat language materials as a fixed factor, and to conduct an F2 analysis, together with an F1 analysis.

3.7.1.3 Multiple regression

MR was used to examine whether learner-related variables predicted the speed of L2 lexical processing, and to examine whether learner-related and lexical feature-related variables predicted the tendency toward a certain pattern of L2 lexical processing. MR, as a general linear modeling approach, enables researchers to assess the strength and direction of the relationship between several IVs and one DV (Tabachnick & Fidell, 2012). Regression and correlation analyses are related in the sense that both deal with associations rather than differences between/among variables; however, as compared to correlation analyses, in regression, such an association is parameterized by a statistical model to provide more detailed quantitative descriptions of the relationships concerned.

There are three main analytical strategies used in MR, namely, the standard MR (simultaneous), the sequential MR (hierarchical), and the statistical MR (stepwise) (Tabachnick & Fidell, 2012). As recommend by Tabachnick and Fidell (2012), which strategy to use should hinge on 'what happens to overlapping variability due to correlated IVs and who determines the order of entry of IVs into the equation' (p.136). In the standard MR, no single variable is given priority than is another, and the contribution of each IV is assessed as if it had entered the regression after all other IVs had entered. Consider Figure 3.19. When the standard MR analysis is opted for,

each IV is only assigned its unique contribution (areas denoted as a, c, and e). It is possible that IV2 appear to be unimportant in the solution but actually is highly correlated with the DV.



Notes: IV: independent variable; DV: dependent variable.
Figure 3. 19 Variance in MR

As for the sequential MR, those IVs of relatively greater theoretical importance are given priority to so that they are entered into the analysis before others. As such, those earlier entered ones are allowed to capture variance in the absence of competition from later entered ones (Field, 2013). By contrast, in statistical regression, the order of entry of IVs is fully determined by statistical criteria.

Note that the standard MR rather than sequential or statistical MR was used in the present study, because there was a lack of theoretical and empirical grounds on which the order of entry of IVs was based. Moreover, which IVs should be included in the MR analysis follows the suggestion given by Tabachnick and Fidell (2012), that is, each IV included should be highly correlated with the DV but not highly correlated

with other IVs, and optimally, each IV can account for a substantial and unique segment of the variability in the DV.

Multiple regression has several assumptions that must be met before this analytical strategy is used.

Sample size and missing data. The following general rules are widely used in determining the size of a sample: a minimum ratio is 5 to 1, an appropriate ratio is 10 to 1, and a preferred ratio is 20 to 1 (Hair, Black, Babin, Anderson, & Tatham, 2005).

Normality, linearity, and homoscedasticity of residuals. Univariate normality was checked and confirmed using skewness and kurtosis values, all of which were well within the recommended limits (+/-2.00). As Tabachnick and Fidell (2012) suggested an examination of the scatterplot of residuals against the predicted DV provides a test of the assumption of normality of residuals (i.e. residuals are normally distributed about the predicted DV), the assumption of linearity of residuals (i.e. residuals have a straight line relationship with predicted DV), and the assumption of homoscedasticity of residuals (i.e. variance of the residuals about the predicted DV is the same for all predicted values).

Outliers. Univariate outliers were checked using standardized scores. Multivariate outliers were checked by examining the Mahalanobis distance of each case to the centroid of all cases. Mahalanobis distance is distributed as a chi-square variable, with degrees of freedom equal to the number of IVs.

Multicollinearity and singularity. Multicollinearity is a high correlation between

two or more independents (.90 and above), and singularity is a perfect or nearly perfect correlation between two or more independents. According to Tabachnick and Fidell (2012), highly correlated independents can lead to a failure of achieving statistical significance of regression coefficients, and a failure of model converge on a solution.

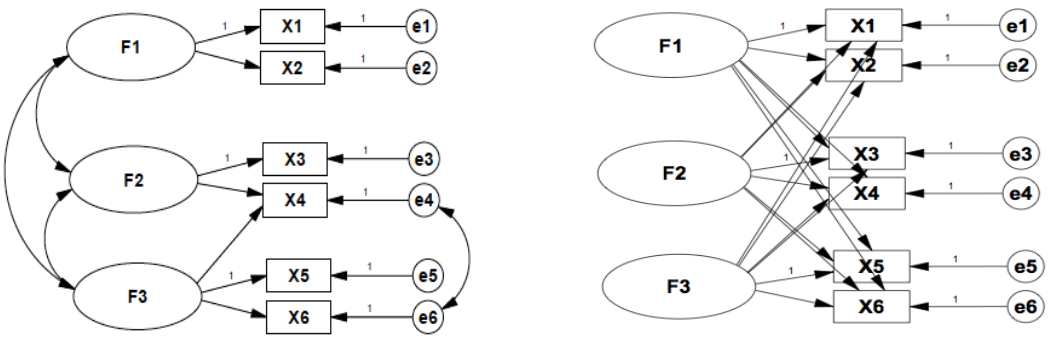
3.7.2 Structural equation modeling

Another important research aim of the present study was to assess the hypothesized structural model based on the identification of the relationships among the variables concerned. To this end, SEM was considered appropriate since it combines *path analysis* (PA), *confirmatory factor analysis* (CFA), and analysis of full structural models, allowing for specifically defining and testing a priori models, for uniquely estimating model parameters, and for comparing the goodness of fit of competing models (Marsh, 1987).

To be specific, PA is used to assess path models, which reflect a priori assumptions regarding direct and/or indirect effects among observed variables (variables can be observed directly and are measureable) as well as their spurious associations (Kline, 2011). Path models are considered as extensions of regression models. This is because like regression models, only observed variables are involved in path models; however, unlike regression models but like structural equation models, 'independents can be both causes and effects of other variables' (Garson, 2012a, p. 19). Therefore, one main advantage of using PA is that it enables researchers to

measure not only the direct but also indirect effects that one variable has upon another (Olobatuyi, 2006).

CFA is another type of SEM in which researchers attempt to examine measurement models, in other words, the relationships between indicators (i.e. observed variables) and latent variables (i.e. unobserved variables) (D. Harrington, 2008). On the other hand, CFA differs from SEM in that the latter allows for assessing structural paths between latent variables, such a function that CFA does not have (D. Harrington, 2008). Moreover, CFA is different from exploratory factor analysis (EFA), because in the former type of analysis, ideally, indicators are only connected with some pre-determined latent variables, some error terms are allowed to correlate, and some of the parameters are allowed to be constrained to certain values; by contrast, in EFA, every indicator is connected with every latent variable, error terms are not correlated, and all parameters are estimated from the data (see Figure 3.20 below, Blunch, 2012).

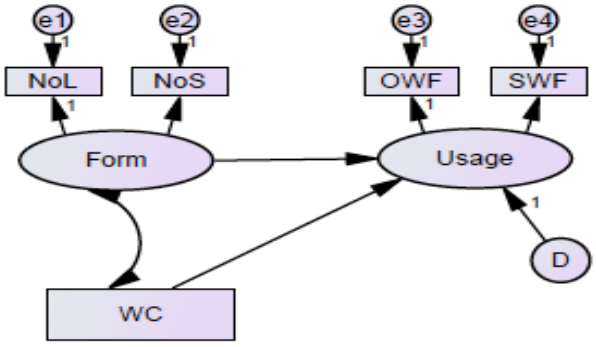


CFA EFA

Figure 3. 20 Exploratory vs. Confirmatory factor analyses (Blunch, 2012, p. 128)

Generally speaking, CFA has many advantages over EFA for testing and confirming a prior theories, for instance, the ability to specifically define and test hypothesized models, to uniquely estimate model parameters, and to compare the goodness of fit of competing models (Marsh, 1987).

Structural equation models consist of two parts: *measurement models* and *structural models*. (An example can be found in Figure 3.21.) The measurement model is a CFA model, which indicates the relationships between indicators (i.e. observed variables) and latent variables (i.e. unobserved variables); whereas the structural model specifies the relationships among latent variables (D. Harrington, 2008).



Measurement model

Structural model

Notes: The oval represents latent variable; the square represents observed variable.
D represents disturbances in structural models, and e represents measurement errors in measurement models.
Double-headed arrows represent hypothesized correlational relationships, and single-headed arrows denote regression paths.

Figure 3. 21 A full model: Measurement models plus structural models

The general structural equation model can be expressed as follows (J. Wang & Wang, 2012):

$$\eta = \mathrm{B}\eta + \Gamma\xi + \zeta$$
$$\mathrm{Y} = \Lambda_y\eta + \varepsilon$$
$$\mathrm{X} = \Lambda_x\xi + \delta$$

Where

η(eta): latent endogenous variable;

ξ(xi): latent exogenous variable;

ζ(zeta): residual terms in equations;

y: endogenous indicators;

x: exogenous indicators;

ε(epsilon): measurement errors of y;

δ(delta): measurement errors of x.

The first equation represents a structural model, in which the endogenous and exogenous variables are connected by a system of linear equations with B (i.e. the effects of some endogenous latent variables on other endogenous latent variables), with Γ (i.e. effects of exogenous latent variables on endogenous latent variables), and with ζ (i.e. regression residual terms). In the second equation, endogenous indicators are connected to endogenous latent variables by factor loadings Λ_y (lambda y), and in the third equation, exogenous indicators are connected to exogenous latent variables by factor loadings Λ_x. A structural equation model is defined by the specification of the structure of eight matrices: the above-mentioned coefficient matrices (Γ, B, $\Lambda_\mathrm{y}, \Lambda_\mathrm{x}$), and four variance/covariance matrices, namely Φ (phi, variance/covariance matrices of ξ), ψ (psi, variance/covariance matrices of ζ), θ_ε (theta-epsilon, variance/covariance matrices of ε), and θ_δ (theta-delta,

variance/covariance matrices of δ).

Maximum likelihood (ML) estimation is most commonly used for fitting function. ML calculates estimates based on that the discrepancy between sample and model variance/covariance is minimized (J. Wang & Wang, 2012). Assumptions of ML include large sample size, indicators with multivariate normality, valid specification of the model, and variables at continuous or interval level (Garson, 2012b).

There exists a misconception about SEM, that is, because SEM allows for inferring causal relationships between the variables concerned, it can be deduced that this statistical technique can be used to demonstrate causation from exogenous to endogenous variables. However, as K. R. Morrison (2009) argued, 'statistics on their own do not prove causation' (p. 165). Similarly, Kline (2011) pointed out that no statistical technique can somehow prove causality in a non-experimental design, because causality can be discerned at best only in experimental or quasi-experimental conditions.

As summarized by Kline (2011), SEM is typically conducted in the following way (see Figure 3.22 below).

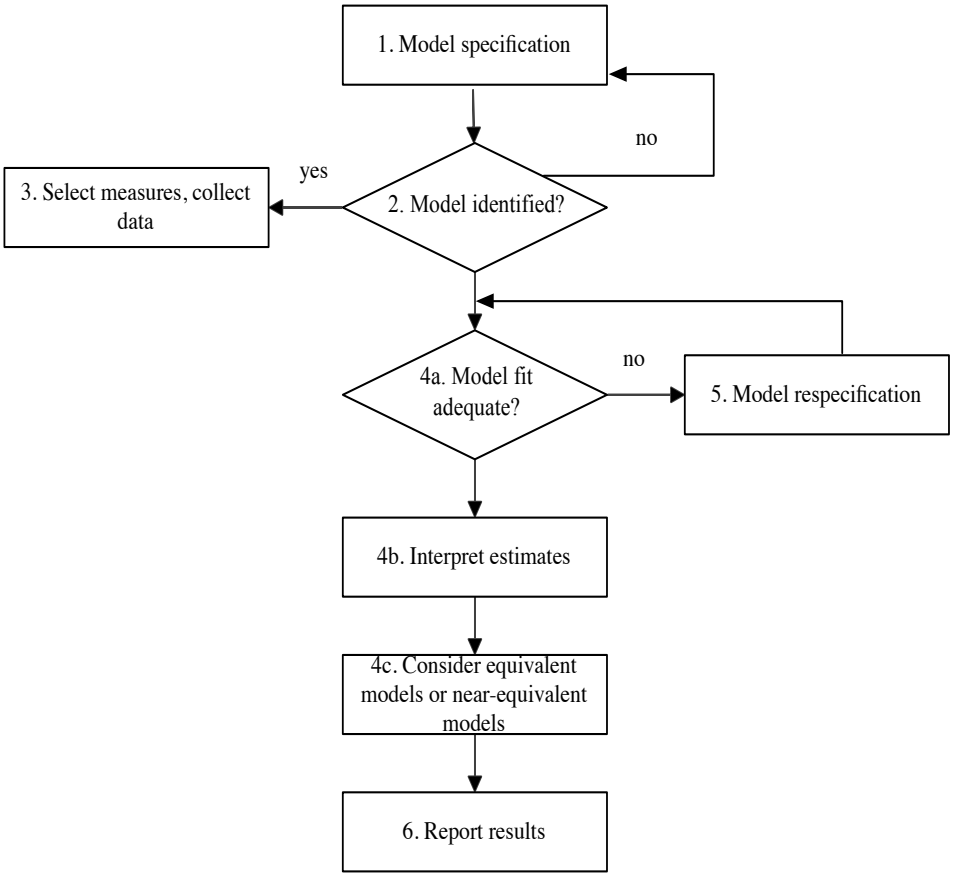


Figure 3. 22 Flow chart of basic steps of SEM (Kline, 2011, p. 92)

To be specific, conducting SEM involves the following six steps (Kline, 2011): (1) model specification: it is a process by which a model diagram is drawn based on previous theoretical and/or empirical grounds; (2) model identification: a model can be identified if it is theoretically possible to assign every model parameter a unique estimate; and if not, the model needs to be re-specified; (3) select good measures, collect the data, and screen the data; (4) model estimation: model parameters and fit of the model are estimated according to a given method of estimation; (5) model evaluation and modification: whether the model fits the data and generates interpretable results needs to be assessed. Corrections, if needed, are made to the

model with poor fit by deleting, adding and modifying parameters so as to obtain a model with acceptable fit; and (6) reporting results: analyses should be accurately and completely described in written reports.

The recommended practice is to conduct the SEM according to the following procedure, that is, validating the measurement model using the CFA, and only after the measurement model can be seen as valid and reliable, the researcher proceeds to the second step of assessing the structural model (Garson, 2012b; Kline, 2011). In the present study, therefore, this two-step approach was adopted, that is, the validity and reliability of the measurement model were assessed and confirmed by checking factor loadings, unique variances, and goodness of fit indices, before the structural model was estimated. Note that there are different types of validity and reliability, as described below.

Convergent validity refers to the extent to which different indicators are in agreement in measuring the same concept. According to Hair et al. (2005), each factor loading on its corresponding latent variable is considered significant when it is greater than .50.

As for *discriminant validity*, that is, the extent to which a construct is truly distinct from other constructs, it is measured by comparing average extracted variances (AEV) estimates for each factor with squared *interconstruct correlations* (SIC) associated with the factor. Note that AEV is another way to examine convergent validity, which is computed as the sum of the squared standardized factor

loadings divided by the number of items. AEV estimates should exceed the recommended level of .50 to indicate the existence of convergent validity (Fornell & Larcker, 1981). Additionally, discriminant validity can also be defined by low correlations among the concepts (Bagozzi, Yi, & Phillips, 1991).

As for *reliability*, Cronbach's alpha and *construct reliability* (CR) coefficients were estimated. Cronbach's alpha measures how well indicators represent a latent construct through calculating the amount of inter-correlation between all items at the same time. In contrast, CR is a summary of how well a construct is measured by its assigned indicators through computing the sum of factor loadings (λi), squared for each construct and the sum of the error variance terms for a construct (δi) using the formula expressed blow.

$$CR = \frac{(\sum_{i=1}^{n} \lambda_i)^2}{(\sum_{i=1}^{n} \lambda_i)^2 + (\sum_{i=1}^{n} \delta_i)}$$

An alpha value of .70 or above (Nunnally, 1978), and the CR estimate of .70 or above (Hair et al., 2005) suggest good reliability

Goodness of fit indices allow for determining how well the hypothesized model fits with the observed data. The models reported below were estimated using the *Chi-square statistic*, degrees of freedom and associated *p* value, together with other two indices the *Root Mean Square Error of Approximation* (RMSEA), and the *Comparative Fit index* (CFI), since these three are considered to be all important for estimating the overall model fit (B. M. Byrne, 2009). A brief introduction to each

index is presented below.

Chi-square (χ^2) *statistic*: The χ^2 statistic is considered to be the most commonly used index for assessing the overall fit of a model (Arbuckle, 2009). The χ^2 statistic is a sample based fit index, estimating 'the magnitude of the discrepancy between the sample and the model estimated variance/covariance matrixes' (J. Wang & Wang, 2012, p. 17). A non-significant *p* value (> .05) indicates an acceptable fit between a hypothesized model and the observed data. The χ^2 index, however, is highly sensitive to the sample size because the probability of a Type I error (rejecting the correct hypothesis) increases as the sample size increases (J. Wang & Wang, 2012). Moreover, the χ^2 is sensitive to the violation of the assumption of multivariate normality (J. Wang & Wang, 2012). As such, whether or not a model should be rejected is not only determined by the significance of the χ^2 statistics, but also by other model fit indices.

CFI (Bentler, 1990): CFI, which is a population-based relative fit index, compares the fit of a hypothesized model to an independent model in which no covariance among the observed variables is assumed (J. Wang & Wang, 2012). The CFI value ranges from .00 to 1.00, with higher values indicating better model fit. Conventionally, an adequate fitting model should have a CFI value ≥ .90 (B. M. Byrne, 2009); Hu and Bentler (1999), however, suggest that the cut-off point should be a stricter .95. CFI largely depends on the average size of the correlations in the

data, and the higher average correlations, the higher CFI values are (J. Wang & Wang, 2012).

RMSEA (Browne & Cudeck, 1993; Steiger & Lind, 1980): RMSEA is an absolute model fit index that takes into account the error of the approximation in the population (B. M. Byrne, 2009). For an evaluation of model fit, it is generally recommended that an RMSEA value of .05 or less represents 'close fit'; (b) an RMSEA value between .05 and .08 represents a 'reasonable fit'; (c) an RMSEA value between .08 and .10 represents mediocre fit; and (d) an RMSEA value of .10 or more represents poor fit (Browne & Cudeck, 1993; B. M. Byrne, 2009; MacCallum, Browne, & Sugawara, 1996).

Like other analytical strategies mentioned above, SEM is also sensitive to such assumptions as sample size, univariate and multivariate normality, and outliers. In particular, SEM is seen as a large sample technique, and the minimum recommended sample size, that is, 5 cases per parameter (Bentler & Chou, 1987), should be met.

In addition to reporting statistical significance, the effect size, which is a standardized measure of the magnitude of the observed effect (Field, 2013), was calculated and reported, wherever appropriate, in the present study. The cut-off criteria are presented in Table 3.6.

Table 3. 5 Measures of effect size, their uses, and cut-off criteria (based on Field, 2013; Tabachnick & Fidell, 2012)

Effect size	Interpretation of index	Common use	Small	Medium	Large
r	An estimate of the relative strength of a relationship between the two variables	Correlation	.10	.30	.50
η_p^2	An estimate of the variance explained by an effect relative to the combined value of that effect and the unexplained variance	Omnibus effect for ANOVA	.01	.06	.14
R^2	An estimate of the variance in the DV explained by the regression model	Omnibus effect for regression	.002	.13	.26

3.8 Ethical considerations

There are three cornerstones a piece of high-quality research rests on: the theoretical excellence (episteme), the technical excellence (techne), and the practical excellence (phronesis) (Oancea & Furlong, 2007, see Figure 3.23). One component underpinning the theoretical excellence is propriety, which involves 'the degree to which research conforms to legal requirements and to ethical principles' (i.e. commonly known as ethical codes, Oancea & Furlong, 2007, p.130).

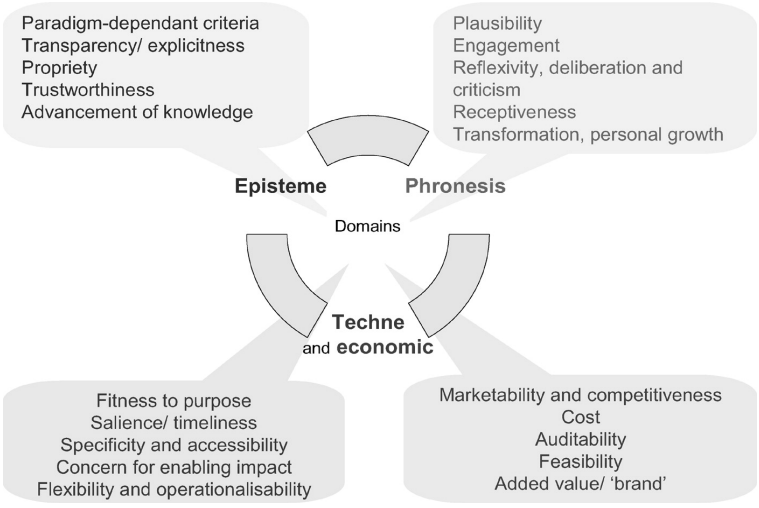


Figure 3. 23 Domains of quality (Oancea & Furlong, 2007, p. 133)

Ethical considerations are particularly significant to educational research. Therefore, they need to be deliberated throughout the whole research process, including research design, data collection, data management, data analysis, and the write-up of a research report. As mentioned by Pring (2005), there are two main principles that deserve close attention: the principle showing respect for those being researched, and the principle of searching for and reflecting the reality as far as possible. As to the specific rules of conduct, the present study complied with the ethical codes set forth by British Educational Research Association (2011) in the following ways.

Compliance with institutional procedures. The research proposal had been submitted and approved by Central University Research Ethics Committee (CUREC) before this study was officially conducted.

Respect. The researcher showed respect for all the participants at all times. For instance, when the participants were available for performing the experiment was scheduled at least a week ahead. A reminder via text message was sent to them two days before the experiment. The researcher carefully selected gifts for the participants to appreciate their participation and contribution to this study.

Disclosure. The main aims and procedures of this research were made clear to all the participants via letter or email. They were enabled to have access to any data in relation to them. The procedures and results of experiments were reported in a clear and transparent manner.

Voluntary informed consent. Opt-in consent was obtained from the student participants and the children's parents. They were clearly informed that they had the right to leave this study at any time if they wanted to, and that they did not have to give a reason for so doing.

Confidentiality and anonymity. All the information provided by the participants was kept confidential. Pseudonyms were used to replace their real names. Each participant was assigned a unique identification number (e.g. Number/Gender/Month/Year: 03MAAPR2013) in the computer records. All data were kept in a secure file to which only the researcher had access over the course of this research.

3.9 Validity, reliability, and generalizability

3.9.1 Validity

High validity and reliability are widely regarded as hallmarks of research quality. As defined by Trochim and Donnelly (2008), validity refers to the strength of conclusions, inferences, or propositions. According to the taxonomy proposed by Campbell and Stanley (1963), there are four types of validity, namely, internal validity, external validity, construct validity, and statistical conclusion validity.

Internal validity. It refers to the confidence one has that the differences in results are attributed to experimental conditions but not to confounders (Cohen et al., 2007). For instance, in the present study, internal validity was secured by ensuring that the observed RL differences were truly due to stimulus conditions rather than to other

variables that might have interfered with the research process.

External validity. It pertains to the extent to which the results of a study can be applicable to a wider population, settings, times, or situations (Cohen et al., 2007). For instance, in the present study, external validity was enhanced by an adequate sample size and appropriate sampling strategies used. Also, the provision of a detailed description of the sample allowed for comparing them with other people within the defined population.

Construct validity. It concerns the degree to which higher-order constructs that research operations represent actually measure what a theory purports to measure (Anastasi & Urbina, 1997). In the present study, great effort was made to enhance construct validity by choosing independent and dependent variables carefully so as to adequately and appropriately capture the conceptual variables of interest as far as possible.

Statistical conclusion validity. It refers to the extent to which statistical tests are used properly and to which conclusions can be drawn confidently (Ary, Jacobs, Razavieh, & Sorensen, 2010). In the present study, for example, statistical conclusion validity was secured by checking the power of a statistical test, by evaluating statistical test assumptions, and by assessing the reliability of the measures used.

An alternative way to have confidence in any conclusions reached is to minimize, if not eliminate, threats to validity. Those four types of validity described above can be threatened for various reasons, as outlined in Figure 3.24. Throughout this study,

the researcher was alert to all possible threats and took specific steps, if needed, to attenuate them, for instance, assigning participants randomly to stimulus sequences to control for the threats of history, employing carefully-considered sampling strategies to attenuate gender bias, and researching the literature extensively to reduce the threat caused by inadequate construct explanations.

Ideally, every effort should be made to attenuate all possible threats as far as possible; however, in reality, attempts made to control one type of validity may increase the possibility of threatening another type of validity. A case in point is that the tighter experimental controls are introduced to a study (i.e. higher internal validity), the more likely it is that its results are of limited usefulness because of the presence of relatively higher artificiality in the study (i.e. lower external validity). What this indicates is that researchers should consider and weigh up various threats and validity issues so as to make decisions and take actions as optimal as possible. Also, research should be viewed as a cumulative enterprise, and no such thing as a piece of perfect research really exists. From a dilemmatic point of view, in the words of McGrath (1982, p. 70), 'All research strategies and methods are seriously flawed, often with their very strengths in regard to one desideratum functioning as serious weaknesses in regard to other, equally important, goals'.

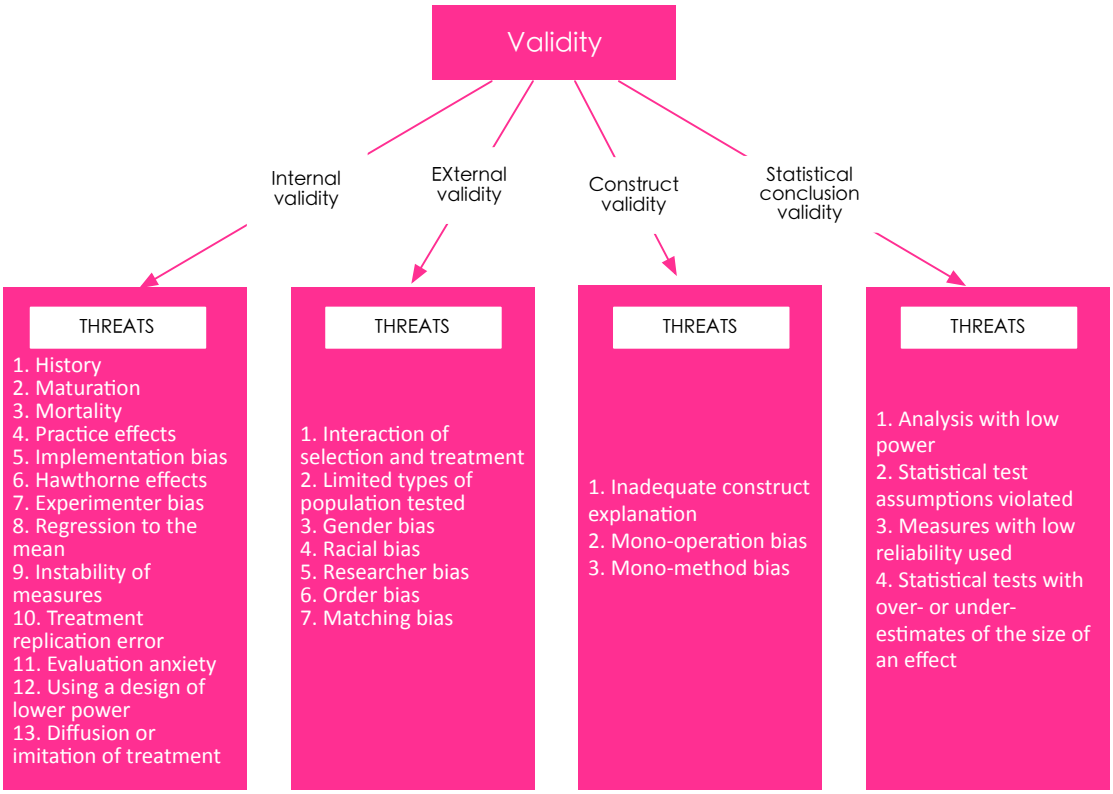


Figure 3. 24 Types of validity and threats to validity (based on Ary et al., 2010; Cohen et al., 2007; Goodwin, 2009)

3.9.2 Reliability

Reliability refers to an estimate of consistency or stability of concepts being examined, in other words, the extent to which a study's findings can be replicated if it is conducted with a similar sample once again (Cohen et al., 2007). Consider throwing darts at a board (see Figure 3.25)—getting all the darts in the same spot means reliability, and getting a dart on the target means validity. According to Cohen et al. (2007), conceivably, reliability is a necessary but not sufficient condition for securing validity, while validity is a sufficient but not necessary condition for establishing reliability.

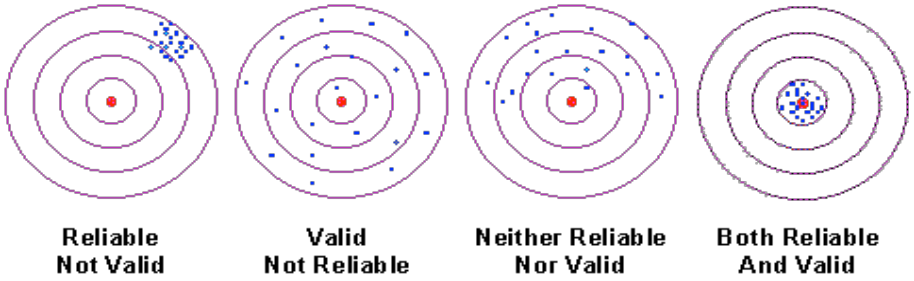


Figure 3. 25 Validity and reliability: Four possibilities

There are two types of reliability, *internal reliability* and *external reliability*. The former refers to the extent to which similar conditions have been applied in collecting and analyzing the data, and in interpreting the results. In the present study, specific steps were taken to secure internal reliability; for instance, the same instructions were given to all the participants, the experimental procedures were standardized, and the data were coded and analyzed in a consistent way.

External reliability concerns the likelihood a research study conducted by another researcher in similar settings with similar samples would draw similar results. To enhance external reliability, in the present study, full details of the design and methods, participants, and experimental procedures were provided.

In brief, throughout the whole process of this research, every choice made, procedure followed, and action taken were all thoroughly considered so as to enhance its validity and reliability, as high as possible.

3.9.3 Generalizability

Generalizability has traditionally been viewed as an important property of a piece of

quantitative research. By definition, generalizability refers to the extent to which the statements grounded in local contexts can apply beyond the time and space in which evidence for those statements are made (D. Byrne, 2002).

There are two types of generalizability, *theoretical generalization* and *empirical generalization*. The former involves the development of concepts, theories and models based on a set of evidence or the data collected and analyzed. In the present study, effort has been made to propose hypothesized models that might be generalizable to all 'like' cases. The second type of generalizability, empirical generalization, concerns the extension of those findings drawn from a sample to a wider population. The way the sample is selected, in most cases, has much impact upon the generalizability of a study. In light of this, sampling strategies were carefully thought out in the present study so as to ensure that the sample was representative of the group to which generalizations could be made within the constraints of resources, personnel, and costs.

3.10 Summary: Chapter 3

As shown in Figure 3.26, this study was conducted in a research question-driven and epistemology-guided fashion. First of all, what issues to be addressed was decided on the basis of a critical appraisal of previous literature. After this, research questions were proposed, and then linked back to the research traditions outlined in Section 3.1. Clearly, the research questions that were posed in the present study could be answered by following a positivist paradigm, because correlations, predictions, and

cross-condition comparisons were examined. The next step was to choose the most appropriate methodology and methods. As indicated previously, positivist research is typified by the use of quantitative methods, and the final decision on which specific one to use were made based on research circumstances and resource availability. The main methods adopted in the present study included response latency-based experiments, assessments, and quantitative questionnaires.

In the third step, a well-considered systematic plan was devised to specify every component and step that were essential to each sub-study, including its aims, participants, task types, task orders and stimuli. Next, taking into account the aims of this research study and characteristics of the participants, research materials, assessments and questionnaires were then developed. All these instruments were piloted with a group of learners who were demographically similar to the main study participants, and modifications were made when needed. When the main stage of the fieldwork started, participants were recruited according to appropriate sampling frames. The sampling choice was made based largely on the extent to which the sample was representative of the population and on several practical considerations. After the desirable sample size was obtained, this research was conducted according to what had previously planned. After the data were collected, the next step involved a preliminary data analysis to gain insight into important characteristics of the data and to prepare the selection of appropriate statistical methods for formal analysis (see Section 3.7). In what follows, inferential analyses were performed to address each

research question.

All these research steps were ethically justifiable by respecting and protecting participants as far as possible. The present study has made every effort to secure methodological rigor in order to produce valid, reliable and generalizable results and findings.

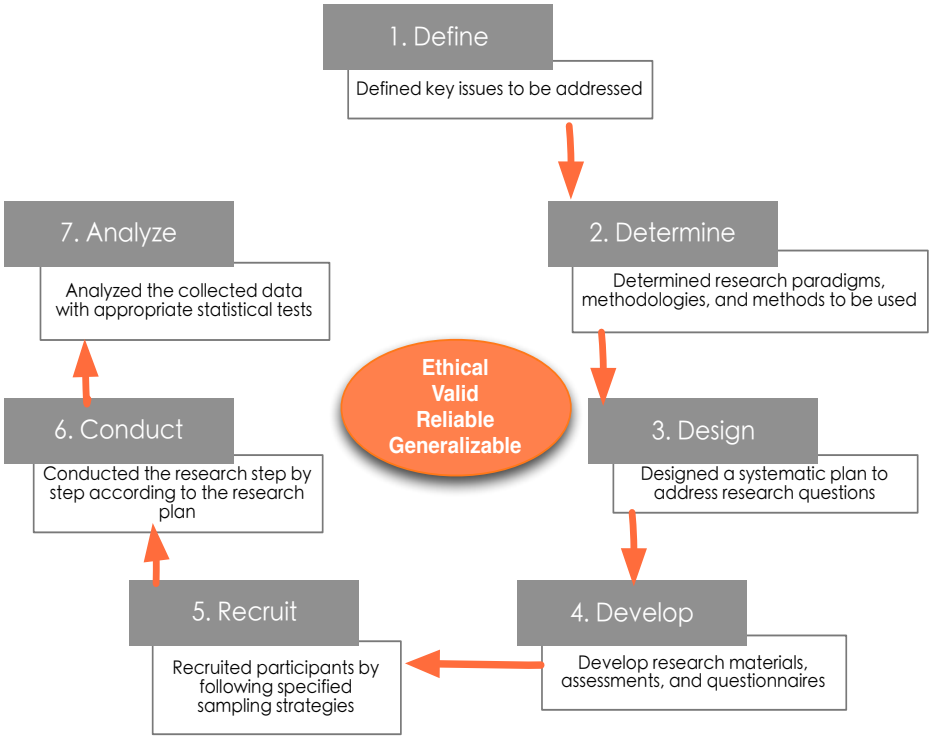
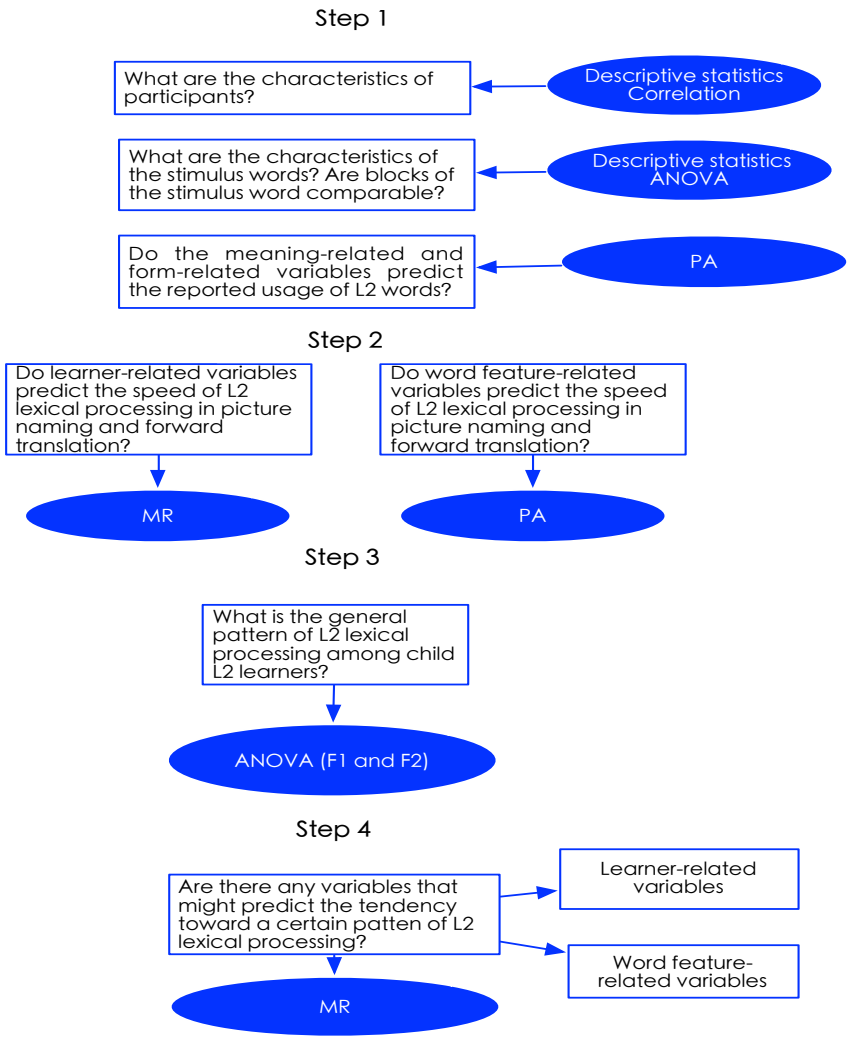


Figure 3. 26 Flowchart of the experiment

Chapter 4 Data Analyses and Results: Child Participants

This chapter first profiles the sample and stimulus words by using descriptive statistics and correlation analysis. After that, the predictive effects of learner-related and lexical-related variables on L2 lexical processing are reported. The following sections address the pattern of L2 lexical processing among the child participants, and the factors that significantly predicted the tendency toward a certain pattern of processing. Figure 4.1 outlines the analytical strategies employed and the procedures followed in analyzing the child participant data. The analyses reported below were conducted using the statistical software SPSS v.20 and IBM SPSS AMOS (Analysis of MOment Structures) v.20.



Notes: ANOVA: analysis of variance; MR: multiple regression; SEM: structural equation modeling; PA: path analysis.

Figure 4. 1 Overview of research questions and methods of data analysis: Child participants

4.1 Data Screening

The child participants responded to 111 stimulus words, and their RLs were recorded. Initially, the data set contained a total of 4329 data points. Before performing statistical analyses, the entire data set was screened for incorrect and omitted responses, outliers, and those participants and stimulus items with an exceptionally high error rate.

First, incorrect and omitted responses that were recorded manually during experiments were discarded for latency analysis. This deletion resulted in a loss of 14.62 % of the data points in PE, of 14.90% in FT, and of 3.33% in PC.

Second, as suggested by Jiang (2012), a participant with an error rate higher than 20% needs to be excluded from further analysis. Since there was no error rate beyond the cut-off criteria, all the participants were retained for analysis. As regards individual items, the cut-off error rate was set at 40% because this is the percentage to which researchers should be alert (Jiang, 2012). By this criterion, six stimulus words (i.e. sweater, moon, shorts, forest, boots, and star) were deleted. After this, the loss of data in PE, FT, and PC became 10.18 %, 10.48 % and 2.93%; and the number of lexical items entered into analysis was 105.

Lastly, outliers were identified and excluded from further statistical analyses. The existence of outliers might be due to an accidental pressing of a voice key, and due to learners' guessing, fatigue, and a lack of attention. Following the common practice recommended by Jiang (2012), as for the low cut-off, those data points that were

below 350 ms were considered as outliers, and the high cut-off was set at 3 standard deviations (SDs) beyond the mean of each participant's RLs. Note that the low and high cut-offs were set on the basis of different criteria, because it was found that the standard scores of some extremely short RLs were still likely to fall between -3.00 SDs and .00.

Prior to the completion of data screening, the number and percentage of valid responses, and those of deleted responses are summarized in Table 4.1. The missing values were replaced by the means of a participant when by-participant analysis was performed or by the means of an item within the stimulus condition when by-item analysis was performed. It is worth mentioning that all the analyses reported below followed the conventional way of treating psycholinguistic data, that is, RLs, RAs, and word-feature ratings were averaged to generate a summary score for each participant and for each lexical item. It was these values that were entered into formal analysis instead of the entire data points.

Table 4. 1 Summary of valid and deleted responses

	Valid		**Deleted**	
	n	**%**	**n**	**%**
PE	1211	88.72	154	11.28
FT	1208	88.50	157	11.50
PC	1319	96.63	46	3.37

4.2 Characteristics of participants and stimulus words

4.2.1 Sample characteristics

The total sample consisted of 39 fifth-grade pupils who had been learning EFL at a public primary school for two years and a half. The researcher accessed the Registrar's student files to obtain the participants' demographic information and academic performance records, for instance, gender, age, and fourth-grade exam scores. In addition, four language-related assessments were administered to the participants, including the vocabulary breadth test, the vocabulary depth test, the YLE reading test, and the YLT listening test. The summary statistics of these variables are presented in Table 4.2. Details about these assessments have been provided in section 3.6.3.

Table 4. 2 Demographics, and language and academic backgrounds of child participants (n = 39)

Demographics				
	Frequency		Percentage	
Gender				
	Male	19		48.7
	Female	20		51.3
Age				
	10	27		69.2
	11	12		30.8
Living area				
	Urban	24		61.5
	Rural	15		38.5
Language backgrounds				
	Mean	SD.	Minimum	Maximum
WASI-III (MR)	60.49	3.66	52	68
Vocabulary breadth	202.08	48.66	115	305
Vocabulary depth	47.38	4.29	40	59
English course scores	89.46	4.54	82	100
Chinese course scores	93.00	2.73	86	97
Math course scores	90.49	5.67	75	100
YLE (Reading)	13.31	2.56	9	22
YLE (Listening)	15.41	3.42	7	19

Notes: MR: matrix reasoning; WASI: Wechsler Abbreviated Scale of Intelligence; YLE: Cambridge Young Learners' English Tests (starter level). The maximum total score for each test is as follows: 500 (vocabulary breadth); 72 (vocabulary depth); 100 (English course); 100 (Chinese course); 100 (math course); 25 (YLE Reading); and 20 (YLE Listening).

The child sample aged from 10 to 11 years old (Mean = 10.31, SD = .47). The proportion of the male participants was 48.7%, while the female participants took up 51.3% of the total. As regards living areas, 61.5 % of the child participants came from

urban areas, and 38.5% from rural areas. The average WASI matrix reasoning score was 60.49 (SD = 3.66). The child participants had a vocabulary size averaging approximately 202 words (SD = 48.66), and their scores on vocabulary depth ranged from 40 to 59, with the mean score being 47.38 (SD = 4.29). The average scores on the English course, the Chinese course, and the math course were 89.46 (SD = 4.54), 93.00 (SD = 2.73), and 90.49 (SD = 5.67) respectively.

The correlation between each pair of test scores is presented in Table 4.3. Clearly, the scores on vocabulary breadth and on vocabulary depth were both significantly and positively correlated with English course grades, as well as with the YLE reading and listening scores. Moreover, English course grades were found to be significantly and positively correlated with Chinese and math course grades.

Table 4. 3 Pearson correlations (r values) between test scores: Child participants (n = 39)

	1	2	3	4	5	6	7	8
1. WASI (MR)	1	—	—	—	—	—	—	—
2. Vocabulary breadth	.23	1	—	—	—	—	—	—
3. Vocabulary depth	.24	.32	1	—	—	—	—	—
4. English scores	-.03	.41*	.39*	1	—	—	—	—
5. Chinese scores	.17	.24	.42**	.49**	1	—	—	—
6. Math scores	.27	.19	.26	.49**	.27	1	—	—
7. YLE (Reading)	.21	.50**	.17	.33*	.28	.20	1	—
8. YLE (Listening)	.38*	.60**	.41**	.60**	.43**	.38*	.62**	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

4.2.2 Characteristics of stimulus words

The present study consisted of three blocks of 35 stimulus words, and each block contained picturable words selected from ten semantic categories, namely: animal (n = 4), nature (n = 3), fruits and vegetables (n = 4), parts of human body (n = 3), color (n = 3), household items (n = 4), clothing (n = 3), actions (n = 4), number (n = 4), and people (n = 3). The values and ratings of each word's lexical variables were collected either through participant-rated questionnaires or directly from psycholinguistic databases (see Appendix 4.1 for detailed results). As for participant-rated estimates, the rate and re-rate reliability on the repeated 10 items were all found to be high (all r values > .77, and $p < .05$).

Comparisons between three blocks of stimuli were made in terms of the lexical variables being examined so as to determine the extent to which the stimulus blocks were comparable. When the statistical assumptions were met, independent measures ANOVAs, an analytical strategy introduced in Section 3.7.1, were used to compare the mean values or ratings of word features across three blocks. The normality of the variable of NoRS, SND, NoL, NoP, and NoS was not met (i.e. skewness or kurtosis values > + 2.00 or < - 2.00); thus, a log-transformation was applied to normalize their distributions. The descriptive statistics of each lexical variable are presented in Table 4.4. The ANOVA results are summarized in Table 4.5 (see Appendix 4.2 for statistical details). In general, there was no statistically significant difference observed between the blocks in terms of all the lexical variables. What this suggests is that all

blocks of stimuli were comparable in meaning-related, form-related, and usage-related variables, and such a similarity ensured the internal validity of the present study.

Table 4. 4 Word features of each block: Child participants (n = 105)

Variables		n	Mean	SD.	Minimum	Maximum
WC (average)	Block 1	35	1.76	.20	1.47	2.24
	Block 2	35	1.79	.17	1.54	2.24
	Block 3	35	1.74	.14	1.56	2.11
WC (total)	Block 1	35	8.82	1.02	7.33	11.22
	Block 2	35	8.95	.84	7.72	11.22
	Block 3	35	8.68	.68	7.81	10.53
WT	Block 1	35	3.52	.49	2.00	4.54
	Block 2	35	3.67	.45	2.46	4.86
	Block 3	35	3.53	.47	2.43	4.43
Log-SND	Block 1	35	.85	.57	0.00	2.31
	Block 2	35	.86	.67	0.00	2.45
	Block 3	35	.83	.60	0.00	2.51
Log-NoRS	Block 1	35	.54	.34	0.00	1.34
	Block 2	35	.55	.22	.30	1.08
	Block 3	35	.49	.28	0.00	1.23
Log-NoL	Block 1	35	.68	.12	.48	.90
	Block 2	35	.69	.12	.48	1.00
	Block 3	35	.67	.13	.48	.95
Log-NoP	Block 1	35	.56	.10	.30	.85
	Block 2	35	.57	.14	.30	.85
	Block 3	35	.57	.13	.30	.85
Log-NoS	Block 1	35	.07	.14	0.00	.48
	Block 2	35	.12	.17	0.00	.48
	Block 3	35	.08	.14	0.00	.48
SWF	Block 1	35	3.31	.35	2.64	4.00
	Block 2	35	3.40	.30	2.94	4.22
	Block 3	35	3.37	.26	2.69	3.83
AoA	Block 1	35	9.81	.83	9.00	11.50
	Block 2	35	9.89	.85	9.00	11.50
	Block 3	35	9.93	.87	9.00	11.50

Table 4. 5 Comparing stimulus words between blocks: Child participants (n = 105)

Variables	df	Mean square	F	Sig.	η_p^2
WC	2	.02	.83	.44	.020
	102	.03			
WT	2	.32	1.06	.35	.020
	102	.14			
Log-SND	2	.000	.001	.99	.000
	102	.37			
Log-NoRS	2	.03	.35	.71	.010
	102	.08			
Log-NoL	2	.003	.20	.82	.004
	102	.02			
Log-NoP	2	.002	.08	.93	.001
	102	.02			
Log-NoS	2	.02	.63	.54	.010
	102	.03			
SWF	2	.07	.75	.48	.010
	102	.10			
AoA	2	.12	.16	.85	.003
	102	.72			

In addition to the comparisons made on the stimulus words across blocks, the visual complexity of pictorial stimuli was also compared across blocks. The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the visual complexity of the pictures, F (2, 102) = .31, p = .73, η_p^2 = .01. Moreover, all the three blocks were matched with regard to the usage frequency of L1 words according to the Language Corpus System of Modern Chinese Study (LCSMCS, www.dwhyyjzx.com/cgi-bin/yuliao/), F (2, 102) = .97, p = .38, η_p^2 = .02.

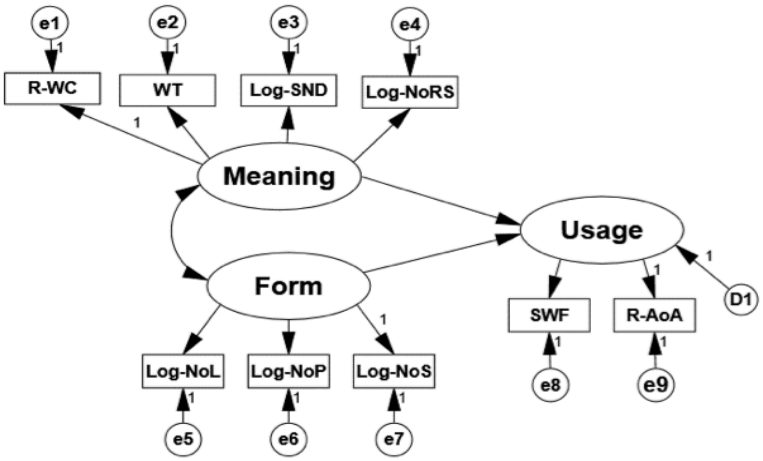
4.2.3 Relationships between meaning-related variables, form, and usage of L2 words: Child L2 learners

This analysis answers the following research question:

Do the meaning-related and form-related variables predict the reported usage of L2 words?

4.2.3.1 Hypothesized SEM model

The hypothesized model is presented in Figure 4.2 below. In this model, three latent factors were hypothesized: a Meaning factor (word concreteness, word typicality, semantic neighborhood density, and the number of related senses), a Form factor (the number of letters, the number of phonemes, and the number of syllables), and a Usage factor (subjective word frequency and age of acquisition). It should be noted that the subjective frequency counts used in the present study allowed for inferring how often the child participants were likely to use the target words although there was a lack of direct measurement of their usage of words in real life due to the involvement of a relatively large sample, as well as due to the fact that the EFL learners in China relied heavily on receptive language skills.



Note: R: recoded.

Figure 4. 2 Hypothesized SEM model: Do Meaning and Form predict the Usage of L2 words among child L2 learners?

It was hypothesized that the factors Meaning and Form both predict the factor Usage. The theoretical underpinning of this hypothesized model was partly derived from the Three Stage Vocabulary Acquisition model (Jiang, 2004), which specifies the route of L2 lexical development in instructional settings; therefore, there appears to be a temporal order of the factors concerned. Moreover, as reviewed in section 2.4.3, usage is considered to be a behavioral outcome, which is under the influence of semantic and form variables.

Following convention, Meaning and Form were allowed, initially, to freely covary, and the errors associated with each lexical variable were not correlated. Note that the following variables were found or supposed to underpin their corresponding latent variable in an opposite direction as compared to the majority of other variables within the same construct, namely: WC and AoA. The ordinal variable WC was recoded, and the interval variable AoA was standardized using a Z-score

transformation and then recoded. A correlation matrix of lexical variables is presented in Table 4.6.

Before performing formal analyses, the assumptions of SEM were checked, including the sample size, missing data, normality and linearity, outliers, and multicollinearity and singularity (see Appendix 4.3 for detailed procedures). The minimum sample size was estimated to be 135 according to Bentler and Chou's recommendation (1987) that a ratio of 5 cases or observations per free parameter is necessary under normal distribution. The sample size of the present study (n = 105) was not larger than the required 135; therefore, composite variables were computed afterwards so as to reduce the complexity of the hypothesized model and to reduce the size required for a reliable model. There were no missing data in the data set. There was no serious violation noted except that one multivariate outlier was identified and removed from further analysis ($p < .001$). Additionally, according to Tabachnick and Fidell (2012), highly correlated independents can lead to a failure of achieving statistical significance of regression coefficients, and a failure of model converge on a solution. A scan of the correlation matrix of the indicants (see Table 4.6 below) showed that none of the correlation coefficients exceeded 0.90, indicating the absence of multicollinearity and linearity.

Table 4. 6 Correlation matrix of word feature-related variables: Child participants (n = 105)

	1	2	3	4	5	6	7	8	9
1. R-WC	1								
2. WT	.17	1							
3. Log-SND	.44**	.15	1						
4. Log-NoRS	.02	-.09	.06	1					
5. Log-NoL	-.23*	-.09	-.23*	-.46**	1				
6. Log-NoP	-.27**	-.02	-.28**	-.34**	.76**	1			
7. Log-NoS	-.31**	.12	-.17	-.41**	.67**	.72**	1		
8. SWF	-.16	.29**	.03	.30**	-.28**	-.21*	-.01	1	
9. R-AoA	.16	.20*	.26**	-.29**	-.08	-.07	-.01	-.11	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

4.2.3.2 Measurement model

Before creating the composite variables to replace the latent variables, the measurement model that was illustrated in Figure 4.3 below was estimated in terms of its validity and reliability first. A CFA with maximum likelihood (ML) estimation method was performed using IBM SPSS AMOS v.20. The ML, which is a normal theory based method, was appropriate for use here as no serious violation of statistical assumptions was noted.

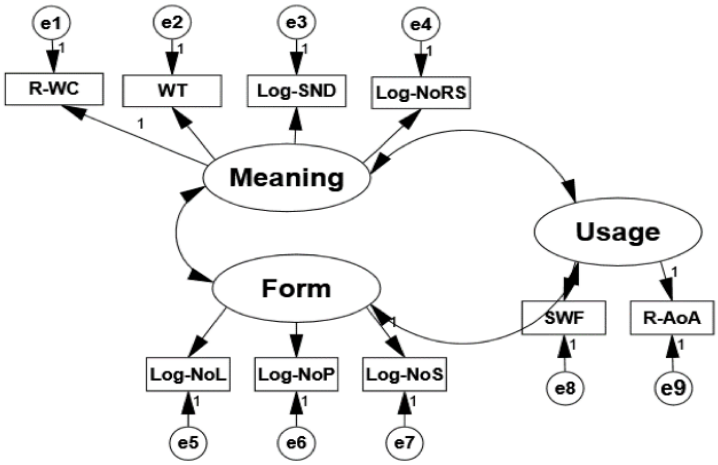


Figure 4. 3 Measurement model: Meaning, Form and Usage of L2 words among bilingual children

Convergent validity is checked against the criteria that each factor loading on its corresponding latent variable should be greater than .50 (Hair et al., 2005). The result suggested that the factor loadings of NoRS and WT were less than .50 (see Appendix 4.4 for the output of IBM SPSS AMOS output). It was therefore decided to treat meaning-related variables as separate variables rather than to create a composite variable. Contrary to our expectation, the variable of recoded AoA was not positively

correlated with SWF, suggesting that early-acquired words were not used more frequently than late-acquired words according to the child participants' estimates. Therefore, SWF was excluded from the specified model on the grounds that (1) the convergent validity of AoA and SWF was not sufficiently high to create a composite variable; (2) rating SWF accurately might be beyond children's abilities to do so; and (3) SWF explained less variance of the outcome variable relative to AoA.

As for internal reliability, *Cronbach's alpha* was estimated. An alpha value of .70 or above (Nunnally, 1978) suggests good internal reliability. The result indicated that the internal reliability of the construct of Form was satisfactory, with $\alpha = .88$.

Composite variables were computed to replace the latent factor Form based on the formula put forward by Joreskog and Sorbom (1989). Factor loadings were multiplied by the corresponding indicator item raw score for each variable and then summed. The formula used was:

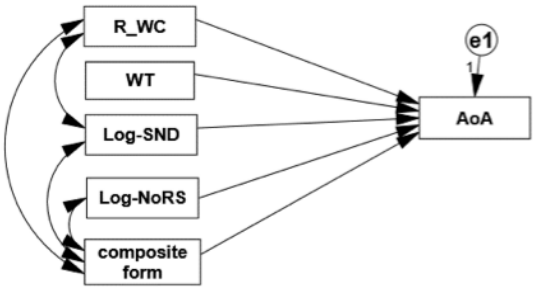
$$\text{Composite variable} = \frac{\sum_{i=1}^{n} w_i * X_i}{\sum_{i=1}^{n} w_i}$$

Where w_i is the factor loading and X_i is the item score.

The use of this formula provides a more realistic interpretation of the data as it appropriately weighted for the contribution of each item to the overall construct they represent. These composite variables were then included in the re-specified path model (see Figure 4.4 below).

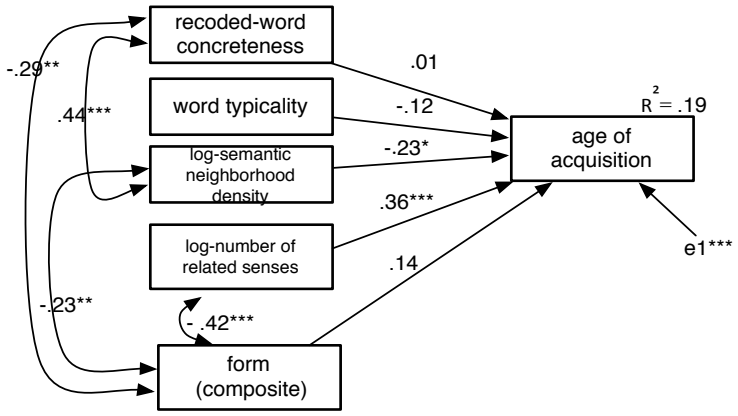
4.2.3.3 Re-specified path model

The above modifications were made to the re-specified path model presented below.



Notes: R: recoded; and C: composite.
Figure 4. 4 Re-specified path model: Form and Usage of L2 words among bilingual children

Path analysis using the ML method was conducted to estimate the overall fit of the re-specified path model. The model fitted the data perfectly, $\chi^2(6, N = 104) = 5.67$, $p = .462$, CFI = 1.000, RMSEA = .000. The model with standardized coefficients is diagrammed in Figure 4.5 below. The output of IBM SPSS AMOS is presented in Appendix 4.5.



Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$.
Figure 4. 5 Re-specified path model: Do meaning and form predict the usage of L2 words among child L2 learners?

As can be seen from Figure 4.5, there are two variables that significantly predicted AoA: one is semantic neighborhood density (standardized coefficient = -.23, $p < .05$), and the other is the number of related senses (standardized coefficient = .36, $p < .001$). The R^2 value was .19 (an effect size between medium and large), indicating that 19 per cent of the variance in AoA was accounted for by the composite variable form, and by WC, WT, SND, and NoRS.

4.2.3.4 Interim summary: Relationships between lexical variables

This analysis showed that early-acquired words had fewer related senses and denser semantic neighborhoods as compared to late-acquired words among child L2 learners. What this suggests is that young children's vocabulary learning in input-limited contexts usually starts with unambiguous words that have more semantic neighbors. Although word typicality negatively predicted age of acquisition, its effect did not reach a statistically significant level, indicating that L2 words that are learned in instructional settings are not arranged in complete compliance with the degree of word typicality. Contrary to our expectation, the contribution of word concreteness to the age of acquisition was not significant, probably because most of the words that the child participants had learned were concrete or nearly concrete, which might make it difficult to observe any significant word concreteness effects. Analysis on the predictive effect of form on the variable of usage showed that early-acquired words did not necessarily have simpler forms.

4.3 The predictive effects of learner- and word feature-related variables on L2 lexical processing among child L2 learners

Standard MR was carried out to examine the predictive effects of learner-related variables on the response latency of L2 lexical processing in picture naming and forward translation. It should be pointed out that MR, rather than SEM, was used, because there was a lack of theoretical and empirical grounds based on which a model can be hypothesized to specify which and in what manner the predictors of interest might be related to the outcomes. The predictive effects of learner-related variables on L2 lexical processing are reported below.

4.3.1 Predictive effects of learner-related variables on L2 lexical processing among child L2 learners

This analysis addresses the following research question:

Do learner-related variables predict the speed of L2 lexical processing in picture naming and forward translation among child L2 learners?

Before formal analyses were performed, descriptive statistics for all the learner-related variables, and the RL of PE and that of FT were examined, including their means, SDs, and frequency histograms wherever appropriate. Also, correlation matrices between continuous predictors and the outcome variable were calculated (see Table 4.7 below). The strength of associations between nominal predictors and the outcome variable was measured using Lambda (see Appendix 4.6). The results suggested that only two predictors (i.e. vocabulary size and Chinese course scores)

were significantly and negatively correlated with PE response latencies, and three predictors (i.e. vocabulary size, reading proficiency level, and listening proficiency level) were significantly and negatively correlated with FT response latencies. These results indicated that those participants that had larger vocabulary sizes and achieved higher Chinese course scores responded to stimulus pictures in English faster than those that had smaller sizes and lower Chinese course scores. Additionally, those participants that had larger vocabulary sizes, and higher reading and listening proficiency levels performed faster in forward translation relative to those that had smaller sizes and lower proficiency levels.

The variables that were significantly correlated with the outcome variable were included in the regression model. Moreover, the assumptions for conducting MR were checked (see Section 3.7.1 for required assumptions), including the sample size, missing data, normality, linearity, and homoscedasticity of residuals, outliers, and multicollinearity and singularity (see Appendix 4.7 for statistical details), with no serious violation noted.

Table 4. 7 Correlations between learner-related variables, and PE and FT performance in child L2 learners (n = 39)

A. Pearson's correlations: PE response latencies

Variables	1	2	3	4	5	6	7	8	9	10
1. PE	1	—	—	—	—	—	—	—	—	—
2. Age	-.16	1	—	—	—	—	—	—	—	—
3. WASI (MR)	-.16	.02	1	—	—	—	—	—	—	—
4. Vocabulary size	-.40*	.08	.23	1	—	—	—	—	—	—
5. Vocabulary depth	-.12	.31	.24	.32*	1	—	—	—	—	—
6. English course scores	-.20	.49**	-.03	.41**	.39*	1	—	—	—	—
7. Chinese course scores	-.41*	.43**	.17	.24	.42**	.49**	1	—	—	—
8. Math course scores	.01	.31	.27	.19	.26	.49**	.27	1	—	—
9. YLE (Reading)	-.31	.23	.22	.50**	.17	.33*	.28	.20	1	—
10. YLE (Listening)	-.29	.25	.38*	.60**	.41**	.60**	.43**	.38*	.62**	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

Table 4. 7 Correlations between learner-related variables, and PE and FT performance in child L2 learners (Continued)

B. Pearson's correlations: FT response latencies

Variables	1	2	3	4	5	6	7	8	9	10
1. FT	1	—	—	—	—	—	—	—	—	—
2. Age	-.10	1	—	—	—	—	—	—	—	—
3. WASI (MR)	-.29	.02	1	—	—	—	—	—	—	—
4. Vocabulary size	-.48**	.05	.23	1	—	—	—	—	—	—
5. Vocabulary depth	-.19	.31	.24	.32*	1	—	—	—	—	—
6. English course scores	-.23	.49**	-.03	.41**	.39*	1	—	—	—	—
7. Chinese course scores	-.15	.43**	.17	.24	.42**	.49**	1	—	—	—
8. Math course scores	-.19	.31	.27	.19	.26	.49**	.27	1	—	—
9. YLE (Reading)	-.35*	.23	.21	.50**	.17	.33*	.28	.20	1	—
10. YLE (Listening)	-.36*	.25	.38*	.60**	.41**	.60**	.43**	.38*	.62**	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

4.3.1.1 Results: Predictive effects of learner-related variables on L2 lexical processing

Table 4.8 and Table 4.9 display the correlations between each pair of variables, the unstandardized regression coefficients (*B*) and intercept, the standardized regression coefficients (β), the semi-partial correlations (sr_i^2), R, R^2, and adjusted R^2. The adjusted R^2 were also reported because it corrects for the sample size and for the number of variables included in the model.

In the regression analysis of learner-related variables that might predict PE response latencies (see Table 4.8 below), the R (.51) was significantly different from zero, F (2, 36) = 6.42, $p < .01$, with R^2 at .26 (a large effect size). The adjusted R^2 value of .22 indicated that 22% of the variability in PE response latencies was predicted by the included learner-related variables. The size and direction of the relationships suggested that those who had larger vocabulary size were likely to process L2 lexical items faster in PE ($\beta = -.32$, $p < .05$), and those who had higher Chinese course scores were likely to process L2 lexical items faster in PE ($\beta = .33$, $p < .05$). Between these two significant predictors, it was found that the Chinese course grades served as a more powerful predictor, as indicated by the squared semi-partial correlations.

In the regression analysis of learner-related variables that might predict FT response latencies (see Table 4.9 below), the R (.50) was significantly different from zero, F (3, 35) = 3.87, $p < .01$, with R^2 at .25 (an effect size between medium and large). The adjusted R^2 value of .19 indicated that nearly 20% of the variability in FT

response latencies was predicted by the learner-related variables included. The size and direction of the relationships suggested that those who had larger vocabulary size ($\beta = -.38$, $p < .05$), and who had higher reading scores ($\beta = -.13$, $p = .50$) and higher listening scores ($\beta = -.06$, $p = .78$) were likely to produce L2 lexical items faster in FT. Among these predictors, only vocabulary size served as a significant predictor.

4.3.1.2 Interim summary: Learner-related variables and L2 lexical processing

A cross-task comparison showed that it was the size of vocabulary that mattered most in predicting the child participants' performance of English picture naming and forward translation. Additionally, Chinese course grades served as a significant predictor in picture naming but not in forward translation, indicating that the mechanisms underlying L1 lexical processing may bear somewhat similarities to what occurs in English picture naming. Lastly, L2 reading and listening competence were found to predict translation speed but not naming latencies. This suggests that there may be a relationship between translation competence and the proficiency levels of L2 receptive skills.

Table 4. 8 Standard MR: Predictive effects of learner-related variables on PE response latencies among child L2 learners (n = 39)

Variables	PE	Vocabulary size	Chinese course scores	*B*	*β*	sr^2 (unique)
Vocabulary size	-.40*			-1.83*	-.32	.10
Chinese course scores	-.41*	.24		-33.32*	-.33	.11
			Intercept= 5271.74			
Means	1803.73	202.08	93.00			
Standard deviations	276.35	48.66	2.73			
					R^2= .26	
					Adjusted R^2= .22	
					R= .51**	

Notes: * $p < .05$, and ** $p < .01$;
B: unstandardized regression coefficients;
β: standardized regression coefficients;
$sr^{2:}$ semi-partial correlations.

Table 4. 9 Standard MR: Predictive effects of learner-related variables on FT response latencies among child L2 learners (n = 39)

Variables	FT	Vocabulary size	Reading proficiency level	Listening proficiency level	*B*	*β*
Vocabulary size	-.48**				-1.80*	-.38
Reading proficiency level	-.35*	.50**			-11.18	-.13
Listening proficiency level	-.36*	.60**	.62**		-3.89	-.06
				Intercept =	2322.50	
Means	1752.07	202.08	13.31	15.41		
Standard deviations	227.94	48.66	2.56	3.42		
						R^2= .25
						Adjusted R^2= .19
						R= .50*

Notes: * $p < .05$, and ** $p < .01$;
B: unstandardized regression coefficients;
β: standardized regression coefficients;
$sr^{2:}$ semi-partial correlations.

4.3.2 Predictive effects of word feature-related variables on L2 lexical processing among child L2 learners

This analysis addresses the following research question:

Do word feature-related variables predict the speed of L2 lexical processing in picture naming and forward translation among child L2 learners?

Two path models were estimated in such a way that was similar to that described in Section 4.2.3.

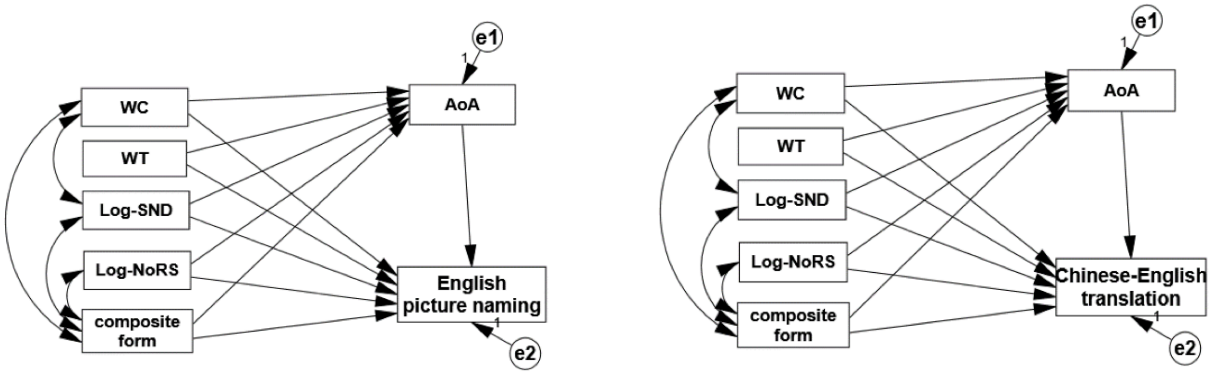
4.3.2.1 Model specification

The hypothesized models (see Figure 4.6 below) consisted of six observed lexical variables, namely: word concreteness, word typicality, semantic neighborhood density, the number of related senses, the composite form, and age of acquisition. These variables were hypothesized to predict PE and FT response latencies. Substantively, the empirical grounds for such hypotheses came from a series of studies examining the impacts of lexical variables on the processing of L1 and L2 words among monolingual and bilingual children (e.g. Bjorklund & Thompson, 1983; D'Amico et al., 2001; Greenberg & Bjorklund, 1981; McFalls et al., 1996; Schwanenflugel & Akin, 1994). This model was over-identified with 6 dfs. It should be noted that the variables that might influence PE and FT response latencies such as the visual stimuli complexity of pictures were not included in the model because the predictive effects of these variables on the outcome variable were not the focus of the present study. A correlation matrix of lexical variables and the child participants' response latencies is presented in Table 4.10.

Table 4. 10 Correlation matrix of lexical variables and child participants' response latencies (n = 104)

	1	2	3	4	5	6	7	8	9	10	11
1. PE	1										
2. FT	.70**	1									
3. WC	.09	-.03	1								
4. WT	-.46**	-.37**	-0.163	1							
5. Log-SND	-.19	-.10	-.44**	0.16	1						
6. Log-NoRS	.11	-.02	-.01	-0.10	.06	1					
7. Log-NoL	.15	.20*	.22*	-0.08	-.23*	-.45**	1				
8. Log-NoP	.11	.10	.26**	0.02	-.30**	-.33*	.77**	1			
9. Log-NoS	-.00	-.03	.30**	0.13	-.17	-.41**	.67**	.76**	1		
10. SWF	-.20*	-.24*	.18	.28**	.03	.28**	-.25*	-.14	.00	1	
11. AoA	.49**	.27**	.15	-0.19	-.26**	.30**	.06	.03	.00	.13	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).



Note: R: recoded; and C: composite.

Figure 4. 6 Hypothesized path models: Predictive effects of word feature-related variables on PE and FT response latencies in child L2 learners

4.3.2.2 Evaluation of assumptions for the specified path models

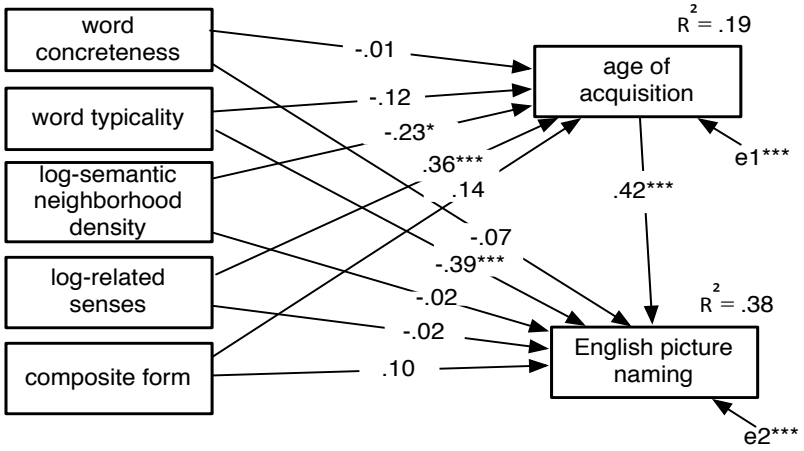
All the assumptions required for path analysis were checked (see Appendix 4.8 for statistical details). It is considered acceptable to perform the path analysis with the obtained size of 104 lexical items although the desirable size was not reached, according to Bentler and Chou's recommendation (1987). There were no univariate and multivariate outliers observed. The skewness and kurtosis values of the variables all fell between +2.00 and -2.00, and no inappropriate level of multicollinearity and singularity was noted. Thus, ML estimates were obtained using IBM SPSS AMOS v.20.

4.3.2.3 Path model estimation and results

PREDICTIVE EFFECTS OF WORD FEATURE-RELATED VARIABLES ON PE RESPONSE LATENCIES

The independence model that tests the hypothesis when all variables are uncorrelated was easily rejected, $\chi^2(21, N = 104) = 135.82, p < .001$, CFI = .000, RMSEA = .230.

The hypothesized model was estimated with ML estimation, and the result suggested that the model fitted the data perfectly, $\chi^2(6, N = 104) = 5.665$, $p = .462$, CFI = 1.000 RMSEA = .000. The model with standardized coefficients is diagrammed in Figure 4.7. The output of IBM SPSS AMOS is presented in Appendix 4.9.



Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$;
R: recoded.

Figure 4. 7 Final path model with standardized coefficients: Predictive effects of word features on PE response latencies among child L2 learners

Direct effects. As shown in Figure 4.7, PE response latencies decreased as word typicality increased (standardized coefficient = -.39, $p < .001$) and as age of acquisition decreased (standardized coefficient = .42, $p < .001$). The composite variable form did not significantly predict the response latency of PE (standardized coefficient = .10, $p = .32$), as well as semantic neighborhood density (standardized coefficient = -.02, $p = .81$), the number of related senses (standardized coefficient = -.02, $p = .93$), and word concreteness (standardized coefficient = -.07, $p = .44$).

Indirect effects. As MacKinnon, Fritz, Williams, and Lockwood (2007)

recommended, assessing the indirect effect using bootstrapping is the most powerful method yielding the lowest Type I error rates. Given this, the bias-corrected bootstrap 95% confidence intervals were employed. The results suggested that the indirect effect from semantic neighborhood density to PE response latencies via usage was significant (standardized coefficient = -.10, $p < .05$), and the indirect effect form the number of related senses to PE response latencies via usage was significant (standardized coefficient = .15, $p < .001$). The indirect effects from word concreteness, word typicality, and the composite variable from to PE response latencies via usage were not significant (standardized coefficient = -.00, $p = 89$, standardized coefficient = -.05, $p = 15$, and standardized coefficient = .06, $p = 18$ respectively)

Total effects. Total effects are defined as the sum of all direct and indirect effects of one variable on another (Kline, 2011). the total effect of word typicality on the RL of L2 picture naming (-.44) was calculated as follows: direct effect (-.39) + indirect effect (-.05 = -.12 × .42).

The results showed that the standardized total effects of word typicality and age of acquisition on PE response latencies was statistically significant (standardized coefficient = -.44, $p < .001$ and standardized coefficient = .42, $p < .001$ respectively). The total effects of word concreteness, semantic neighborhood density, the number of related senses and form on PE response latencies were not statistically significant (standardized coefficients= -.08, -.12, .14 and .15 respectively, p = .45, .24, .17 and .12 respectively). A summary of the decomposition of standardized effects is

presented in Table 4.11.

Table 4. 11 Decomposition of standardized effects for path diagrams of Figures 4.7

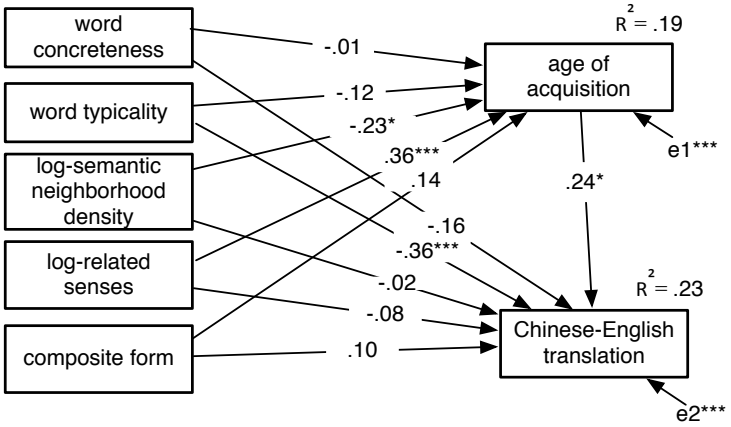
	Direct effects	Indirect effects	Total effects
WC	-.07	-.00	-.08
WT	-.39***	-.05	-.44***
Log-SND	-.02	-.10*	-.12
Log-NoRS	-.02	.15***	.14
Form	.10	.06	.15
AoA	.42***	–	.42***

Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$.

Per cent of variances accounted for. The R^2 value was .38 (a large effect size), suggesting that 38% of the variability in PE response latencies was accounted for by a set of lexical variables included.

PREDICTIVE EFFECTS OF WORD FEATURE-RELATED VARIABLES ON FT RESPONSE LATENCIES

The independence model that tests the hypothesis that all variables are uncorrelated was easily rejected, $\chi^2(21, N = 104) = 109.83$, $p < .001$, CFI = .000, RMSEA = .203. The hypothesized model was then estimated with ML estimation, and the result suggested that the model fitted the data perfectly, $\chi^2(6, N = 104) = 5.665$, $p < .001$, CFI = 1.000, RMSEA = .000. The model with standardized coefficients is diagrammed in Figure 4.8. The output of IBM SPSS AMOS is presented in Appendix 4.9.



Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$;
R: recoded.

Figure 4. 8 Final path model with standardized coefficients: Predictive effects of word features on FT response latencies among child L2 learners

Direct effects. As shown in Figure 4.8, FT response latencies decreased as age of acquisition decreased (standardized coefficient = .24, $p < .05$) and WT increased (standardized coefficient = -.36, $p < .001$). The lexical variable word concreteness did not significantly predict the response latency of FT (standardized coefficient = -.16, $p = .12$), as well as semantic neighborhood density (standardized coefficient = -.02, $p = .87$), the number of related senses (standardized coefficient = -.08, $p = .51$), and form (standardized coefficient = .10, $p = .35$).

Indirect effects. The indirect effect from semantic neighborhood density to FT response latencies via usage was significant (standardized coefficient = -.06, $p < .05$), and the indirect effect form the number of related senses to FT response latencies via usage was significant (standardized coefficient = .09, $p < .05$). The indirect effect from word concreteness, word typicality, and form to FT response latencies via usage

was not significant (standardized coefficient = -.00, -.03, and .03 respectively, *p* = .87, .12, and .14 respectively).

Total effects. The standardized total effect of word typicality and age of acquisition on FT response latencies was statistically significant (standardized coefficient = -.40, and .24, both *ps* < .05). The total effects of word concreteness, semantic neighborhood density, the number of related senses, and form on FT response latencies were not statistically significant (standardized coefficients= -.17, -.07, .005 and .14 respectively, *p* = .12, .47, .95 and .19 respectively). A summary of the decomposition of standardized effects is presented in Table 4.12.

Table 4. 12 Decomposition of standardized effects for path diagrams of Figures 4.8

	Direct effects	**Indirect effects**	**Total effects**
WC	-.16	-.00	-.17
WT	-.36***	-.03	-.40***
Log-SND	-.02	-.06*	-.07
Log-NoRS	-.08	.09*	.005
Form	.10	.03	.14
AoA	.24*	–	.24*

Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$.

Percent of variances accounted for. The R^2 value was .23 (an effect size between medium and large) suggested that 23% of the variability in FT response latencies was accounted for by a set of the lexical variables included.

4.3.2.4 Interim summary: Predictive effects of lexical variables on processing efficiency

A cross-task comparison demonstrated that lexical variables were more involved in

picture naming than in forward translation. At the conceptual-semantic level, semantic neighborhood density and the number of related senses significantly predicted the RLs via word usage in both naming and translation tasks. The indirect effect of NoRS was found to be larger than that of SND, possibly because children's semantic networks have not been fully developed but instead they are more able to grasp related senses of individual words. Word typicality served as the strongest predictor of naming and translation speed over and above other variables, with a total effect of .44 and .40 in naming and translation respectively. Contrary to our expectation, word concreteness did not have significant effects on the outcomes. This may be because most of the words that the participants had learned at this stage were concrete or nearly concrete. At the form level, none of the predictors were found to be significant, a result that is compatible with some previous studies (e.g. Bonin et al., 2002; Hirsh et al., 2003). At the usage level, the mediating role that the age of acquisition played between SND/NoRS and L2 lexical processing confirms the existence of a semantic locus of AoA effects (Van Loon-Vervoorn et al., 1988).

4.4 General pattern of L2 Lexical processing among child L2 learners

This analysis answers the following research questions:

What is the general pattern of L2 lexical processing among child L2 learners?

Before examining the general pattern of L2 lexical processing among child L2 learners, the results of the preparation study are reported first. The main aim of that study was to compare children's encoding speed of pictures with that of Chinese

characters (see Section 3.6.2 above for details on the experiment).

4.4.1 Speed and accuracy of encoding pictures and Chinese characters

The speed and accuracy of encoding pictures and Chinese characters were compared across conditions among the child participants (n = 28) with the use of F1 (by participants) and of F2 (by items) repeated-measures ANOVAs. The statistical assumptions were checked. The skewness and kurtosis values of the variables all fell between +2.00 and -2.00, without any serious violations of the normality noted. There were no outliers identified far beyond +/- 3.00 SD. The treatment of missing values here were the same as the way in which RL data were treated, as reported in Section 4.1.

4.4.1.1 Speed of encoding

BY PARTICIPANTS

The descriptive statistics of the speed of encoding by the child participants across two stimulus conditions are presented in Table 4.13.

Table 4. 13 Descriptive statistics of encoding speed (in ms) by child participants between conditions

Conditions	n	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Pictures	28	813.82	163.19	571.52	1147.76	.91	.59
Characters	28	802.96	212.41	526.17	1282.11	1.21	1.20

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-subjects factor. The ANOVA result suggested that there was no significant effect of stimulus conditions on the speed of encoding by participants, F1 (1, 27) = .30, p = .59, with a small effect size (η_p^2 = .01), indicating that the speed of encoding pictures was not significantly different from that of encoding Chinese characters.

BY ITEMS

The descriptive statistics of the speed of encoding by participants between two stimulus conditions are presented in Table 4.14.

Table 4. 14 Descriptive statistics of encoding speed (in ms) by items between conditions: Child participants

Conditions	n	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Pictures	60	833.52	104.62	661.64	1112.00	.73	.08
Characters	60	855.13	119.55	645.43	1159.56	.33	-.22

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-items factor. The ANOVA result suggested that there was no significant effect of stimulus conditions on the speed of encoding by items, F2 (1, 59) = 1.10, p = .30, with an effect size between small and medium (η_p^2 = .02), indicating that the speed of encoding pictures was not significantly different from that of encoding Chinese characters by items.

4.4.1.2 Accuracy of encoding

BY PARTICIPANTS

The descriptive statistics of the accuracy of encoding by participants between two stimulus conditions are summarized in Table 4.15.

Table 4. 15 Descriptive statistics of accuracy of encoding (in percentage) by child participants between conditions

Conditions	n	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Pictures	28	98.07	2.58	93.33	100	-1.02	-.49
Characters	28	97.99	2.43	91.67	100	-.87	-.50

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-subjects factor. The ANOVA result showed that there was no significant effect of stimulus conditions on the accuracy of encoding by participants, F1 (1, 27) = .02, p = .88, with a small effect size (η_p^2 = .001), suggesting that the accuracy of encoding did not vary significantly with the stimulus conduction.

BY ITEMS

The descriptive statistics of the accuracy of encoding by items across two stimulus conditions are presented in Table 4.16.

Table 4. 16 Descriptive statistics of accuracy of encoding (in percentage) by items between conditions: Child participants

Conditions	n	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Pictures	60	98.57	2.39	92.86	100	-1.43	1.26
Characters	60	98.33	2.75	89.28	100	-1.50	.76

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-items factor. The ANOVA result showed that there was no significant effect of stimulus conditions on the accuracy of encoding by items, F2 (1, 59) = .30, p = .58, with a small effect size (η_p^2 = .005), suggesting that encoding picture was not significantly different from encoding Chinese characters by items in terms of the accuracy.

To sum up, the preparation study indicates that encoding Chinese characters was comparable to encoding pictures among the child participants in terms of the speed and accuracy, the result of which would support the comparability of the beginning stage of PN and FT so as to corroborate the internal validity of this research and validate the results to be obtained.

4.4.2 General pattern of L2 lexical processing among child L2 learners

Response latencies and accuracies were compared across conditions among the child participants (n = 39) with the use of F1 (by participants) and of F2 (by items) mixed-model ANOVAs. The statistical assumptions were checked, indicating that almost all the skewness and kurtosis values of the variables fell between +2.00 and -2.00 although a moderate departure from normality was noted when response latencies were compared across conditions by participants (see Table 4.17 below). The results of the *Shapiro–Wilk* tests, however, indicated that the assumption of normality was met as all the p values were greater than .05. Moreover, there were no outliers beyond the recommended limits (+/-3.00 SD). The assumption of

homogeneity was tested with the Levene's test, and violations of the homogeneity of variance were noted in comparing the RAs across blocks by items. However, according to Howell (2011), ANOVA procedure was considered robust to violations of the homogeneity assumption, provided that group sizes are approximately equal. The assumption of sphericity was checked with the Mauchly's test, and Greenhouse-Geisser or Huynh-Feldt corrections for the violation of sphericity of covariance were made wherever appropriate. Additionally, it should be pointed out that by-item analyses were reported in a similar way to by-participant analyses in order to assess statistical reliability across these two types of analysis but also to check whether or not stimulus blocks were comparable based on the participants' performance.

4.4.2.1 Response latencies

BY PARTICIPANTS

The descriptive statistics of RLs by participants across three stimulus conditions in each group are presented in Table 4.17.

Table 4. 17 Descriptive statistics of response latencies (in ms) by participants across stimulus conditions in each group: Child L2 learners

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Group 1	PE	13	1773.29	1173.12	2249.97	275.38	-.59	.92
	FT	13	1717.61	1195.71	2191.00	287.23	-.30	-.61
	PC	13	1070.30	874.77	1314.31	157.95	.11	-1.75
Group 2	PE	13	1825.05	1311.14	2382.00	334.82	-.10	-.77
	FT	13	1753.70	1253.15	2285.00	238.76	.16	2.29
	PC	13	972.07	819.74	1180.54	129.09	.55	-1.12
Group 3	PE	13	1812.86	1355.13	2302.18	228.69	-.16	2.00
	FT	13	1784.89	1397.17	2032.09	150.20	-1.10	3.48
	PC	13	995.20	695.29	1263.39	173.63	.11	-.50

Notes: PE: picture naming in English; FT: forward translation; and PC: picture naming in Chinese.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-subjects factor (tested at 3 levels: PE, FT, and PC), and with the grouping as a between-subjects factor (tested at 3 levels: Group 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RLs by participants, F1 (2, 72) = 277.28, $p < .001$, with a large effect size ($\eta_p^2 = .89$), and there was no significant effect of grouping on the RLs by participants F1 (2, 36) = .02, $p = .98$, with a small effect size ($\eta_p^2 = .001$). The interaction effect between stimulus conditions and the group was not significant, F1 (4, 72) = 1.02, $p = .40$, with an effect size between small and medium ($\eta_p^2 = .05$). Post-hoc comparisons using the Bonferroni-correction were conducted. Note that Bonferroni correction serves to control for family-wise error rate (Field, 2013). The results suggested that PE response latencies were not significantly different from FT response latencies by participants (all *ps* > .05) in all three groups, and PC was significantly faster than FT and than PE by participants (all *ps* < .001) in

all there groups (see Table 4.18 for details). These effects are illustrated in Figure 4.9.

Table 4. 18 Post-hoc pairwise comparisons between stimulus conditions in each group (n = 39): Response latencies (in ms) among child L2 learners

		MD	SE	Sig.
Group 1	PE vs. FT	55.68	64.38	1.000
	PE vs. PC	702.99***	67.21	.000
	FT vs. PC	647.31***	63.70	.000
Group 2	PE vs. FT	71.36	64.38	.825
	PE vs. PC	852.99***	67.21	.000
	FT vs. PC	781.63***	63.70	.000
Group 3	PE vs. FT	27.97	64.38	.000
	PE vs. PC	817.65***	67.21	.000
	FT vs. PC	789.69***	63.70	.000

Notes: *** $p < .001$;
MD: meaning difference; and SE: standard error.

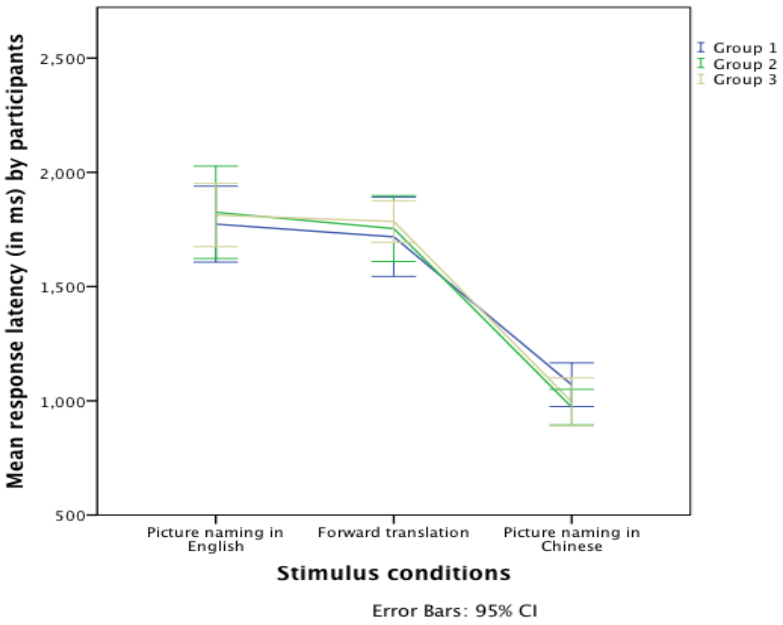


Figure 4. 9 Comparisons (by participants) across stimulus conditions and groups: response latencies among child L2 learners

BY ITEMS

The descriptive statistics of RLs by items across stimulus conditions in each block are presented in Table 4.19.

Table 4. 19 Descriptive statistics of response latencies (in ms) by items across stimulus conditions in each block: Child L2 learners

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Block 1	PE	35	1801.07	1089.92	2457.27	371.03	-0.06	-0.84
	FT	35	1717.79	1039.83	2520.36	386.29	0.28	-0.85
	PC	35	972.82	665.67	1545.36	183.95	0.81	1.37
Block 2	PE	35	1774.45	962.58	2868.91	516.41	0.46	-0.53
	FT	35	1716.84	1007.75	2768.27	400.5	0.28	0.19
	PC	35	955.95	697.08	1603.33	214.41	1.54	1.83
Block 3	PE	35	1810.79	1081.08	2515.33	366.96	-0.35	-0.56
	FT	35	1740.95	995.07	2505.07	431.89	-0.12	-1.05
	PC	35	1087.75	713.58	1659.00	257.57	0.68	-0.41

Notes: PE: picture naming in English; FT: forward translation; and PC: picture naming in Chinese.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-items factor (tested at 3 levels: PE, FT, and PC), and with the block of stimulus words as a between-items factor (tested at 3 levels: Block 1, Block2, and Block 3). The ANOVA result suggested a significant effect of stimulus conditions on RLs by items, F2 (1.87, 190.73) = 259.84, $p < .001$, with a large effect size ($\eta_p^2 = .72$). There was no significant effect of stimulus blocks on RLs by items, F2 (2, 102) = .49, $p = .61$, with a small effect size ($\eta_p^2 = .01$), and no significant interactions between the condition and the stimulus block, F2 (3.74, 370.10) = .52, $p = .71$, with a small effect size ($\eta_p^2 = .01$). Bonferroni-corrected post-hoc comparisons were conducted, and the results suggested that PE response latencies were not significantly different

from FT response latencies by items in all three blocks (all *ps* > .05), and PC was significantly faster than FT and than PE by items in all three blocks (all *ps* < .001) (see Table 4.20 for details). These effects are illustrated in Figure 4.10.

Table 4. 20 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 105): Response latencies (in ms) by items among child L2 learners

		MD	SE	Sig.
Block 1	PE vs. FT	83.28	54.47	.388
	PE vs. PC	828.25***	70.60	.000
	FT vs. PC	744.97***	72.73	.000
Block 2	PE vs. FT	57.61	54.47	.878
	PE vs. PC	818.50***	70.60	.000
	FT vs. PC	760.89***	72.73	.000
Block 3	PE vs. FT	69.84	54.47	.608
	PE vs. PC	723.04***	70.60	.000
	FT vs. PC	653.20***	72.73	.000

Notes: *** $p < .001$;
MD: meaning difference; and SE: standard error.

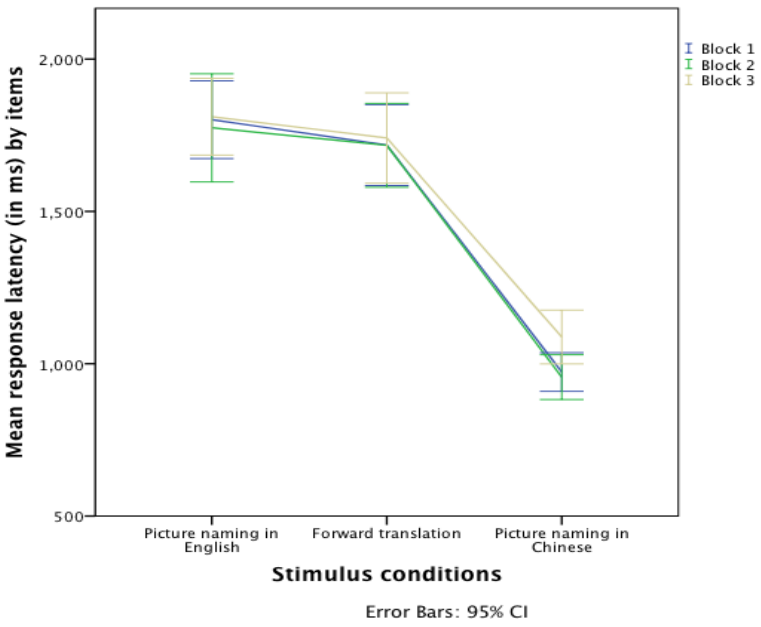


Figure 4. 10 Comparisons across stimulus conditions and blocks: Response latencies by items among child L2 learners

4.4.2.2 Response accuracies

BY PARTICIPANTS

The descriptive statistics of RAs across stimulus conditions in each group are presented in Table 4.21.

Table 4. 21 Descriptive statistics of response accuracies (in percentage) by participants across stimulus conditions in each group: Child L2 learners

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Group 1	PE	13	88.79	85.71	91.43	1.83	-.05	.06
	FT	13	86.81	82.86	91.43	2.98	.10	-.98
	PC	13	96.26	91.43	100.00	2.15	-.78	1.22
Group 2	PE	13	88.13	80.00	91.43	3.67	-1.07	.54
	FT	13	89.01	82.86	94.29	3.27	-.74	.65
	PC	13	96.26	91.43	100.00	2.44	-.24	-.04
Group 3	PE	13	89.23	85.71	94.29	3.72	-.30	-1.80
	FT	13	89.67	85.71	94.29	2.98	.10	-.1.00
	PC	13	97.36	94.29	100.00	-.05	.06	1.74

Notes: PE: picture naming in English; FT: forward translation; and PC: picture naming in Chinese.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-subjects factor (tested at 3 levels: PE, FT, and PC), and with the grouping as a between-subjects factor (tested at 3 levels: Group 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RAs by participants, F1 (2, 72) = 104.25, p < .001, with a large effect size partial (η_p^2 = .74). There was no significant effect of the variable grouping on RAs by participants, F1 (2, 36) = 1.96, p = .15, with a medium effect size (η_p^2 = .10), and no significant interactions observed between the stimulus condition and the group, F1 (4, 72) = 1.04, p = .39, with an effect size between small and medium (η_p^2 = .06). Bonferroni-corrected post-hoc

comparisons were conducted, which suggested that the participants made more errors in PE and FT as compared to PC in all three groups (all *ps* < .001). However, there were no significant differences found between the accurate rate of PE and that of FT by participants in all three groups (all *ps* > .05) (see Table 4.22 for details). These effects are illustrated in Figure 4. 11.

Table 4. 22 Pairwise comparisons between stimulus conditions in each group (n = 39): Response accuracies (in percentage) among child L2 learners

		MD	SE	Sig.
Group 1	PE vs. FT	1.98	1.21	.328
	PE vs. PC	-7.47	1.03	.000
	FT vs. PC	-9.45***	1.10	.000
Group 2	PE vs. FT	-.88	1.21	1.000
	PE vs. PC	-8.13***	1.03	.000
	FT vs. PC	-7.25***	1.10	.000
Group 3	PE vs. FT	.44	1.21	1.000
	PE vs. PC	-8.13***	1.03	.000
	FT vs. PC	-7.69***	1.10	.000

Notes: *** $p < .001$;
MD: meaning difference; SE: standard error.

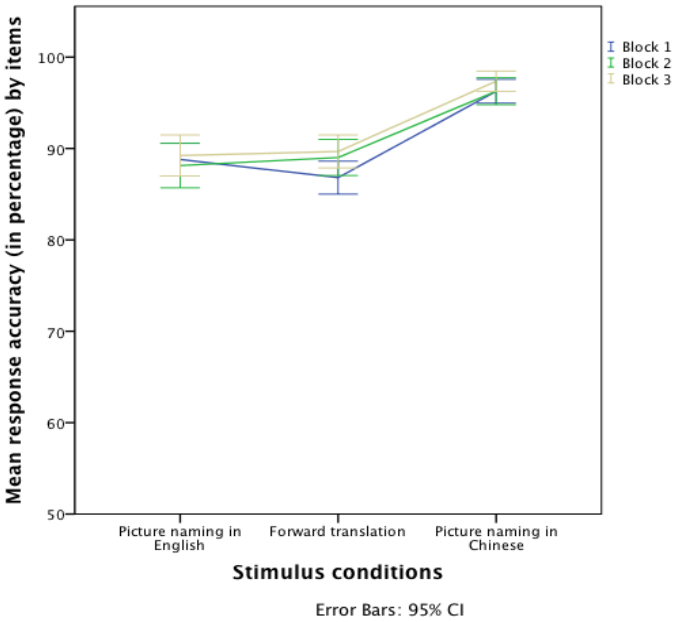


Figure 4. 11 Comparisons (by participants) across stimulus conditions and groups:

Response accuracies (in percentage) among child L2 learners

BY ITEMS

The descriptive statistics of RAs across stimulus conditions in each block are presented in Table 4.23.

Table 4. 23 Descriptive statistics of response accuracies (in percentage) by items across stimulus conditions in each block: Child L2 learners

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Block 1	PE	35	89.23	69.23	100.00	7.03	-.80	.89
	FT	35	88.57	69.23	100.00	8.82	-.41	-.91
	PC	35	97.14	92.31	100.00	3.78	-.56	-1.80
Block 2	PE	35	89.67	69.23	100.00	6.97	-.75	1.06
	FT	35	87.91	76.92	100.00	7.29	.11	-.81
	PC	35	96.70	84.62	100.00	4.29	-.84	-.31
Block 3	PE	35	89.01	76.92	100.00	7.05	-.47	-.57
	FT	35	90.11	76.92	100.00	6.61	-.28	-.40
	PC	35	96.48	92.31	100.00	3.89	-.18	-.20

Notes: PE: picture naming in English; FT: forward translation; and PC: picture naming in Chinese.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-items factor (tested at 3 levels: PE, FT, and PC), and with the block as a between-items factor (tested at 3 levels: Block 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RAs by items, F2 (1.92, 195.18) = 86.34, $p < .001$, with a large effect size ($\eta_p^2 = .45$). There was no significant effect of the block on RAs by items, F2 (2, 102) = .14, $p = .87$, with a very small effect size ($\eta_p^2 = .001$), and no significant interactions between the stimulus condition and the stimulus block, F2 (3.83, 195.18) = .98, $p = .42$, with a small effect size ($\eta_p^2 = .02$). Bonferroni-corrected post-hoc comparisons suggested that more

errors were made in PE and FT as compared to PC by items in all three blocks (all *ps* < .001). However, there were no significant differences found between the accurate rate of PE and that of FT by items in all three blocks (all *ps* > .05) (see Table 4.24 for details). These effects are illustrated in Figure 4.12.

Table 4. 24 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 105): Response accuracies by items among child L2 learners

		MD	SE	Sig.
Block 1	PE vs. FT	.66	1.01	1.000
	PE vs. PC	-7.91***	1.19	.000
	FT vs. PC	-8.57***	1.31	.000
Block 2	PE vs. FT	1.76	1.01	.252
	PE vs. PC	-7.03***	1.19	.000
	FT vs. PC	-8.79***	1.31	.000
Block 3	PE vs. FT	-1.10	1.01	.833
	PE vs. PC	-7.47***	1.19	.000
	FT vs. PC	-6.37***	1.31	.000

Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$;
MD: meaning difference; SE: standard error.

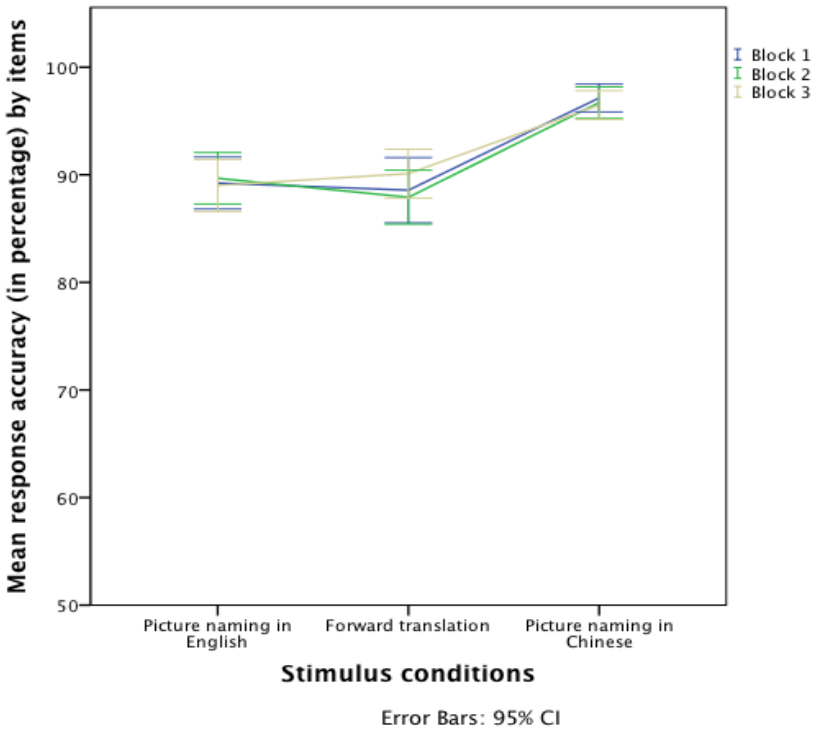


Figure 4. 12 Comparisons across stimulus conditions and blocks: Response accuracies by items among child L2 learners

4.4.2.3 Interim summary: General pattern of children's L2 lexical processing

The absence of significant differences between PE and FT response latencies was observed with the child participants, indicating that young children's L2 lexical processing is likely to be characterized by conceptual mediation, this is to say, there exists a direct link between L2 words and their concepts. This result is consistent with the study by Chen and Leung (1989), in which target stimulus words were found to be conceptually processed among the 2nd grade children who had studied English for 2 years. However, this result seems to run against the *Developmental Hypothesis* (Kroll & De Groot, 1997), which postulates that the ability to conceptually process L2 words has not been ready for less fluent learners. Possible explanations for such an inconsistency may relate to vocabulary learning and teaching strategies used in

input-limited contexts and to the less firmly established L1 knowledge in children.

4.5 Tendency toward a certain pattern of L2 lexical processing among child L2 learners

This analysis addresses the following research questions:

Do learner-related variables predict the tendency toward a certain pattern of L2 lexical processing among child L2 learners?

Do word feature-related variables predict the tendency toward a certain pattern of L2 lexical processing among child L2 learners?

As discussed earlier in Section 3.6.3, when the response latency of PE is significantly longer than that of FT, learners are more likely to process L2 words through lexical routes or equivalent translations. On the contrary, when the PE response latency is not significantly different from that of FT, L2 words are more likely to be conceptually processed by learners. Following this logic, it was hypothesized that the larger PE response latencies minus FT response latencies, the more likely it is that the word association model is supported. In order to identify which variables might predict the tendency toward a certain pattern of L2 lexical processing, standard MR was conducted, with learner-related and word feature-related variables as the predictors, and PE-FT response latency differences as the outcomes. The analysis steps used here were similar to those described in Section 4.3.1.

4.5.1 Learner-related variables and processing tendency

Before performing formal analyses, descriptive statistics for all predictor variables and the outcome variable were examined, and a correlation matrix among continuous predictors, and between each predictor variable and the outcome variable was estimated (see Table 4.25). The strength of association between nominal predictors and the outcome variable was estimated using Lambda, and the results are reported in Appendix 4.10. The correlation results suggested that only one predictor variable (i.e. Chinese course scores) was significantly correlated with PE-FT response latency differences. This variable was included in the regression model to examine the extent to which it predicted the PE-FT response latency differences. Moreover, the assumptions of running MR were checked, including the sample size, missing data, normality, linearity, and homoscedasticity of residuals, outliers, and multicollinearity and singularity (see Appendix 4.11 for details), with no serious violation noted.

Table 4. 25 Correlation matrices: Relationships between learner-related variables, and PE-FT response latency differences among child L2 learners (n = 39)

Pearson's correlations

Variables	1	2	3	4	5	6	7	8	9	10
1. PE-FT differences	1	—	—	—	—	—	—	—	—	—
2. Age	-.09	1	—	—	—	—	—	—	—	—
3. WASI (MR)	.11	.02	1	—	—	—	—	—	—	—
4. Vocabulary size	-.002	.08	.23	1	—	—	—	—	—	—
5. Vocabulary depth	.05	.31	.24	.32	1	—	—	—	—	—
6. English course scores	-.02	.49**	-.03	.41*	.39*	1	—	—	—	—
7. Chinese course scores	-.34*	.43**	.17	.24	.42**	.49**	1	—	—	—
8. Math course scores	.20	.31	.27	.19	.26	.49**	.27	1	—	—
9. Reading proficiency level	-.01	.23	.21	.50**	.17	.33*	.28	.20	1	—
10. Listening proficiency level	.01	.25	.38*	.60**	.41**	.60**	.43**	.38*	.62**	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

The MR result suggested that Chinese course scores significantly predicted the PE-FT response latency differences (see Table 4.26), the R (.34) was significantly different from zero, F (1, 37) = 4.79, $p < .05$, with R^2 at .12 (an effect size between medium and large). The adjusted R^2 value of .09 indicated that only 9% of the variability in the PE-FT response latency difference was predicted by Chinese course scores. The direction of the relationship suggested that those participants who had lower Chinese grades were more likely to process L2 words through word associations.

Table 4. 26 Standard multiple regression of Chinese course scores on the PE-FT response latency difference (n = 39)

Variables	PE-FT differences	Chinese course scores	*B*	*β*
Chinese course scores	- .34*		-28.06*	-.34
			Intercept = 2660.20	
Means	51.65	93.00		
Standard deviations	226.51	2.73		
				R^2= .12
				Adjusted R^2= .09
				R= .34*

Notes: * $p < .05$;
B: unstandardized regression coefficients;
β: the standardized regression coefficients;
$sr^{2:}$ the semi-partial correlations.

4.5.2 Word feature-related variables and processing tendency

As in the above analysis, descriptive statistics for all the predictor variables and the outcome variable were examined, and a correlation matrix among continuous predictors, and between each predictor and the outcome variable was estimated (see Table 4.27). It was found that only the variable of age of acquisition was significantly correlated with PE-FT response latency differences. This variable was therefore included in the regression model to examine the extent to which it predicted the PE-FT response latency differences. Moreover, the assumptions of MR were checked (see Appendix 4.11 for details), with no serious violation noted.

Table 4. 27 Correlation matrix: Correlations between word feature-related variables, and PE-FT response latency differences among child L2 learners (n = 105)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. PE-FT differences	1	—	—	—	—	—	—	—	—	—
2. WC	.17	1	—	—	—	—	—	—	—	—
3. WT	-.14	-.17	1	—	—	—	—	—	—	—
4. Log-SND	-.13	-.44**	.15	1	—	—	—	—	—	—
5. Log-NoRS	.15	-.02	-.09	.06	1	—	—	—	—	—
6. Log-NoL	-.04	.23*	-.09	-.22*	-.46**	1	—	—	—	—
7. Log-NoP	.04	.27**	-.02	-.28**	-.34**	.76**	1	—	—	—
8. Log-NoS	.04	.31**	.12	-.17	-.41**	.67**	.72**	1	—	—
9. SWF	.03	.16	.29**	.03	.30**	-.28**	-.21*	-.01	1	—
10. AoA	.31**	.16	-.20*	-.26**	.29**	.08	.07	.01	.11	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

In the regression analysis of the variable age of acquisition that might predict PE-FT response latency differences (see Table 4.28), the R (.31) was significantly different from zero, F (1, 103) = 10.92, $p < .01$, with R^2 at .10 (an effect size between medium and large). The adjusted R^2 value of .09 indicated that 9% of the variability in PE-FT response latency differences was significantly predicted by the age of acquisition. The direction of the relationship suggested that those words that were learned at a later age were more likely to be processed through lexical associations.

Table 4. 28 Standard multiple regression of age of acquisition on the PE-FT response latency difference among child L2 learners (n = 105)

Variables	PE-FT differences	Age of acquisition	*B*	*β*
Age of acquisition	.31**		117.27**	.31
			Intercept = 1087.89	
Means	70.24	9.88		
Standard deviations	319.31	.84		
				R^2= .10
				Adjusted R^2= .09
				R= .31**

Notes: ** $p < .01$;
B : unstandardized regression coefficients;
β: the standardized regression coefficients;
$sr^{2:}$ the semi-partial correlations.

4.5.3 Interim summary: Factors predicting processing tendency

Multiple regression analysis on the predictive effects of Chinese course grades and age of acquisition on response latency differences between English picture naming and forward translation showed that those participants that had lower L1 competence and those words that were learned later tended to support the word association model.

The results suggest that L1 and L2 abilities are underpinned by a common system (Cummins, 1991), and that L1 competence may play a crucial role in advancing L2 words to be conceptually processed. Additionally, late-acquired words are less likely to be conceptually processed than early-acquired words, probably because young children have more chances to practice early-acquired ones or they have learned these words within the critical period.

4.6 Summary: Chapter 4

In brining this chapter to a close, it is instructive to return to Figure 4.1 presented at the beginning of this chapter, and how far those research questions have been answered are summarized in Figure 4.13 below. The main results obtained in the present study indicate that several learner and lexical predictors accounted for a considerable amount of variance in L2 picture naming and forward translation, but in comparison, lexical variables mattered more in predicting the processing efficiency relative to learner variables. Moreover, the absence of significant differences in response latencies between picture naming and forward translation, both by participants and by items, gave support to the concept mediation model. However, it should be noted that the existence of this general processing pattern does not deny the fact that some words may still be lexically accessed among young children. Lastly, L1 competence and a lexical item's age of acquisition best predicted the tendency toward a certain pattern of lexical processing over and above other learner and lexical variables. Theoretical and practical implications of these findings will be discussed in

Chapter 6 and 7.

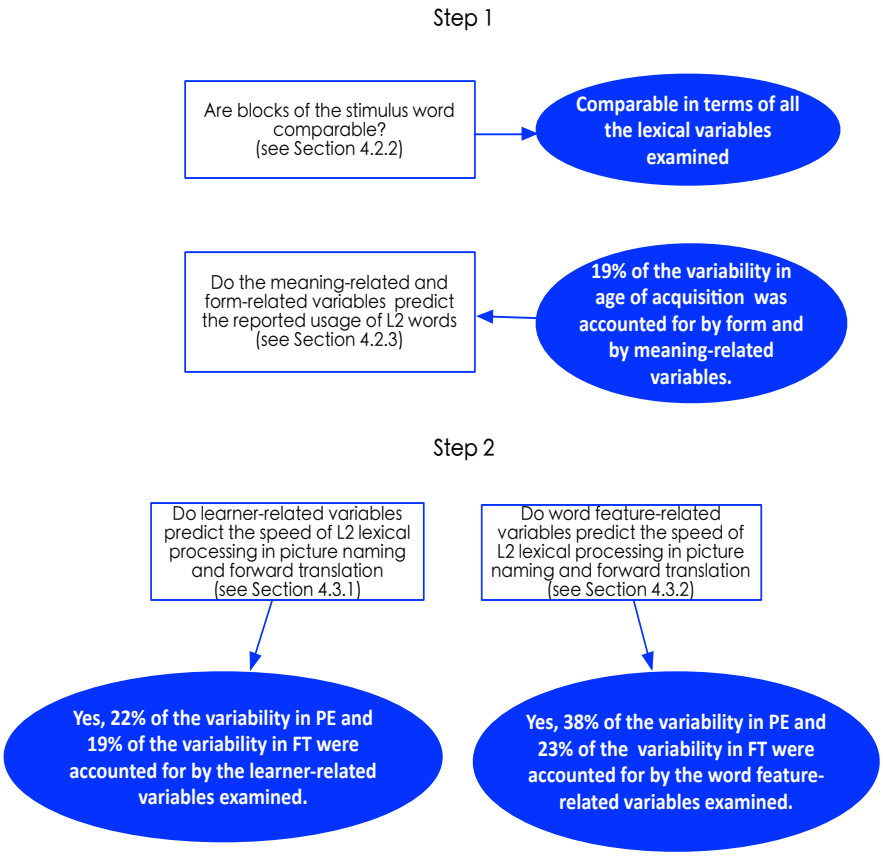


Figure 4. 13 Summary of Chapter 4: Research questions and results

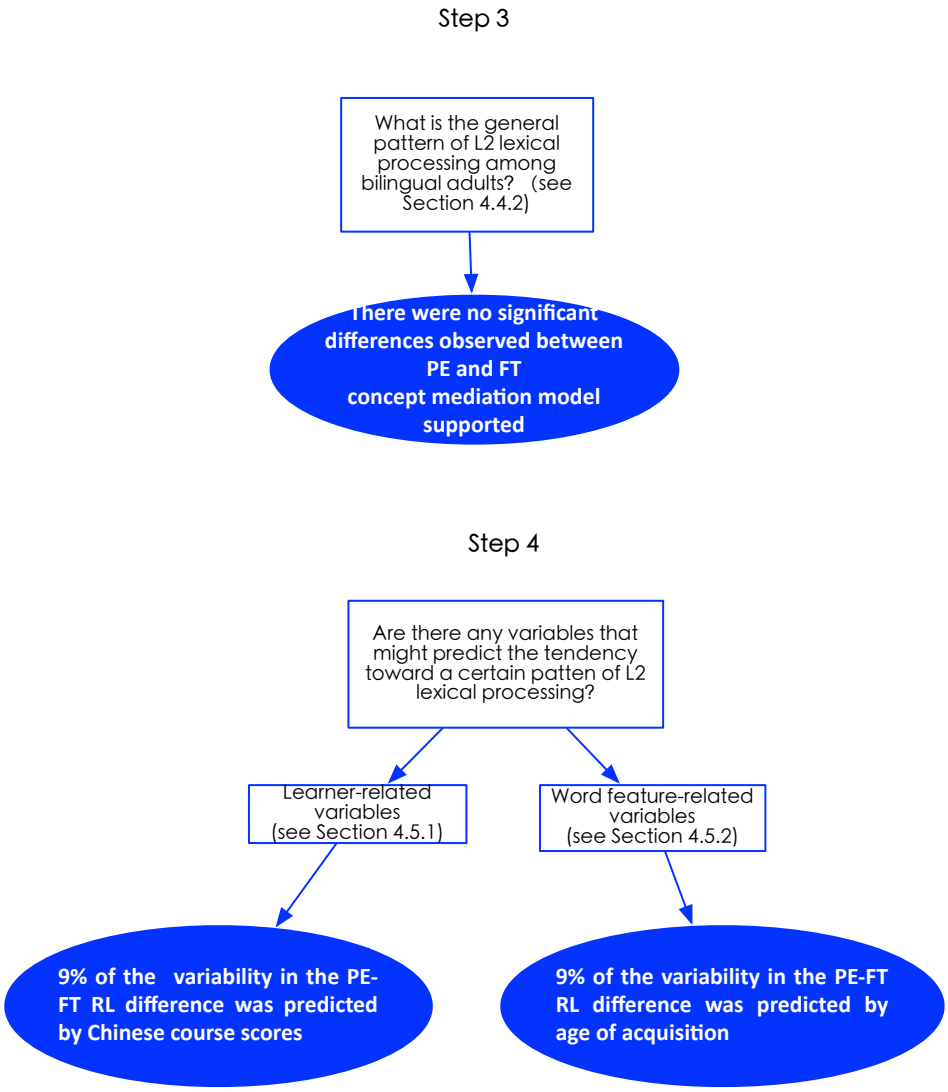
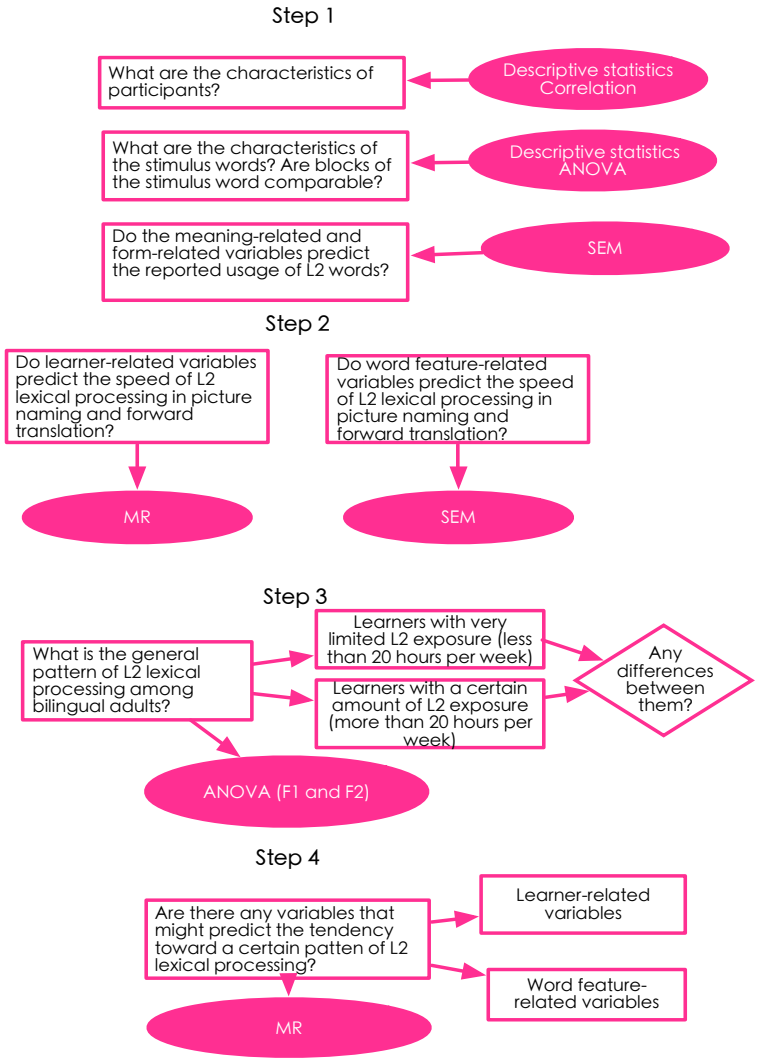


Figure 4. 13 Summary of Chapter 4: Research questions and results (Continued)

Chapter 5 Data Analysis and Results: Adult Participants

Analyses on the adult part were conducted similarly to the child part in terms of statistical choices, analysis procedures, and steps.



Notes: ANOVA: analysis of variance; SEM: structural equation modeling; MR: multiple regression.

Figure 5.1 Overview of research questions and methods of data analysis

5.1 Data screening

In the present study, the adult participants responded to 201 stimulus words, and their response latencies (RL) were recorded. Initially, the data set contained a total of 18894 data points (non-English majors: 12060 data points, and English majors: 6834 data points). Before performing statistical analyses, the entire data set was screened for inaccurate responses, those participants and stimulus items with an exceptionally high error rate, and outliers.

First, incorrect and omitted responses that were recorded manually during experiments were discarded for latency analysis. This deletion resulted in a loss of 7.51% of data points in PE, of 7.69% in FT, and of 1.87% in PC among non-English majors, and resulted in a loss of 5.88% of data points in PE, of 6.63% in FT, and of 1.76% in PC among English majors.

Second, it was found that the highest error rate among non-English majors was 21%, and among English-majors was 17%. Since there was no error rate well beyond the cut-off criteria mentioned in Jiang (2012), all the participants were retained for analysis. Individual items were treated in a similar way to the child participant data, that is, the cut-off error rate was set at 40%. By this criterion, three stimulus words (i.e. butcher, carpenter and priest) were deleted. After this, the percentages of discarded data points were: 6.95% in PE, 7.02% in FT, and 1.76% in PC among non-English majors; and 5.30% in PE, 6.11% in FT, and 1.53% in PC among English majors; and the number of lexical items entered into analysis was 198.

Lastly, outliers were identified and excluded from further statistical analyses similarly to the way in which the child participant data were treated.

After the data screening was complete, the number and percentage of valid responses, and those of deleted responses are summarized in Table 5.1. The missing values were replaced by the means of a participant or by the means of an item within the stimulus condition. Response latencies and accuracies, and word-feature ratings were averaged to generate a summary score for each participant and for each lexical item. It was these values that were entered into analysis instead of the entire data points.

Table 5. 1 Summary of valid and deleted responses

Condition	Non-English majors				English majors			
	Valid		Deleted		Valid		Deleted	
	n	%	n	%	n	%	n	%
PE	3647	92.10	313	7.90	2097	93.47	147	6.53
FT	3640	91.92	320	8.08	2083	92.81	161	7.19
PC	3871	97.75	89	2.25	2195	97.82	49	2.18

5.2 Characteristics of participants and stimulus words

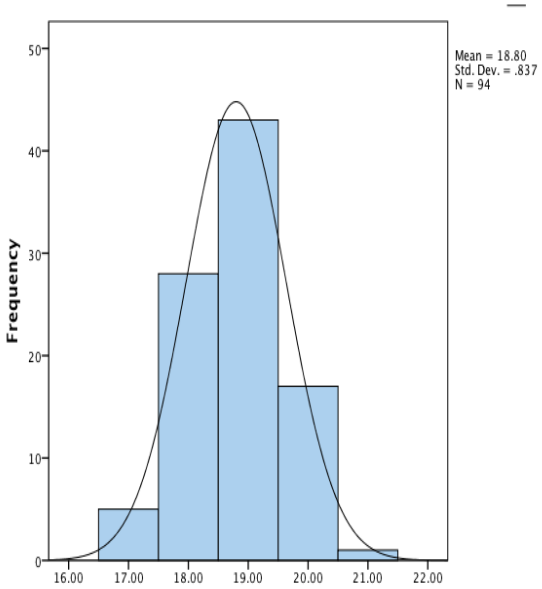
5.2.1 Sample characteristics

A total sample of 94 undergraduates was recruited in the present study. A short questionnaire was designed and used to profile the sample in terms of their demographic characteristics (e.g. age, gender, and ethnicity), and language background (e.g. the amount of L2 exposure per week, the age of starting L2 learning,

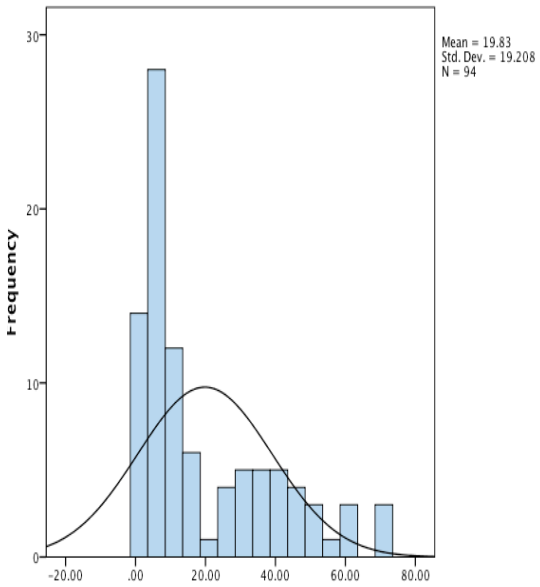
and ways of English learning). The distribution of these learner-related variables is presented in Table 5.2 and Figure 5.2.

Table 5. 2 Distributions of sample characteristics: Adult participants (n = 94)

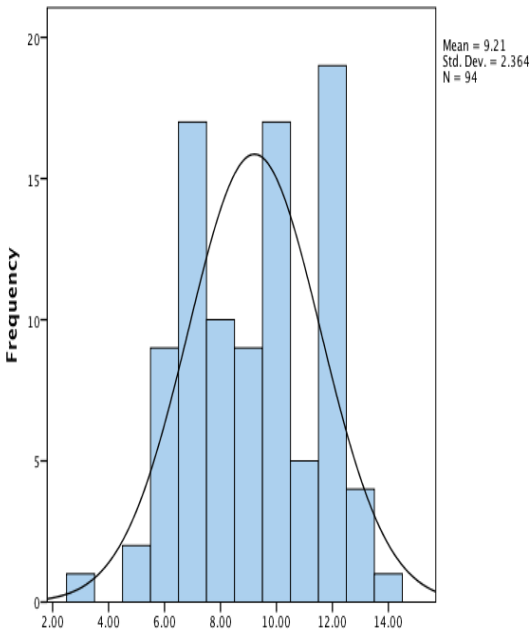
		Frequency	Percentage
Gender			
	Male	33	35.11
	Female	61	64.89
Ethnicity			
	Han	87	92.55
	Minority	7	7.45
Place of birth			
	North China	27	28.72
	East China	18	19.15
	South China	17	18.09
	Southwest China	8	8.51
	Northwest China	2	2.13
	Northeast China	14	14.89
	Southeast China	2	2.13
	Central China	6	6.38
Urban/Rural area			
	Urban	74	78.72
	Rural	20	21.28
Main way of English learning			
	English courses	87	92.55
	Cram courses	5	5.32
	Interaction with Native speakers	2	2.17



Age



Amount of L2 exposure



Age of starting L2 learning

Figure 5. 2 Distributions of sample characteristics

The sample aged from 17 to 20 years (Mean = 18.80, SD = .84), with 60 participants reading for non-English subjects, and the other 34 participants majoring in English. Nearly half of the non-English majors were females (n = 29); by contrast,

the participants majoring in English were overwhelmingly female (n = 32). It should be pointed out that such a gender imbalance among English majors is very commonplace in almost all the Departments of English Language and Literature in China Mainland (Wu, 2012). Moreover, as regards the participants' ethnicity, it was found that the majority were Han people, and the ethnic minority participants took up 7.4% of the total sample. Such a proportion is representative of the overall population in China (Levinson, 1998). As for places of birth, most of the participants came from North China, East China, South China, and Northeast China. Most of the participants had lived in urban areas before they entered universities, and learned English through courses offered at schools and universities. The amount of their L2 exposure ranged widely from 1 hour to 70 hours per week, with the mean being 19.83 hours and with the SD being 19.21 hours. Also, the sample started L2 learning at different ages, ranging from 3 to 14 years old (Mean = 9.21, SD = 2.36).

Moreover, the scores that participants achieved in vocabulary assessments and university entrance examinations were collected so as to profile their L2 vocabulary knowledge, English proficiency levels, as well as academic achievements. Descriptive statistics of the scores, and correlations between each pair of the scores are presented in Tables 5.3 and 5.4.

Taking non-English and English majors together, they had a receptive vocabulary size averaging 4153 words. The vocabulary breadth (VB) and vocabulary depth (VD) were found to be positively correlated with each other ($r = .31$, $p < .001$), but note that

the correlation coefficient was not as high as the one previously reported in the study by Qian (1999) (r = .82). Furthermore, the mean scores of the English proficiency test and of the overall academic test were 130.57 (SD = 8.26), and 599.30 (SD = 29.51) respectively. English proficiency scores were positively correlated with VD (r = .21, $p < .05$), as well as with overall academic achievements (r = .40, $p < .001$).

Table 5. 3 Descriptive statistics of test scores among non-English majors and among English-majors

	Non-English majors (n = 60)				**English majors (n = 34)**				**Total (n = 94)**			
	Mean	SD.	Minimum	Maximum	Mean	SD.	Minimum	Maximum	Mean	SD.	Minimum	Maximum
Vocabulary breadth	4148.33	384.22	3200	4850	4163.24	390.67	3250	4750	4153.72	384.53	3200	4850
Vocabulary depth	73.38	5.29	61	82	74	44.67	53	88	73.61	5.81	53	88
English proficiency level	129.48	9.02	107	145	132.50	6.42	120	143	130.57	8.26	107	145
Academic achievement	606.82	30.93	510	654	586.03	21.43	545	622	599.30	29.51	510	654

Notes: Details of each assessment are provided in Section 3.6.3. The maximum total score for each test is as follows: 5000 (vocabulary breadth), 90 (vocabulary depth), 150 (English proficiency level), and 750 (academic achievement).

Table 5. 4 Pearson correlations (r values) between test scores: Adult participants (n = 94)

	VB	VD	English proficiency	Academic achievement
Vocabulary breadth	1.00	—	—	—
Vocabulary depth	.31***	1.00	—	—
English proficiency	.14	.21*	1.00	—
Academic achievement	-.02	.06	.40***	1.00

Notes: * $p < .05$ (two-tailed), and *** $p < .001$ (two-tailed).

5.2.2 Characteristics of stimulus words

The present study consisted of three blocks of 66 stimulus words, and each block contained the picturable words selected from ten semantic categories, namely: animal (n = 7), nature (n = 7), fruits and vegetables (n = 7), body (n = 7), colors and shapes (n = 6), vehicles (n = 5), clothing (n = 7), actions (n = 7), numbers (n = 7), and occupations (n = 6). Comparisons between three blocks of stimuli were made in terms of a variety of word features so as to determine the extent to which the stimulus blocks were comparable. The values of word features (see Appendix 5.1 for details) were collected either through participant-rated questionnaires or directly from psycholinguistic databases. As for the ratings given by the participants, the rate and re-rate reliability on the repeated 20 items were all found to be high (all r values > .84, and $p < .01$).

The summary statistics of each word feature are presented in Table 5.5. When the statistical assumptions were met, independent-measures ANOVAs were used to compare the mean values of word features across three blocks. The normality of the variable NoRS and SND was seriously violated (i.e. skewness or kurtosis values > + 2.00 or < - 2.00); therefore, a log-transformation was applied to normalize these variables. A summary of the ANOVA results is provided in Table 5.6 (see Appendix 5.2 for statistical details). In general, no statistically significant difference was found between blocks in terms of WC, WT, Log-SND, Log-NoRS, NoL, NoP, NoS, SWF, and AoA, indicating that ach block of stimuli was comparable to the others in terms

of all the word features examined, and such a similarity would corroborate any conclusions to be reached in the present study.

Table 5. 5 Summary statistics: Lexical features of each block (n = 198)

Variables		n	Mean	SD.	Minimum	Maximum
WC	Block 1	66	4.64	.92	2.66	5.74
	Block 2	66	4.40	.99	2.30	5.56
	Block 3	66	4.32	.98	2.25	5.46
WT	Block 1	66	3.60	.72	2.12	4.80
	Block 2	66	3.49	.75	1.84	4.80
	Block 3	66	3.48	.72	1.86	4.74
Log-SND	Block 1	66	.58	.62	.00	2.40
	Block 2	66	.65	.51	.00	1.77
	Block 3	66	.71	.56	.00	2.81
Log-NoRS	Block 1	66	.48	.26	.00	1.34
	Block 2	66	.45	.27	.00	1.18
	Block 3	66	.44	.28	.00	1.15
NoL	Block 1	66	5.23	1.73	3.00	10.00
	Block 2	66	5.52	1.64	3.00	10.00
	Block 3	66	5.48	1.76	3.00	10.00
NoP	Block 1	66	4.22	1.42	2.00	8.00
	Block 2	66	4.30	1.59	1.00	8.00
	Block 3	66	4.29	1.43	2.00	9.00
NoS	Block 1	66	1.45	.64	1.00	3.00
	Block 2	66	1.55	.75	1.00	4.00
	Block 3	66	1.52	.66	1.00	4.00
SWF	Block 1	66	3.94	.69	2.55	5.69
	Block 2	66	3.83	.72	2.53	5.65
	Block 3	66	4.08	.75	2.78	5.75
AoA	Block 1	66	11.39	2.50	9.00	17.00
	Block 2	66	12.17	2.87	9.00	17.00
	Block 3	66	12.08	2.98	8.00	17.00

Table 5. 6 Comparing stimulus words between blocks: ANOVA results (n = 198)

Variables	df	Mean square	F	Sig.	η_p^2
WC	2	1.86	2.00	.14	.020
	195	.93			
WT	2	.29	.55	.58	.006
	195	.54			
Log-SND	2	.27	.85	.43	.010
	195	.32			
Log-NoRS	2	.03	.43	.65	.004
	195	.07			
NoL	2	1.65	.56	.57	.010
	195	2.94			
NoP	2	.11	.05	.95	.000
	195	2.19			
NoS	2	.14	.30	.74	.003
	195	.49			
SWF	2	.97	1.88	.16	.020
	195	.52			
AoA	2	11.77	1.51	.22	.020
	195	7.81			

Together with the comparisons made on the stimulus words across blocks, the visual complexity of pictorial stimuli was also compared across blocks. The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the visual complexity of the pictures, F (2, 195) = .008, p = .99, η_p^2 = .000. Moreover, all the three blocks were matched with regard to the usage frequency of L1 words according to the Language Corpus System of Modern Chinese Study (LCSMCS, www.dwhyyjzx.com/cgi-bin/yuliao/), F (2, 195) = .15, p = .86, η_p^2 = .002.

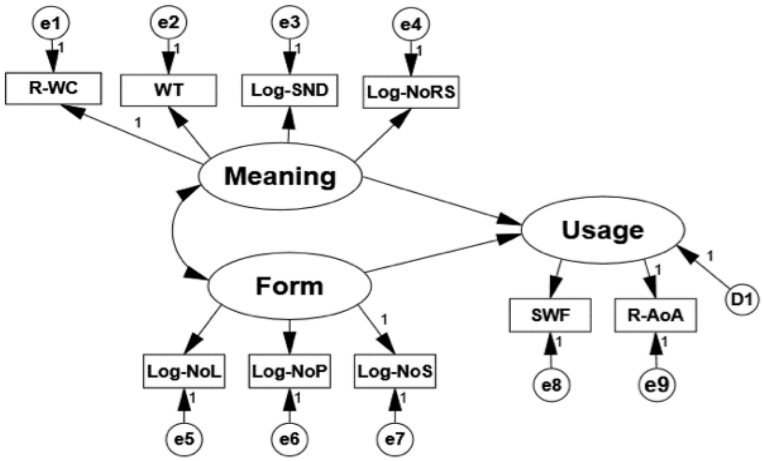
5.2.3 Relationships between meaning-related variables, Form level and Usage level of L2 words: Adult L2 learners

This analysis answers the following research question:

Do the meaning-related and form-related variables predict the reported usage of L2 words?

5.2.3.1 Hypothesized SEM model

The hypothesized model is presented in Figure 5.3 below, including a Meaning factor (word concreteness, word typicality, semantic neighborhood density, and the number of related senses), a Form factor (the number of letters, phonemes, and syllables), and a Usage factor (subjective word frequency and age of acquisition).



Note: R: recoded

Figure 5. 3 Hypothesized SEM model: Do Meaning and Form predict the Usage of L2 words among adult L2 learners?

Note that WC and AoA were found to underpin their corresponding latent variable in an opposite direction as compared to other variables within the same construct. Thus, the ordinal variable WC was recoded, and the interval variable AoA

were standardized using a Z-score transformation and then recoded; in so doing, all the variables measuring the same construct were scored in the same direction for analysis. How lexical variables are correlated with each other is presented in Table 5.7.

Table 5. 7 Correlation matrix of word feature-related variables: Adult participants (n = 198)

	1	2	3	4	5	6	7	8	9
1. R-WC	1								
2. WT	.10	1							
3. Log-SND	.38**	.30**	1						
4. Log-NoRS	.27**	.07	.21**	1					
5. Log-NoL	-.06	-.28**	-.30**	-.40**	1				
6. Log-NoP	-.11	-.20**	-.30**	-.36**	.78**	1			
7. Log-NoS	-.15*	-.13	-.23**	-.32**	.64**	.71**	1		
8. SWF	.26**	.56**	.41**	.28**	-.24**	-.26**	-.16*	1	
9. R-AoA	.16*	.55**	.36**	-.02	-.13	-.09	-.07	.53**	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

Before performing formal analyses, the assumptions of SEM were checked, including the sample size, missing data, normality and linearity, outliers, and multicollinearity and singularity (see Appendix 5.3 for details). There was no serious violation noted except that one multivariate outlier was identified and removed from further analysis ($p < .001$), and that the critical ratios of NoL, NoP, and NoS were beyond the recommended limits +/-3.00 (Tabachnick & Fidell, 2012) so log-transformations were applied to normalize them.

5.2.3.2 Measurement model

As mentioned earlier, a two-step procedure was used to assess how well the hypothesized model described the observed data. Therefore, the measurement model (see Figure 5.4) was estimated first in terms of its construct, convergent, discriminant validity, internal reliability, and construct reliability before the structural model was evaluated.

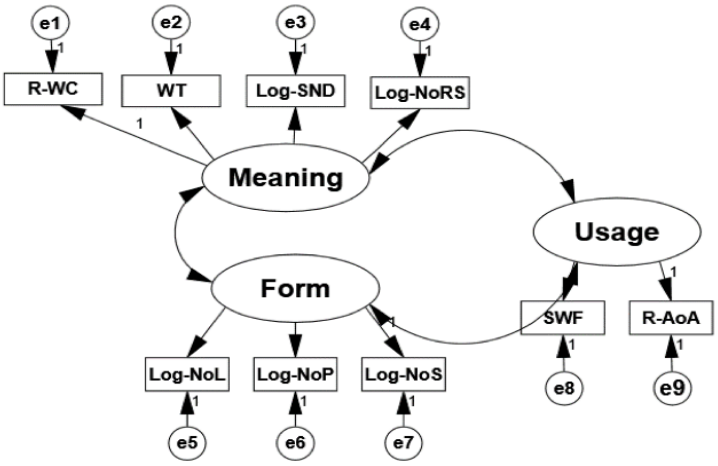


Figure 5. 4 Measurement model: Meaning, Form and Usage of L2 words among adult L2 learners

To test the measurement model, a CFA with maximum likelihood (ML)

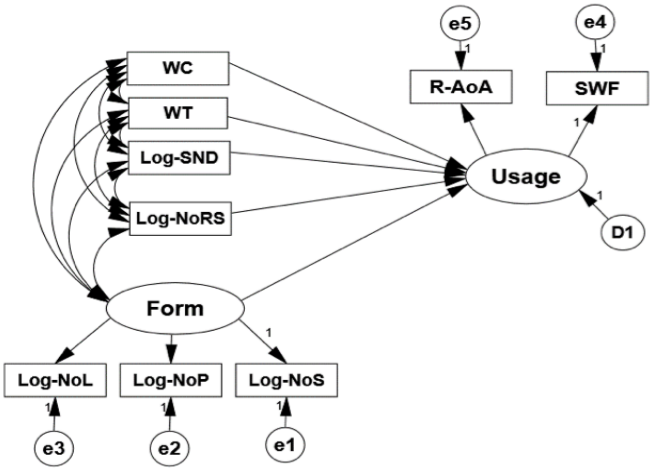
estimation method was conducted using IBM SPSS AMOS v.20. The ML was appropriate for use here as no serious violation of statistical assumptions was noted. The overall model fit was not fully satisfactory: χ^2(24, N = 197) = 87.29, p = .000, CFI = .904, RMSEA = .116. The model with standardized coefficients is presented in Appendix 5.4. The factor loadings of WC and Log-NoRS to the underlying latent variable were less than .50, indicating that convergent validity of the construct of meaning is not acceptable. Thus, the researcher decided to treat meaning-related variables as separate manifest variables rather than to combine them behind a latent factor. After meaning-related variables were excluded from the model, a very good overall model fit was noted (i.e. construct validity), χ^2(4, N = 197) = 5.291, p = .259, CFI = .997, RMSEA = .041.

In order to assess reliability, *Cronbach's alpha* and *construct reliability* (CR) coefficients were estimated. An alpha value of .70 or above (Nunnally, 1978), and the CR estimate of .70 or above (Hair et al., 2005) suggest good reliability. Sufficient or nearly sufficient alpha and CR coefficients were observed with Usage, $\alpha =$ .69, CR = .79, and with Form $\alpha =$ .88, CR = .84. There was a low correlation between the latent variable Form and Usage (r = -.27), indicating that discriminant validity was secured in this measurement model. How these validity and reliability estimates were defined and calculated has been provided in Section 3.7.2. It should be pointed out, as B. M. Byrne (2009) warned, that once modifications are made to the hypothesized model, all analyses henceforth tend to be exploratory rather than confirmatory per se.

After securing the measurement models' validity and reliability, analysis moved on to the structural model.

5.2.3.3 Re-specified SEM model

The re-specified SEM model is presented in Figure 5.5.



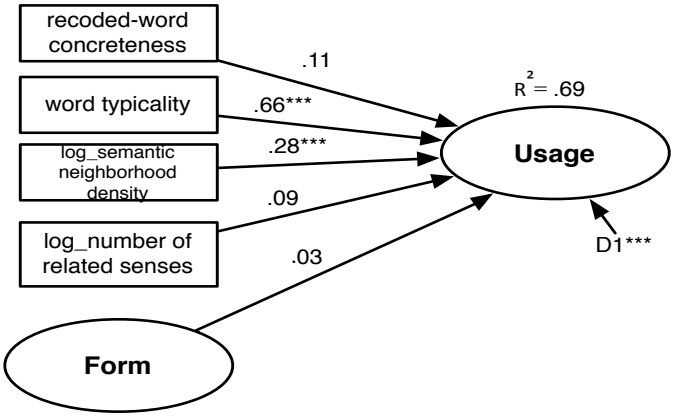
Note: R: recoded

Figure 5. 5 Re-specified SEM model: Do Meaning and Form predict the Usage of L2 words among adult L2 learners?

ML method was employed to estimate the overall fit of the re-specified SEM model. The fit indices were satisfactory, $\chi^2(16, N = 197) = 35.432$, $p = .003$, CFI = .970, RMSEA = .079. The model with standardized coefficients is diagrammed in Figure 5.6 below. The output of IBM SPSS AMOS is presented in Appendix 5.5.

As is evident from Figure 5.6, Usage increased as the degree of word typicality and semantic neighborhood density increased (standardized coefficient = .66 and .28 respectively, $p < .001$). Word concreteness, the number of related senses, and Form did not make significant contributions to predicting Usage (standardized coefficient = .11, .09, and .03 respectively, $p = .106$, .207, and .725 respectively). The R^2 value

was .69 (a very large effect size), indicating that more than half (69%) of the variance in Usage was accounted for by the lexical variable included.



Notes: *** $p < .001$; for clarity reasons, the measurement portion of the model and correlations between manifest lexical variables are not presented.

Figure 5. 6 Final SEM model with standardized coefficients: Do Meaning and Form predict the Usage of L2 words among adult L2 learners?

5.2.3.4 Interim summary: Relationships between lexical variables

This analysis showed that word typicality and semantic neighborhood density had significant predictive effects on Usage. To illustrate, those words that highly represent its category are usually used more often and for a longer period of time. Additionally, the adult data suggested that used-more-often and early-acquired L2 words had higher neighborhood density relative to used-less-often and late-acquired L2 words. Contrary to our expectation, word concreteness effects were not observed in this experiment probably due to the predominant use of picturable stimulus words. The contribution made by the number of related sense to word usage was not significant among the adult participants. Analysis on the predictive effect of form on word usage suggests

that words with simpler forms are not always used more often.

5.3 The predictive effects of learner- and word feature-related variables on L2 lexical processing among adult L2 learners

This analysis focuses on examining the specific relationships between several learner-related variables and the speed of L2 lexical processing in picture naming and forward translation by using standard MR. This analytical strategy was used, as mentioned earlier, because there was a lack of theory to hypothesize a model specifying which and in what manner predictors might be related to the outcomes. MR analyses were followed by an investigation of the predictive effects of a set of word feature-related variables on the speed of L2 lexical processing through the use of SEM.

5.3.1 Predictive effects of learner-related variables on L2 lexical processing among adult L2 learners

This analysis addresses the following research question:

Do learner-related variables predict the speed of L2 lexical processing in picture naming and forward translation among adult L2 learners?

Before performing formal analyses, descriptive statistics for all the learner-related variables, and the RL of PE and that of FT were examined, including their means, SDs, and frequency histograms wherever appropriate. Also, a correlation matrix among continuous predictors and the outcome variables was calculated and provided in Table 5.8 below. The strength of association between nominal predictors and the

outcome variables was measured using Lambda, and the results are reported in Appendix 5.6. These three IVs (age, vocabulary depth, and the age of starting L2 learning) were significantly correlated with PE response latencies, and these four IVs (age, vocabulary depth, the age of starting L2 learning and the amount of exposure) were significantly correlated with FT response latencies. The variables that were significantly correlated with the DVs were included in the regression model. Moreover, the assumptions of MR were checked, including the sample size, missing data, normality, linearity, and homoscedasticity of residuals, outliers, and multicollinearity and singularity (see Appendix 5.7 for statistical details), with no serious violation found.

Table 5. 8 Correlations between learner-related variables, and PE and FT performance in adult L2 learners (n = 94)

A. Pearson's r correlations: PE response latencies

Variables	1	2	3	4	5	6	7	8
1.PE	1	—	—	—	—	—	—	—
2. Age	.29*	1	—	—	—	—	—	—
3. Vocabulary depth	-.30**	-.17	1	—	—	—	—	—
4. Vocabulary size	-.18	.01	.31**	1	—	—	—	—
5. English proficiency	-.11	.12	.21*	.14	1	—	—	—
6. Academic achievement	.01	.07	.06	-.02	.40**	1	—	—
7. Age of starting L2 learning	.20*	.15	.09	.03	.38**	.18	1	—
8. Amount of exposure	-.14	.00	.09	-.02	.11	-.40**	.12	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

B. Pearson's r correlations: FT response latencies

Variables	1	2	3	4	5	6	7	8
1. FT	1	—	—	—	—	—	—	—
2. Age	.25*	1	—	—	—	—	—	—
3. Vocabulary depth	-.34**	-.17	1	—	—	—	—	—
4. Vocabulary size	-.18	.00	.31**	1	—	—	—	—
5. English proficiency	-.15	.12	.21*	.14	1	—	—	—
6. Academic achievement	.17	.07	.06	-.02	.40**	1	—	—
7. Age of starting L2 learning	.21*	.15	.09	.03	.38*	.18	1	—
8. Amount of exposure	-.25*	.00	.09	-.02	.11	-.40**	.12	1

Notes: * $p < .05$ (two-tailed), and ** $p < .01$ (two-tailed).

5.3.1.1 Results: Predictive effects of learner-related variables on L2 lexical processing

Table 5.9 and Table 5.10 display the correlations between each pair of variables, the unstandardized regression coefficients (*B*) and intercept, the standardized regression coefficients (β), the semi-partial correlations (sr_i^2), R, R^2, and adjusted R^2. The adjusted R^2 were also reported because it corrects for the sample size and for the number of variables included in the model.

In the regression analysis of learner-related variables that might predict PE response latencies (see Table 5.9 below), the R (.41) was significantly different from zero, F (3, 90) = 6.05, $p < .01$, with R^2 at .17 (an effect size between medium and large). The adjusted R^2 value of .14 indicated that 14% of the variability in PE response latencies was predicted by the included learner-related variables. The size and direction of the relationships suggested that those who had deeper vocabulary knowledge were likely to process L2 lexical items faster in PE ($\beta = -.29$, $p < .01$), and those who started L2 learning earlier were likely to process L2 lexical items faster in PE ($\beta = .20$, $p < .05$). Between those two significant predictors, it was found that the VD served as a more powerful predictor, as indicated by the squared semi-partial correlations.

In the regression analysis of learner-related variables that might predict FT response latencies (see Table 5.10 below), the R (.51) was significantly different from zero, F (4, 89) = 7.89, $p < .01$, with R^2 at .26 (a large effect size). The adjusted R^2 value of .23 indicated that more than 20% of the variability in FT response latencies

was predicted by the learner-related variables included. The size and direction of the relationships suggested that those who had deeper vocabulary knowledge ($\beta = -.31$, $p < .01$), had greater amount exposure to the L2 ($\beta = -.25$, $p < .01$), and started L2 learning earlier ($\beta = .25$, $p < .05$) were likely to produce L2 lexical items faster in FT. Among those three significant predictors, it was found that the VD served as a more power predictor, as indicated by the squared semi-partial correlations.

5.3.1.2 Interim summary: Learner-related variables and L2 lexical processing

MR analyses indicated that it was L2 adult learners' vocabulary depth that best predicted the speed of their lexical processing. This result seems to be in contrast to that of child L2 learners in whom vocabulary size was found to be the strongest predictor over and beyond other learner variables. In addition to vocabulary depth, age of onset was significantly predictive of naming and translation latencies in adult L2 learners, a result that consists with the important role that age of acquisition played in determining how fast individual lexical items were processed. Additionally, the amount of L2 exposure contributed significantly to predicting translation times but not to picture naming speed probably due to the essential differences that exist between these two types of tasks.

Table 5. 9 Standard MR: Predictive effects of learner-related variables on PE response latencies among adult L2 learners (n = 94)

Variables	PE	Age	Vocabulary depth	Age of starting L2 learning	*B*	*β*	sr^2 (unique)
Age	.29*				51.37	.17	
Vocabulary depth	-.30**	-.17			-12.60**	- .29	.08
Age of starting L2 learning	.20*	.15	.09		21.49*	.20	.04
					Intercept= 1311.13		
Means	1547.28	18.80	73.61	9.21			
Standard deviations	256.72	.84	5.81	2.36			
						R^2= .17	
						Adjusted R^2= .14	
						R= .41**	

Notes: * $p < .05$, and ** $p < .01$;

B: unstandardized regression coefficients;

β: standardized regression coefficients;

$sr^{2:}$ semi-partial correlations.

Table 5. 10 Standard MR: Predictive effects of learner-related variables on FT response latencies among adult L2 learners (n = 94)

Variables	FT	Age	Vocabulary depth	Amount of exposure	Age of starting L2 learning	*B*	*β*	sr^2 (unique)
Age	.25*					44.60	.16	
Vocabulary depth	-.34**	-.17				-12.26**	-.31	.09
Amount of exposure	-.25*	.003	.09			-3.00**	-.25	.06
Age of starting L2 learning	.21*	.15	.09	.12		23.86*	.25	.06
					Intercept = 1232.58			
Means	1328.41	18.80	73.61	19.83	9.21			
Standard deviations	229.64	.84	5.81	19.21	2.36			
							R^2= .26	
							Adjusted R^2= .23	
							R= .51**	

Notes: * $p < .05$, and ** $p < .01$;
B: unstandardized regression coefficients;
β: standardized regression coefficients;
$sr^{2:}$ semi-partial correlations.

5.3.2 Predictive effects of word feature-related variables on L2 lexical processing among adult L2 learners

This analysis answers the following research question:

Do word feature-related variables predict the speed of L2 lexical processing in picture naming and forward translation among adult L2 learners?

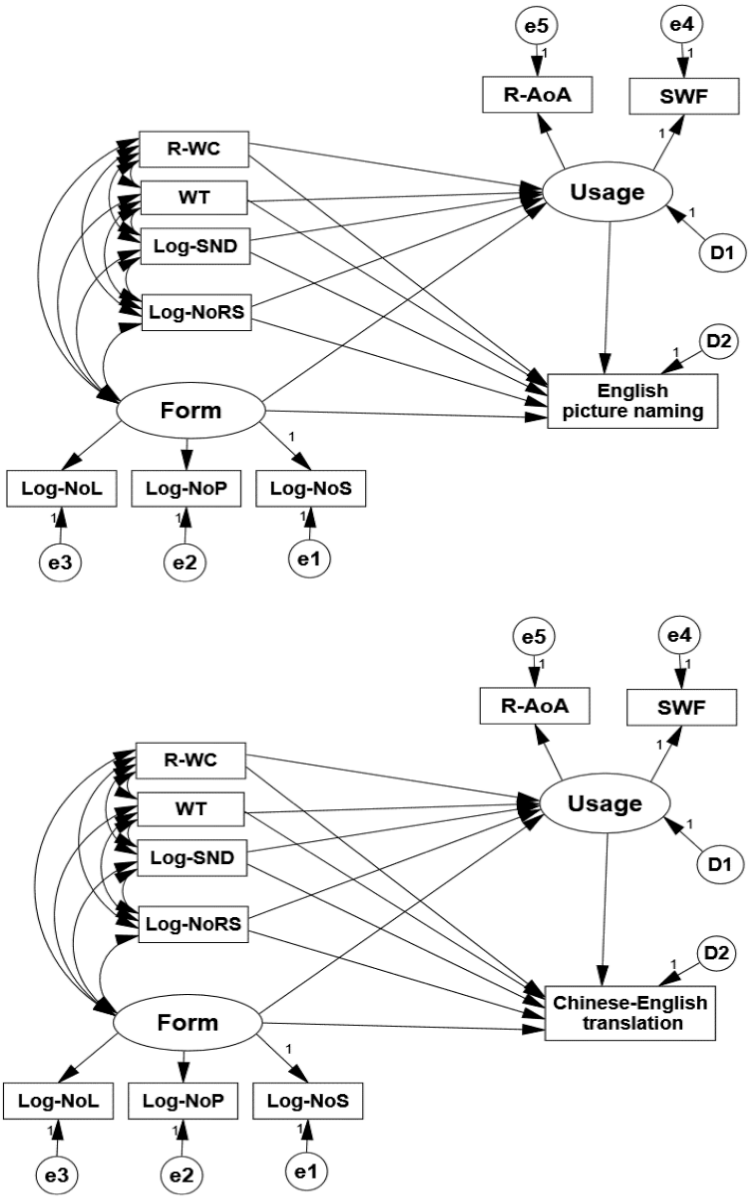
Two SEM models were estimated in such a way that was similar to that described in Section 5.2.3.

5.3.2.1 Model specification

The hypothesized model (see Figure 5.7 below) contained two latent factors: Form (the number of letters, phonemes, and syllables) and Usage (subjective word frequency and age of acquisition), and four meaning-related manifest variables, namely: word concreteness, word typicality, semantic neighborhood density, and the number of related senses.

These word feature-related latent and observed variables were assumed to predict PE and FT response latencies. Substantively, this assumption is empirically supported by a series of studies examining the impacts of lexical variables on the processing of L1 and L2 words in monolingual and bilingual adults (e.g. Alario et al., 2004; Balota & Chumbley, 1984; Cuetos et al., 1999; De Groot, 1992a; De Groot et al., 1994; De Groot & Keijzer, 2000; Hino & Lupker, 1996; Lichacz et al., 1999; Prior, MacWhinney, & Kroll, 2007). As is conventionally done, five variables (i.e. WC, WT, SND, NoRS, and Form) were allowed, initially, to freely covary. This model was

over-identified with 19 dfs. As pointed out in analyzing the child participant data, variables such as the visual complexity of stimulus pictures were not included in the model because they were not the focus of the present study.



Note: R: recoded

Figure 5. 7 Hypothesized SEM model: Predictive effects of word feature-related variables on PE and FT response latencies in adult L2 learners

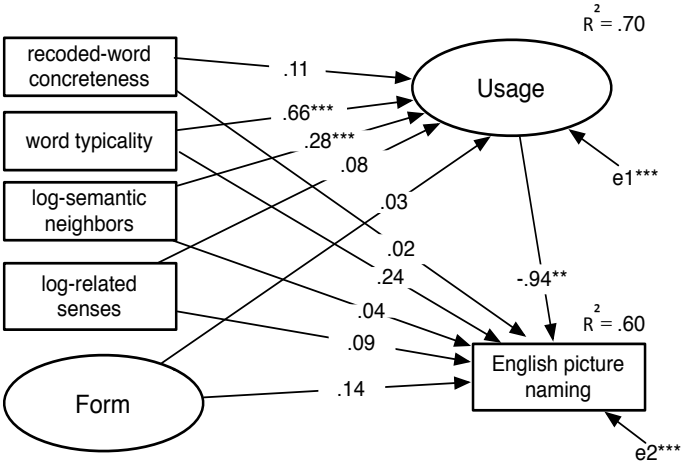
5.3.2.2 Evaluation of assumptions for the specified SEM model

All the assumptions necessary for the conduct of SEM were checked (see Appendix 5.8 for details), which indicated that it was acceptable to proceed with the SEM with the obtained size of 197 words although the desirable size was not reached, according to Bentler and Chou's recommendation (1987). There were no univariate and multivariate outliers observed. The skewness and kurtosis values of the variables all fell between +2.00 and -2.00, and no inappropriate level of multicollinearity and singularity was noted. Thus, for this analysis, ML estimates were obtained using IBM SPSS AMOS v.20.

5.3.2.3 SEM model estimation and results

PREDICTIVE EFFECTS OF WORD FEATURE-RELATED VARIABLES ON PE RESPONSE LATENCIES

The independence model that tests the hypothesis that all variables are uncorrelated was easily rejected, $\chi^2(45, N = 197) = 810.94$, $p < .001$, CFI = .000, RMSEA = .295. The hypothesized model was estimated with ML estimation, and the result suggested that the overall model fit was considered acceptable, $\chi^2(19, N = 197) = 41.29$, $p = .002$, CFI = .971, RMSEA = .077. The model with standardized coefficients is diagrammed in Figure 5.8. The output of IBM SPSS AMOS is presented in Appendix 5.9.



Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$; for clarity reasons, the measurement portion of the model is not presented; R: recoded.

Figure 5. 8 Final SEM model with standardized coefficients: Predictive effects of word features on PE response latencies among adult L2 learners

Direct effects. As shown in Figure 5.8, PE response latencies decreased significantly as the latent variable Usage increased (standardized coefficient = -.94, $p < .01$). Word concreteness, word typicality, semantic neighborhood density, the number of related senses, and Form did not significantly predict the response latency of PE (standardized coefficient = .02, .24, .04, .09, and .14 respectively, $p = .78$, .13, .77, .31, and .11 respectively).

Indirect effects. The results suggested that the indirect effects from word typicality and semantic neighborhood density to PE response latencies via Usage were significant (standardized coefficient = -.63 and -.27, both *ps* < .001). The indirect effects from word concreteness, the number of related senses, and Form to PE response latencies via Usage were not significant (standardized coefficient = -.10, -.08, and -.03 respectively, $p = .10$, .34, and .67 respectively).

Total effects. The standardized total effect of word typicality on PE response latencies was statistically significant (standardized coefficient = -.39, $p < .001$), the total effect of semantic neighborhood density on PE response latencies was statistically significant (standardized coefficient = -.23, $p < .001$), and the total effect of Usage on PE response latencies was significant (standardized coefficient = -.94, $p < .01$). The total effects of word concreteness, the number of related senses, and Form on PE response latencies were not statistically significant (standardized coefficients= -.08, .01, and .11 respectively, p = .22, .89 and .14 respectively). A summary of the decomposition of standardized effects is presented in Table 5.11.

Table 5. 11 Decomposition of standardized effects for structural models of Figure 5.8

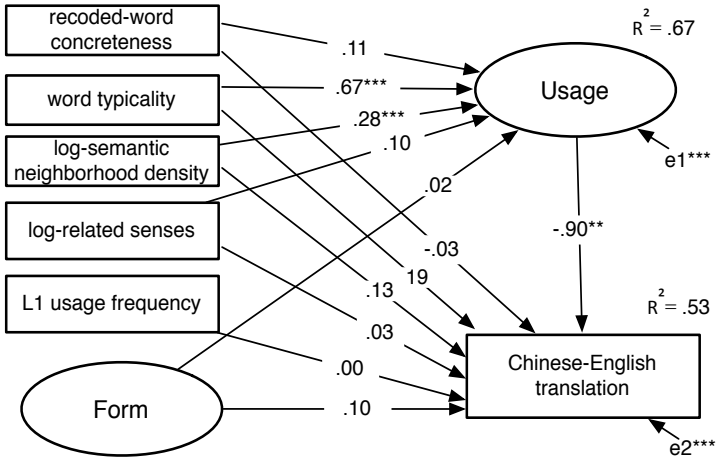
	Direct effects	**Indirect effects**	**Total effects**
R-WC	.02	-.10	-.08
WT	.24	-.63***	-.39 ***
Log-SND	.04	-.27***	-.23***
Log-NoRS	.09	-.08	.01
Form	.14	-.03	.11
Usage	-.94**	–	-.94**

Note: * $p < .05$; ** $p < .01$, *** $p < .001$.

Per cent of variances accounted for. The R^2 value was .60 (a large effect size), suggesting that 60% of the variability in PE response latencies was accounted for by the included meaning-related variables, Form and Usage.

PREDICTIVE EFFECTS OF WORD FEATURE-RELATED VARIABLES ON FT RESPONSE LATENCIES

The independence model that tests the hypothesis that all variables are uncorrelated was easily rejected, $\chi^2(45, N = 197) = 791.11$, $p < .001$, CFI = .000, RMSEA = .291. The hypothesized model was then estimated with ML estimation, and the result suggested that the overall model fit was considered reasonable, $\chi^2(19, N = 197) = 37.19$, $p < .001$, CFI = .976, RMSEA = .070. The model with standardized coefficients is diagrammed in Figure 5.9. The output of IBM SPSS AMOS is presented in Appendix 5.9.



Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$; for clarity reasons, the measurement portion of the model is not presented.

Figure 5. 9 Final SEM model with standardized coefficients: Predictive effects of word features on FT response latencies among adult L2 learners

Direct effects. As shown in Figure 5.8, PE response latencies decreased significantly as the latent variable Usage increased (standardized coefficient = -.90, p

< .01). Word concreteness, word typicality, semantic neighborhood density, the number of related senses, and Form did not significantly predict the response latency of PE (standardized coefficient = -.03, .19, .13, .03, and .10 respectively, p = .72, .22, .18, .74, and .24 respectively).

Indirect effects. Given that the paths from WC, WT, SND, NoRS, and Form to PE response latencies via Usage were statistically significant. So the indirect effects of these variables on PE response latencies via Usage were explored. The results suggested that the indirect effect from word typicality to FT response latencies via Usage was statistically significant (standardized coefficient = -.59, $p < .001$) and the indirect effect from semantic neighborhood density to FT response latencies (standardized coefficient = -.25, $p < .01$). The indirect effect from word concreteness, the number of related senses, and Form to FT response latencies via Usage was not significant (standardized coefficient = -.10, -.09, and -.02 respectively, and p = .09, .23, and .73 respectively).

Total effects. The total effect of word typicality on FT response latencies was statistically significant (standardized coefficient = -.40, $p < .001$) and the total effect of Usage on FT response latencies was statistically significant (standardized coefficients = -.90, $p < .01$). The total effect of word concreteness, semantic neighborhood density, the number of semantic related senses, and Form on FT response latencies was not statistically significant (standardized coefficient = -.13, -.12, -.05, and .08 respectively, p = .07, .10, .44 and .32 respectively). A summary of

the decomposition of standardized effects is presented in Table 5.12.

Table 5. 12 Decomposition of standardized effects for structural models of Figure 5.9

	Direct effects	Indirect effects	Total effects
R-WC	-.03	-.10	-.13
WT	.19	-.59***	-.40***
Log-SND	.13	-.25**	-.12
Log-NoRS	.03	-.09	-.05
Form	.10	-.02	.08
Usage	-.90**	–	-.90**

Note: * $p < .05$; ** $p < .01$, *** $p < .001$.

Percent of variances accounted for. The R^2 value was .53 (a large effect size) suggested that 53% of the variability in FT response latencies was accounted for by the included meaning-related, form-related and usage-related variables.

5.3.2.4 Interim summary: Predictive effects of lexical variables on processing efficiency

A cross-task comparison showed that the included lexical variables accounted for similar amounts of variance across picture naming and translation. At the conceptual-semantic level, the participants' semantic networks were found to have significant predictive effects on L2 lexical processing via word usage. Conceivably, as the strength of semantic nodes becomes stronger, learners are more able to access target items rapidly as long as these items have been practiced sufficiently before. Word typicality contributed only indirectly to naming and translation latencies, suggesting the reduced role that the conceptual structure plays in directly accessing L2 words with increasing age and proficiency. Word usage acted as a mediator

through which typicality effects occurred. The latent variable Form, represented by a combination of Log-NoL, Log-NoP, and Log-NoS, did not serve as a significant predictor of naming and translation speed. At the usage level, its total effects on response latencies were significant, with β = -.94 and β = -.90 in picture naming and translation respectively.

5.4 General pattern of L2 Lexical processing among adult L2 learners

This analysis answers the following research questions:

What is the general pattern of L2 lexical processing among adult learners with very limited exposure to L2 (less than 20 hours per week)?

What is the general pattern of L2 lexical processing among adult learners with a certain amount of exposure to L2 (more than 20 hours per week)?

Before examining the general pattern of L2 lexical processing among bilinguals, the results of the preparation study are reported first. The main aim of that study was to compare the encoding speed of pictures with that of Chinese characters (see Section 3.6.2 above for details on the experiment).

5.4.1 Speed and accuracy of encoding pictures and Chinese characters

The speed and accuracy of encoding PE and FT were compared across conditions among the adult participants (n = 28) with the use of F1 (by participants) and of F2 (by items) repeated-measures ANOVAs. The statistical assumptions were checked. The skewness and kurtosis values of the variables all fell between +2.00 and -2.00,

without any serious violations of the normality noted. There were no outliers identified far beyond +/- 3.00 SD. The treatment of missing values here were the same as the way in which RL data were treated, as reported in Section 5.1.

Speed of encoding

BY PARTICIPANTS

The descriptive statistics of the speed of encoding by participants across two stimulus conditions are presented in Table 5.13.

Table 5. 13 Descriptive statistics of encoding speed (in ms) by adult participants between conditions

Conditions	n	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Pictures	28	799.85	212.34	523.27	1378.61	1.21	1.20
Characters	28	766.14	162.33	526.52	1197.76	.91	.57

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-subjects factor. The ANOVA result showed that there was no significant effect of stimulus conditions on the speed of encoding by participants, F1 (1, 27) = 2.88, p = .10, with an effect size between medium and large (η_p^2 = .10), indicating that the speed of encoding pictures was not significantly different from that of encoding Chinese characters.

BY ITEMS

The descriptive statistics of the speed of encoding by participants between two stimulus conditions are presented in Table 5.14.

Table 5. 14 Descriptive statistics of encoding speed (in ms) by items between conditions: Adult participants

Conditions	n	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Pictures	60	824.26	104.86	632.79	1069.21	.17	-.32
Characters	60	788.33	119.04	615.43	1129.56	.97	.59

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-items factor. The ANOVA result showed that there was no significant effect of stimulus conditions on the speed of encoding by items, F2 (1, 59) = 2.41, p = .13, with an effect size between small and medium (η_p^2 = .04), indicating that the speed of encoding pictures was not significantly different from that of encoding Chinese characters by items.

Accuracy of encoding

BY PARTICIPANTS

The descriptive statistics of the accuracy of encoding by participants between two stimulus conditions are summarized in Table 5.15.

Table 5. 15 Descriptive statistics of accuracy of encoding (in percentage) by adult participants between conditions

Conditions	n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Pictures	28	98.21	93.33	100	2.64	-1.07	-.47
Characters	28	98.45	93.33	100	2.48	-1.29	.19

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-subjects factor. The ANOVA result showed that there was no significant effect of stimulus conditions on the accuracy of encoding by participants, F1 (1, 27) = .16, p = .69, with a small effect size (η_p^2 = .01). The result suggested that the accuracy of encoding did not vary significantly with the stimulus condition.

BY ITEMS

The descriptive statistics of the accuracy of encoding by items across two stimulus conditions are presented in Table 5.16.

Table 5. 16 Descriptive statistics of accuracy of encoding (in percentage) by items between conditions: Adult participants

Conditions	n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Pictures	60	99.39	96.67	100	1.06	-1.54	2.01
Characters	60	99.47	96.67	100	.99	-1.75	1.19

A one-way repeated-measures ANOVA was conducted, with the stimulus condition as a within-items factor. The ANOVA result showed that there was no significant effect of stimulus conditions on the accuracy of encoding by items, F2 (1, 59) = .23, p = .64, with a small effect size (η_p^2 = .004). The result suggested that encoding picture was not significantly different from encoding Chinese characters by items in terms of the accuracy.

To sum up, the preparation study indicated that encoding Chinese characters was comparable to encoding pictures among the adult participants in terms of the speed

and accuracy, the result of which would corroborate the internal validity of the research, as well as validate the results to be reached.

5.4.2 General pattern of L2 lexical processing among non-English majors

Response latencies and accuracies were compared across conditions among non-English majors (n = 60) with the use of F1 (by participants) and of F2 (by items) mixed-model ANOVAs. The statistical assumptions were checked, indicating that the skewness and kurtosis values of the variables all fell between +2.00 and -2.00, without any serious violations of the normality observed. There were no outliers beyond the recommended limits (+/-3.00 SD). The assumption of homogeneity was tested with the Levene's test, and violations of the homogeneity of variance were noted in comparing the RAs across blocks by items. However, according to Howell (2011), ANOVA procedure was considered to be robust to violations of the homogeneity assumption, provided that group sizes are approximately equal. The assumption of sphericity was checked with the Mauchly's test, and Greenhouse-Geisser or Huynh-Feldt corrections for the violation of sphericity of covariance were made wherever appropriate.

5.4.2.1 Response latencies: Non-English majors

BY PARTICIPANTS

The descriptive statistics of RLs by participants across three stimulus conditions in each group are presented in Table 5.17.

Table 5. 17 Descriptive statistics of response latencies (in ms) by participants across stimulus conditions in each group: Non-English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Group 1	PE	20	1526.42	1224.37	1949.73	197.04	.57	-.47
	FT	20	1374.51	901.19	1798.80	244.35	-.04	-.65
	PC	20	970.28	749.78	1304.08	144.85	.67	-.33
Group 2	PE	20	1565.25	1054.10	2210.84	316.33	.024	-.35
	FT	20	1330.09	1009.24	1956.93	253.78	.79	.28
	PC	20	974.90	691.14	1348.37	161.21	.41	.02
Group 3	PE	20	1584.72	1166.40	2269.19	277.19	.52	.35
	FT	20	1370.84	924.37	1818.85	237.00	.14	-.52
	PC	20	917.94	817.52	1278.68	130.51	1.11	1.63

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-subjects factor (tested at 3 levels: PE, FT, and PC), and with the grouping as a between-subjects factor (tested at 3 levels: Group 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RLs by participants, F1 (1.76, 100.28) = 247.18, $p < .001$, with a large effect size ($\eta_p^2 = .82$), and there was no significant effect of grouping on the RLs by participants F1 (2, 57) = .03, $p = .97$, with a small effect size ($\eta_p^2 = .001$). The interaction effect between stimulus conditions and the group was not significant, F1 (3.52, 100.30) = .67, $p = .60$, with an effect size between small and medium ($\eta_p^2 = .023$). Post-hoc comparisons using the Bonferroni-correction were conducted. Note that Bonferroni correction serves to control for family-wise error rate (Field, 2013). The results suggested that PE was significantly slower than FT by participants (all *ps* < .001) in all three groups, and PC was significantly faster than FT and than PE by participants (all *ps* < .001) in all there

groups (see Table 5.18 for details). These effects are illustrated in Figure 5. 10.

Table 5. 18 Post-hoc pairwise comparisons between stimulus conditions in each group (n = 60): Response latencies (in ms) among non-English majors

		MD	SE	Sig.
Group 1	PE vs. FT	151.90***	35.14	.000
	PE vs. PC	556.14***	53.98	.000
	FT vs. PC	404.24***	49.79	.000
Group 2	PE vs. FT	235.16***	35.14	.000
	PE vs. PC	590.35***	53.96	.000
	FT vs. PC	355.19***	49.79	.000
Group 3	PE vs. FT	213.88***	35.14	.000
	PE vs. PC	666.77***	53.19	.000
	FT vs. PC	452.90***	49.79	.000

Notes: *** $p < .001$;
MD: meaning difference; SE: standard error.

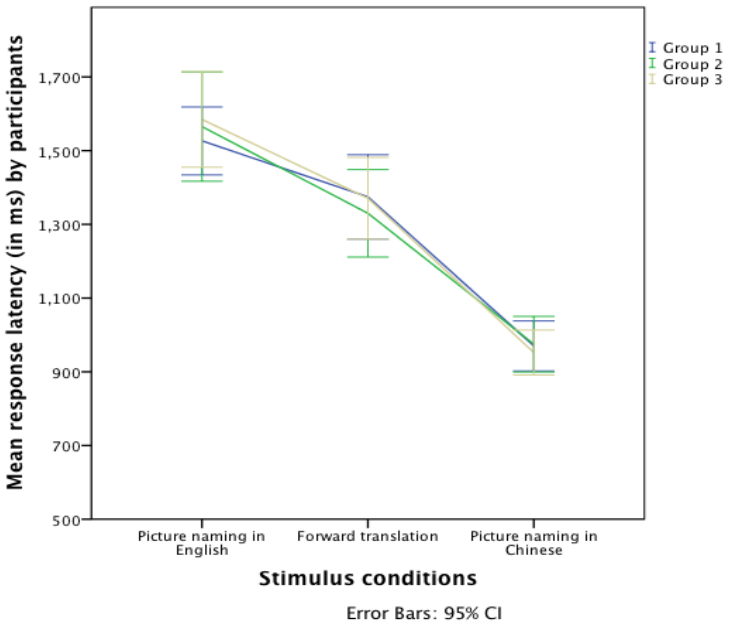


Figure 5. 10 Comparisons (by participants) across stimulus conditions and groups: Response latencies among non-English majors

BY ITEMS

The descriptive statistics of RLs by items across stimulus conditions in each block are presented in Table 5.19.

Table 5. 19 Descriptive statistics of response latencies (in ms) by items across stimulus conditions in each block: Non-English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Block 1	PE	66	1556.71	743.75	2420.45	411.12	.26	-.44
	FT	66	1343.76	753.38	2472.29	368.23	.78	.20
	PC	66	989.35	682.91	1515.07	190.56	.52	-.02
Block 2	PE	66	1607.59	709.90	2615.04	491.63	.16	-.10
	FT	66	1374.57	746.31	2089.69	345.54	.25	-.67
	PC	66	1039.54	558.63	1765.06	257.19	.64	.58
Block 3	PE	66	1566.17	731.66	2492.25	420.72	.11	-.56
	FT	66	1325.19	798.63	2124.57	315.75	.26	-.55
	PC	66	1017.03	636.59	1505.63	192.12	.30	-.23

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-items factor (tested at 3 levels: PE, FT, and PC), and with the block of stimulus words as a between-items factor (tested at 3 levels: Block 1, Block2, and Block 3). The ANOVA result suggested a significant effect of stimulus conditions on RLs by items, F2 (1.92, 373.89) = 343.89, $p < .001$, with a large effect size ($\eta_p^2 = .62$). There was no significant effect of stimulus blocks on RLs by items, F2 (2, 195) = .43, $p = .65$, with a small effect size ($\eta_p^2 = .004$), and no significant interactions between the condition and the stimulus block, F2 (3.84, 373.89) = .19, $p = .94$, with a very small effect size ($\eta_p^2 = .002$). Bonferroni-corrected post-hoc comparisons were

conducted, and the results suggested that PE was significantly slower than FT by items in all three blocks (all *ps* < .001), and PC was significantly faster than FT and than PE by items in all three blocks (all *ps* < .001) (see Table 5.20 for details). These effects are illustrated in Figure 5. 11.

Table 5. 20 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response latencies (in ms) by items among non-English majors

		MD	SE	Sig.
Block 1	PE vs. FT	212.95***	33.99	.000
	PE vs. PC	567.36***	42.50	.000
	FT vs. PC	354.41***	38.82	.000
Block 2	PE vs. FT	233.02***	33.99	.000
	PE vs. PC	568.06***	42.50	.000
	FT vs. PC	335.03***	38.82	.000
Block 3	PE vs. FT	240.98***	33.99	.000
	PE vs. PC	549.14***	42.50	.000
	FT vs. PC	308.16***	38.82	.000

Notes: *** $p < .001$;
MD: meaning difference; and SE: standard error.

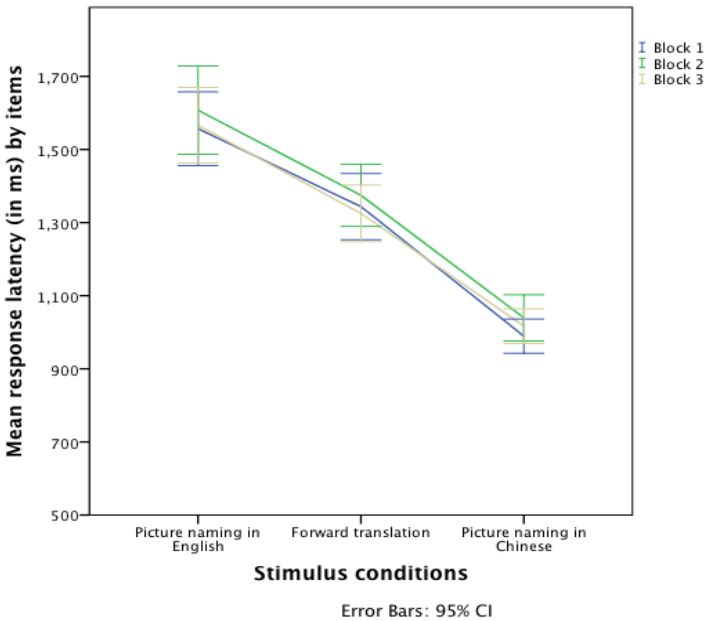


Figure 5. 11 Comparisons across stimulus conditions and blocks: Response latencies by items among non-English majors

5.4.2.2 Response accuracies: Non-English majors

BY PARTICIPANTS

The descriptive statistics of RAs across stimulus conditions in each group are presented in Table 5.21.

Table 5. 21 Descriptive statistics of response accuracies (in percentage) by participants across stimulus conditions in each group: Non-English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Group 1	PE	20	92.82	89.39	96.97	2.59	.24	-1.01
	FT	20	91.13	81.82	96.97	4.03	-.87	.31
	PC	20	97.42	93.94	100	1.85	-.06	-1.02
Group 2	PE	20	91.67	80.30	98.48	4.24	-.74	1.43
	FT	20	91.89	81.82	98.48	4.51	-.53	-.32
	PC	20	97.52	93.94	100	1.57	-.55	.06
Group 3	PE	20	91.82	78.79	98.48	5.33	-1.24	.86
	FT	20	92.75	87.88	96.97	2.17	-.19	.37
	PC	20	98.32	93.94	100	1.68	-1.06	.97

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-subjects factor (tested at 3 levels: PE, FT, and PC), and with the grouping as a between-subjects factor (tested at 3 levels: Group 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RAs by participants, F1 (2, 114) = 79.93, $p < .001$, with a large effect size partial ($\eta_p^2 = .58$). There was no significant effect of the variable grouping on RAs by participants, F1 (2, 57) = .35, $p = .71$, with a small effect size ($\eta_p^2 = .01$), and no significant interactions observed between the stimulus condition and the group, F1 (4, 114) = 1.18, $p = .32$, with an effect size between small and medium ($\eta_p^2 = .04$). Bonferroni-corrected post-hoc

comparisons were conducted, which suggested that the participants made more errors in PE and FT as compared to PC in all three groups (all *ps* < .001). However, there were no significant differences found between the accurate rate of PE and that of FT by participants in all three groups (all *ps* > .05) (see Table 5.22 for details). These effects are illustrated in Figure 5. 12.

Table 5. 22 Pairwise comparisons between stimulus conditions in each group (n = 60): Response accuracies (in percentage) among non-English majors

		MD	SE	Sig.
Group 1	PE vs. FT	1.69	.95	.240
	PE vs. PC	-4.60***	.96	.000
	FT vs. PC	-6.29***	.81	.000
Group 2	PE vs. FT	-.23	.95	1.00
	PE vs. PC	-5.86***	.96	.000
	FT vs. PC	-5.63***	.81	.000
Group 3	PE vs. FT	-.94	.95	.980
	PE vs. PC	-6.50***	.96	.000
	FT vs. PC	-5.56***	.81	.000

Notes: *** $p < .001$;
MD: meaning difference; and SE: standard error.

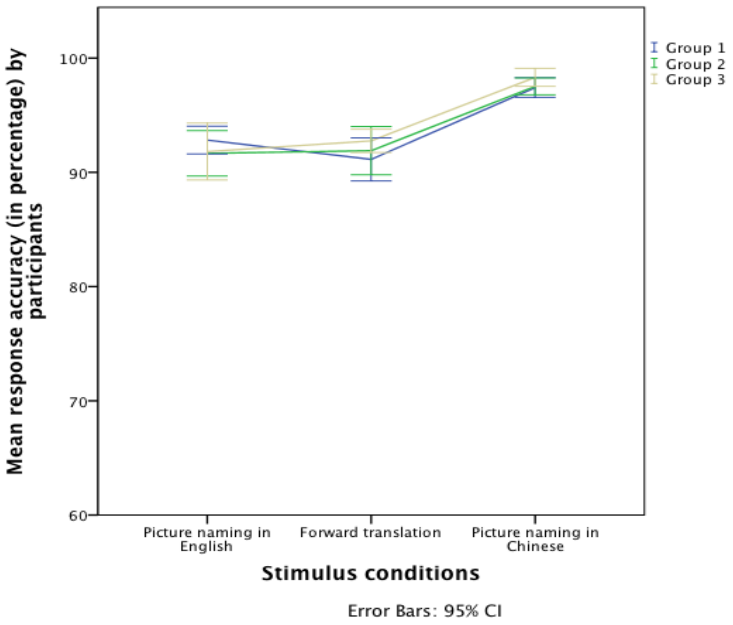


Figure 5. 12 Comparisons (by participants) across stimulus conditions and groups: Response accuracies (in percentage) among non-English majors

BY ITEMS

The descriptive statistics of RAs across stimulus conditions in each block are presented in Table 5.23.

Table 5. 23 Descriptive statistics of response accuracies (in percentage) by items across stimulus conditions in each block: Non-English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Block 1	PE	66	93.60	68.42	100	8.73	-1.56	1.79
	FT	66	94.24	75.00	100	6.86	-.1.22	.72
	PC	66	97.38	85.71	100	3.99	-.1.65	2.22
Block 2	PE	66	92.32	71.43	100	8.46	-1.21	.60
	FT	66	92.94	68.42	100	8.54	-1.46	1.54
	PC	66	97.35	85.00	100	4.13	-1.59	1.88
Block 3	PE	66	92.80	70.00	100	7.12	-.89	.58
	FT	66	91.63	61.90	100	9.38	-1.20	.93
	PC	66	98.52	92.00	100	2.56	-1.38	.34

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-items factor (tested at 3 levels: PE, FT, and PC), and with the block as a between-items factor (tested at 3 levels: Block 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RAs by items, F2 (2, 390) = 46.52, $p < .001$, with a large effect size ($\eta_p^2 = .19$). There was no significant effect of the block on RAs by items, F2 (2, 195) = .54, $p = .58$, with a very small effect size ($\eta_p^2 = .006$), and no significant interactions between the stimulus condition and the stimulus block, F2 (4, 390) = .13, $p = .11$, with an effect size between small and medium ($\eta_p^2 = .02$). Bonferroni-corrected post-hoc comparisons suggested that more errors were made in PE and FT as compared to PC by items in all three blocks (all *ps* < .01). However, there were no significant differences found between the accurate rate of PE and that of FT by items in all three blocks (all *ps* > .05) (see Table 5.24 for details). These effects are illustrated in Figure 5. 13.

Table 5. 24 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response accuracies by items among non-English majors

		MD	SE	Sig.
Block 1	PE vs. FT	-.64	.98	1.00
	PE vs. PC	-3.78***	.99	.001
	FT vs. PC	-3.14**	1.04	.008
Block 2	PE vs. FT	-.61	.98	1.00
	PE vs. PC	-5.03***	.99	.000
	FT vs. PC	-4.42**	1.04	.000
Block 3	PE vs. FT	1.17	.98	.690
	PE vs. PC	-5.71***	.99	.000
	FT vs. PC	-6.88***	1.04	.000

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$;
MD: meaning difference; SE: standard error.

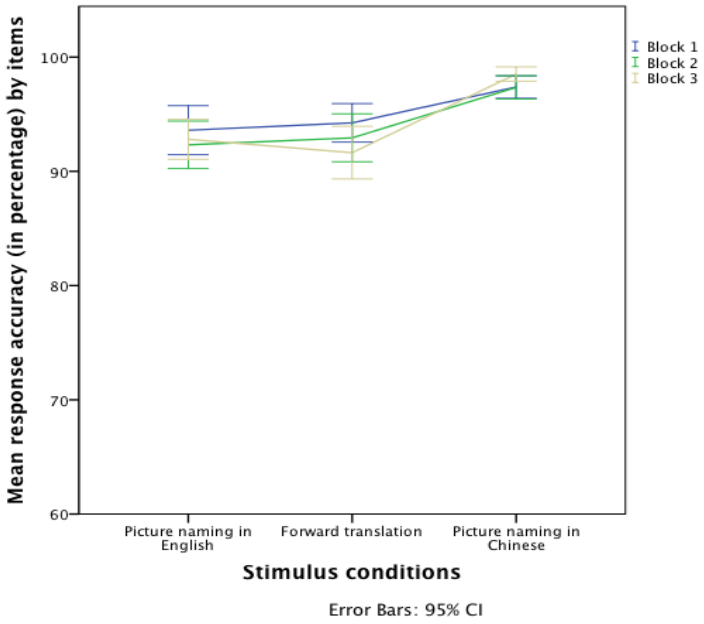


Figure 5. 13 Comparisons across stimulus conditions and blocks: Response accuracies by items among non-English majors

5.4.2.3. Interim summary: General pattern of non-English majors' L2 lexical processing

There were significant differences observed between PE response latencies and FT response latencies, both by participants and by items, among non-English majors, the results of which served to support the general pattern of L2 lexical processing characterized by the presence of L1-L2 lexical associations.

5.4.3 General pattern of L2 lexical processing among English majors

As in the above analysis, mixed-model F1 (by participants) and F2 (by items) ANOVAs were conducted with English-majors (n = 34) so as to identify the general pattern of L2 lexical processing among learners with L2 exposure more than 20 hours per week, as well as to examine whether or not there existed any significant

differences in the general pattern of L2 lexical processing between English majors and non-English majors. Statistical assumptions were checked in the same way as described above, and corrections were made, wherever appropriate, if the assumption violated.

5.4.3.1 Response latencies: English majors

BY PARTICIPANTS

Table 5.25 displays the descriptive statistics of RLs by participants across three stimulus conditions in each group.

Table 5. 25 Descriptive statistics of response latencies (in ms) by participants across stimulus conditions: English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Group 1	PE	12	1548.08	1245.54	1954.05	220.43	.70	.27
	FT	11	1300.08	986.83	1557.70	177.83	.001	-.87
	PC	11	1072.04	882.97	1277.34	139.25	.47	-1.39
Group 2	PE	12	1518.51	1149.89	1796.00	239.99	-.28	-1.50
	FT	11	1278.65	1020.20	1816.20	256.02	1.02	.32
	PC	11	1000.80	857.79	1210.28	122.44	.18	-1.26
Group 3	PE	12	1514.31	1034.58	1820.18	275.79	-.32	-1.26
	FT	11	1247.27	972.32	1532.21	169.60	-.02	-.39
	PC	11	993.96	825.82	1297.65	136.09	1.10	1.30

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was performed, with the stimulus condition as a within-subjects factor (tested at 3 levels: PE, FT, and PC), and with the grouping as a between-subjects factor (tested at 3 levels: Group 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RLs by participants, F1 (2, 62)

= 133.93, $p < .001$, with a large effect size ($\eta_p^2 = .81$). There was no significant effect of the group on RLs by participants, F1 (2, 31) = .32, $p = .73$, with an effect size between small and medium ($\eta_p^2 = .02$). The interaction effect between the stimulus condition and the group was not significant, F1 (4, 62) = .17, $p = .95$, with a very small effect size ($\eta_p^2 = .001$). Post-hoc comparisons using Bonferroni-correction were conducted, and the results suggested that PE response latencies were significantly longer than FT response latencies by participants in all three groups (all *ps* < .001). Moreover, PC was significantly faster than FT and than PE in all three groups (all *ps* < .001) (see Table 5.26 for details). These effects are illustrated in Figure 5.14.

Table 5. 26 Post-hoc pairwise comparisons between stimulus conditions in each group (n = 34): Response latencies (in ms) among English majors

		MD	SE	Sig.
Group 1	PE vs. FT	247.99***	44.47	.000
	PE vs. PC	476.03***	55.29	.000
	FT vs. PC	228.04***	55.09	.001
Group 2	PE vs. FT	239.86***	46.45	.000
	PE vs. PC	517.71***	57.74	.000
	FT vs. PC	277.85***	57.54	.000
Group 3	PE vs. FT	267.05***	46.45	.000
	PE vs. PC	520.35***	57.74	.000
	FT vs. PC	253.31***	57.54	.000

Notes: *** $p < .001$;
MD: meaning difference; and SE: standard error.

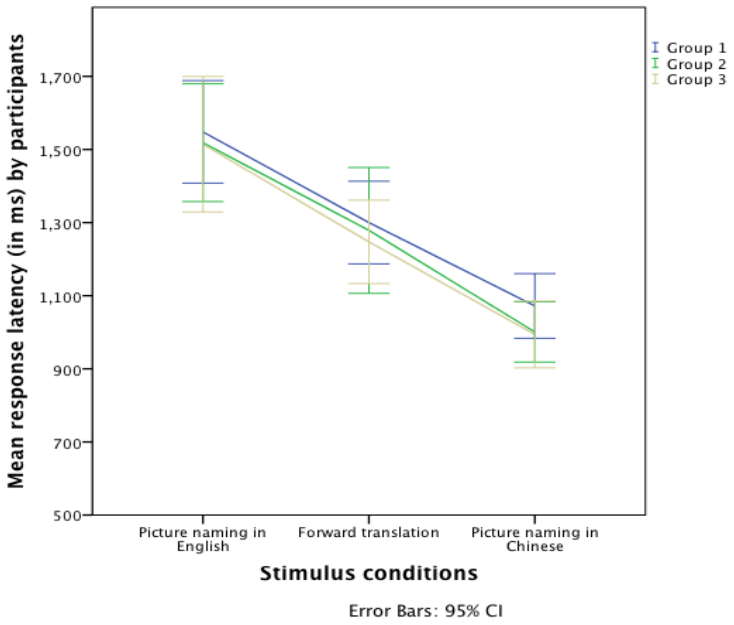


Figure 5. 14 Comparisons (by participants) across stimulus conditions and groups: Response latencies among English majors

BY ITEMS

Table 5.27 shows the descriptive statistics of RLs by items across stimulus conditions in each block.

Table 5. 27 Descriptive statistics of response latencies (in ms) by items across stimulus conditions in each block: English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Block 1	PE	66	1530.48	785.45	2850.56	442.95	.62	.29
	FT	66	1301.31	798.33	2328.29	334.56	.82	.40
	PC	66	998.36	685.00	1602.00	204.58	.90	.84
Block 2	PE	66	1534.72	738.55	2445.67	425.05	.21	-.67
	FT	66	1323.72	717.73	2553.86	419.99	.81	.39
	PC	66	1027.18	568.50	1627.82	264.74	.49	-.20
Block 3	PE	66	1568.63	713.58	2589.00	462.61	.05	-.73
	FT	66	1255.14	792.64	2024.11	278.1	.51	.30
	PC	66	1069.17	625.27	1825.67	259.51	.95	1.08

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was conducted, with the stimulus condition as a

within-items factor (tested at 3 levels: PE, FT, and PC), and with the block of stimulus words as a between-items factor (tested at 3 levels: Block 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RLs by items, F2 (2, 390)= 210.84, p < .001, with a large effect size (η_p^2 = .52). There was no significant effect of the block on RLs by items, F2 (2, 195) = .10, p = .90, with a very small effect size (η_p^2 = .001), and no significant interactions between the stimulus condition and the stimulus block, F2 (4, 390)= 1.35, p = .25, with a small effect size (η_p^2 = .01). Bonferroni-corrected post-hoc comparisons were conducted, and the results suggested that PE was significantly slower than FT by items in all three blocks (all *ps* < .001), and PC was significantly faster than FT and than PE by items in all three blocks (all *ps* < .001) (see Table 5.28 for details). These effects are illustrated in Figure 5. 15.

Table 5. 28 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response latencies (in ms) by items among English majors

		MD	SE	Sig.
Block 1	PE vs. FT	229.17***	41.53	.000
	PE vs. PC	532.12***	45.85	.000
	FT vs. PC	302.95***	42.34	.000
Block 2	PE vs. FT	210.10***	41.53	.000
	PE vs. PC	507.54***	45.85	.000
	FT vs. PC	296.54***	42.34	.000
Block 3	PE vs. FT	313.49***	41.53	.000
	PE vs. PC	499.46***	45.85	.000
	FT vs. PC	185.97***	42.34	.000

Notes: *** p < .001;
MD: meaning difference; and SE: standard error.

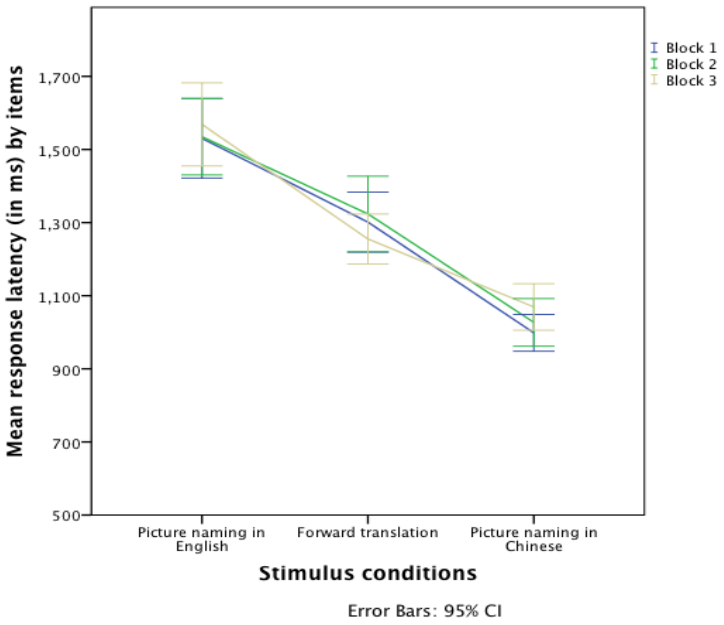


Figure 5. 15 Comparisons across stimulus conditions and blocks: Response latencies by items among English majors

5.4.3.2 Response accuracies: English majors

BY PARTICIPANTS

Table 5.29 displays the descriptive statistics of RAs across stimulus conditions in each group.

Table 5. 29 Descriptive statistics of response accuracies (in percentage) by participants across stimulus conditions in each group: English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Group 1	PE	12	94.03	89.39	98.48	2.80	-.17	-1.04
	FT	11	91.79	83.33	96.97	3.90	-.76	.33
	PC	11	96.76	92.42	100.00	2.20	-.87	.98
Group 2	PE	12	92.84	86.36	98.48	3.90	-.24	-.90
	FT	11	93.25	86.36	98.48	4.85	-.12	-.1.81
	PC	11	97.93	95.45	100.00	1.23	-.54	.66
Group 3	PE	12	93.53	87.88	96.97	2.96	-.42	-.42
	FT	11	93.39	84.85	98.48	4.08	-1.07	.59
	PC	11	98.76	96.97	100.00	.91	-.02	.39

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was performed, with the stimulus condition as a within-subjects factor (tested at 3 levels: PE, FT, and PC), and with the grouping as a between-subjects factor (tested at 3 levels: Group 1, 2, and 3). The ANOVA result suggested that there was a significant effect of stimulus conditions on RAs by participants, F1 (2, 62) = 34.33, $p < .001$, with a large effect size ($\eta_p^2 = .53$). There was no significant effect of the group on RLs by items, F1 (2, 31) = .55, $p = .58$, with an effect size between small and medium ($\eta_p^2 = .03$), and no significant interactions between the stimulus condition and the group, F1 (4, 62) = 1.07, $p = .38$, with a medium effect size ($\eta_p^2 = .07$). Bonferroni-corrected post-hoc comparisons were conducted, and the results suggested that more errors were made by participants in PE and FT relative to those in PC in all three groups (all *ps* < .05). However, there were no significant differences in RAs found between PE and FT in all three groups (all *ps* > .05) (see Table 5.30 for details). These effects are illustrated in Figure 5.16.

Table 5. 30 Pairwise comparisons between stimulus conditions in each group (n = 34): Response accuracies (in percentage) among English majors

		MD	SE	Sig.
Group 1	PE vs. FT	2.24	1.06	.130
	PE vs. PC	-2.73*	1.00	.030
	FT vs. PC	-4.99***	1.25	.000
Group 2	PE vs. FT	-.41	1.10	1.000
	PE vs. PC	-5.10***	1.04	.000
	FT vs. PC	-4.68***	1.30	.000
Group 3	PE vs. FT	.14	1.10	1.000
	PE vs. PC	-5.23***	1.04	.000
	FT vs. PC	-5.37**	1.30	.000

Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$;
MD: meaning difference; and SE: standard error.

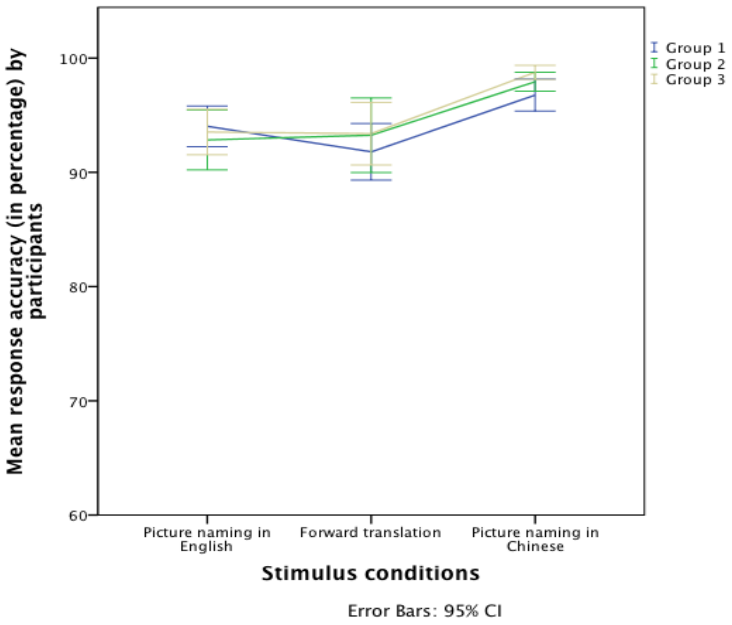


Figure 5. 16 Comparisons (by participants) across stimulus conditions and groups: Response accuracies (in percentage) among English majors

BY ITEMS

Table 5.31 shows the descriptive statistics of RAs across stimulus conditions in each block.

Table 5. 31 Descriptive statistics of response accuracies (in percentage) by items across stimulus conditions in each block: English majors

Condition		n	Mean	Minimum	Maximum	SD	Skewness	Kurtosis
Block 1	PE	66	95.27	72.73	100	7.65	-1.45	.98
	FT	66	93.98	75.00	100	8.83	-.1.24	.25
	PC	66	98.71	93.00	100	1.80	-.1.12	.09
Block 2	PE	66	93.99	72.73	100	8.92	-1.23	.10
	FT	66	93.44	63.64	100	8.80	-1.33	1.28
	PC	66	97.48	83.33	100	4.62	-1.21	.89
Block 3	PE	66	93.98	75.00	100	8.07	-1.28	.82
	FT	66	94.42	72.73	100	7.24	-1.68	1.98
	PC	66	98.49	90.00	100	3.30	-1.92	1.84

Notes: FT: forward translation; PC: picture naming in Chinese; and PE: picture naming in English.

A mixed-model ANOVA was conducted, with the stimulus condition as a within-items factor (tested at three levels: PE, FT, and PC), and with the block of stimulus words as a between-items factor (tested at three levels: Block 1, 2, and 3). The ANOVA result suggested a significant effect of stimulus conditions on RAs by items, F2 (2, 390)= 25.14, $p < .001$, with an effect size between medium and large ($\eta_p^2 = .11$). There was no significant effect of the block on RAs by items, F2 (2, 195) = .86, $p = .43$, with a small effect size ($\eta_p^2 = .01$), and no significant interactions between the stimulus condition and the stimulus block, F2 (4, 390)= 1.35, $p = .32$, with a very small effect size ($\eta_p^2 = .003$). Bonferroni-corrected post-hoc comparisons suggested that more errors were made in PE and FT as compared to PC by items in all three blocks (all *ps* < .01). However, there were no significant differences found in between the accurate rate of PE and that of FT by items in all three blocks (all *ps* > .05) (see Table 5.32 for details). These effects are illustrated in Figure 5. 17.

Table 5. 32 Post-hoc pairwise comparisons between stimulus conditions in each block (n = 198): Response accuracies by items among English majors

		MD	SE	Sig.
Block 1	PE vs. FT	1.29	1.21	.860
	PE vs. PC	-3.43**	1.09	.006
	FT vs. PC	-4.73***	1.14	.000
Block 2	PE vs. FT	.55	1.21	1.000
	PE vs. PC	-3.49**	1.09	.005
	FT vs. PC	-4.04**	1.14	.001
Block 3	PE vs. FT	-.44	1.21	1.000
	PE vs. PC	-4.51***	1.09	.000
	FT vs. PC	-4.07***	1.14	.001

Notes: * $p < .05$, ** $p < .01$, and *** $p < .001$;
MD: meaning difference; and SE: standard error.

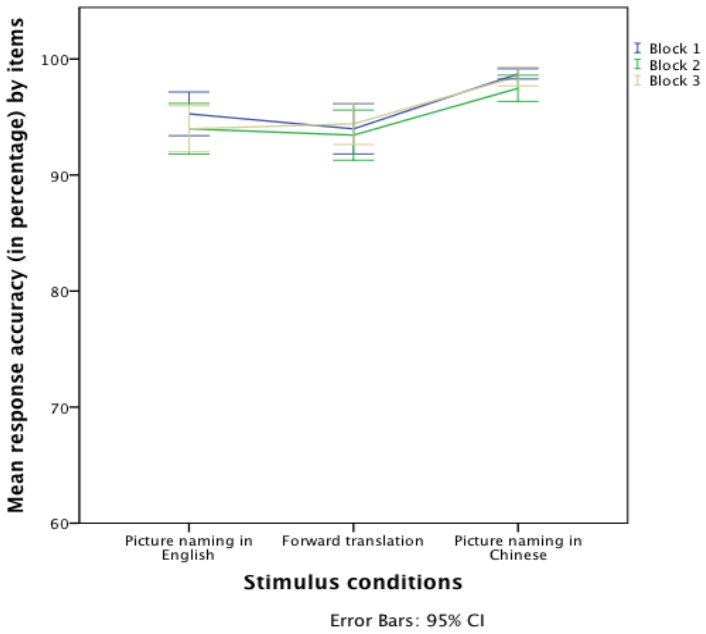


Figure 5. 17 Comparisons across stimulus conditions and blocks: Response accuracies by items among English majors

5.4.3.3 Interim summary: General pattern of English majors' L2 lexical processing

The above analyses showed that there were statistically significant RL differences between PE and FT among English majors, indicating that the mechanisms underlying their L2 lexical processing are lexically mediated. The general processing pattern observed with English majors is similar to that observed with non-English majors. Therefore, it can be argued that lexical routes are likely to underpin L2 lexical processing among adult L2 learners regardless of the amount of target language exposure that they have received in non-native English speaking instructional settings.

5.5 Tendency toward a certain pattern of L2 lexical processing among adult L2 learners

This analysis answers the following research questions:

Do learner-related variables predict the tendency toward a certain pattern of L2 lexical processing among adult L2 learners?

Do word feature-related variables predict the tendency toward a certain pattern of L2 lexical processing among adult L2 learners?

The logic behind the following analysis has been specified in Section 4.5. Briefly speaking, the larger the differences between PE response latencies and FT response latencies are, the more likely it is that the word association model is supported.

5.5.1 Learner-related variables and processing tendency

Before performing formal analyses, descriptive statistics for all predictor variables and the outcome variable were examined, and a correlation matrix among continuous predictors, and between each predictor variable and the outcome variable was estimated (see Table 5.33 below). The strength of association between nominal predictors and the outcome variable was estimated using Lambda, and the results are reported in Appendix 5.10. The correlation results suggested that only one predictor variable (i.e. academic achievement) was significantly correlated with PE-FT response latency differences. This variable was included in the regression model to examine the extent to which it predicted the PE-FT response latency difference. Moreover, the assumptions of running MR were checked, including the sample size,

missing data, normality, linearity, and homoscedasticity of residuals, outliers, and multicollinearity and singularity (see Appendix 5.11 for details), with no serious violation noted.

Table 5. 33 Correlation matrices: Relationships between learner-related variables, and PE-FT response latency differences among adult L2 learners (n = 94)

Pearson's correlations

Variables	1	2	3	4	5	6	7	8
1. PE-FT differences	1	—	—	—	—	—	—	—
2. Age	.03	1	—	—	—	—	—	—
3. Vocabulary depth	.01	-.17	1	—	—	—	—	—
4. Vocabulary size	-.03	.02	.31**	1	—	—	—	—
5.English proficiency	.04	.12	.21*	.14	1	—	—	—
6. Academic achievement	-.24*	.07	.06	-.02	.40**	1	—	—
7. Age of starting L2 learning	.01	.15	.09	.03	.38**	.18	1	—
8. Amount of L2 exposure	.13	.00	.09	-.02	.19	-.40**	.12	1

Notes: * $p < .05$ (two-tailed); ** $p < .01$ (two-tailed)

The MR result suggested that academic achievement might significantly predict the PE-FT response latency difference (see Table 5.34), the R (.24) was significantly different from zero, F (1, 92) = 5.62, $p < .05$, with R^2 at .06 (an effect size between small and medium). The adjusted R^2 value of .05 indicated that only 5% of the variability in the PE-FT response latency difference was predicted by academic achievement scores. The size and direction of the relationship suggested that those who had better academic achievements were more likely to process L2 words through conceptual mediation.

Table 5. 34 Standard multiple regression of academic achievement on the PE-FT response latency difference (n = 94)

Variables	PE-FT differences	Academic achievement	*B*	*β*	sr^2 (unique)
Academic achievement	- .240*		-1.27*	-.24	.06
			Intercept = 981.85		
Means	218.87	599.30			
Standard deviations	156.55	29.51			
				R^2= .06	
				Adjusted R^2= .05	
				R= .24*	

Notes: * $p < .05$;
B: unstandardized regression coefficients;
β: the standardized regression coefficients;
$sr^{2:}$ the semi-partial correlations.

5.5.2 Word feature-related variables and processing tendency

As in the above analysis, descriptive statistics for all the predictor variables and the outcome variable were examined, and a correlation matrix among continuous predictors, and between each predictor and the outcome variable was estimated (see Table 5.35). Clearly, three predictor variables (i.e. WT, Log-SND, and AoA) were significantly correlated with PE-FT response latency differences. These lexical variables were included in the regression model to examine the extent to which they predicted the PE-FT response latency differences. Moreover, the assumptions of MR were checked (see Appendix 5.11 for details), with no serious violation found.

Table 5. 35 Correlation matrix: Correlations between word feature-related variables, and PE-FT response latency differences among adult L2 learners (n = 198)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. PE-FT differences	1	—	—	—	—	—	—	—	—	—
2. WC	.03	1	—	—	—	—	—	—	—	—
3. WT	-.15*	-.10	1	—	—	—	—	—	—	—
4. Log-SND	-.21**	-.38**	.30**	1	—	—	—	—	—	—
5. Log-NoRS	-.03	-.27**	.07	.21**	1	—	—	—	—	—
6. Log-NoL	.12	.06	-.28**	-.30**	-.40**	1	—	—	—	—
7. Log-NoP	.10	.11	-.20**	-.30**	-.36**	.78**	1	—	—	—
8. Log-NoS	.01	.15*	-.13	-.23**	-.32**	.64**	.71**	1	—	—
9. SWF	-.12	-.26**	.56**	.41**	.28**	-.24**	-.26**	-.16*	1	—
10. AoA	.23**	.16*	-.55**	-.36**	.02	.13	.09	.07	-.53**	1

Notes: * $p < .05$ (two-tailed); ** $p < .01$ (two-tailed)

In the regression analysis of word feature-related variables that might predict PE-FT response latency differences (see Table 5.36), the R (.27) was significantly different from zero, F (3, 194) = 4.94, $p < .01$, with R^2 at .07 (an effect size between small and medium). The adjusted R^2 value of .06 indicated that 6% of the variability in PE-FT response latency differences was significantly predicted by the following two lexical variables: SND and AoA. The size and direction of the relationships suggested that those words that had higher semantic neighborhood density ($\beta = -.14$, $p < .05$) and were learned earlier ($\beta = .18$, $p < .05$) were likely to be processed through conceptual mediation. Between these two significant predictors, it was found that AoA was slightly more important, as indicated by the squared semi-partial correlations.

5.5.3 Interim summary: Factors predicting processing tendency

Multiple regression analysis on the predictive effects of a learner variable and three lexical variables on response latency differences between English picture naming and forward translation suggest that those participants that have lower academic achievements and those words that have lower semantic neighborhood density and are learned later have a tendency to support the word association model. Academic achievements, as an indicator of intellectual competence, were found to play an important role in influencing the way in which L2 words are processed. Additionally, the semantic network that has been fully developed by adults likely influences their L2 lexical processing mechanisms. Late-acquired words are more likely to be

lexically processed than early-acquired words.

Table 5. 36 Standard multiple regression of word feature-related variables on the PE-FT response latency difference among adult L2 learners (n = 198)

Variables	PE-FT	WT	Log-SND	AoA	*B*	*β*	sr^2 (unique)
Word typicality	-.15*				-3.31	-.01	
Log-semantic neighborhood density	-.21**	.30**			-73.17*	-.14	.02
Age of acquisition	.23**	-.55**	-.36**		18.58*	.18	.03
					Intercept = 7.81		
Means	237.48	3.53	.64	11.88			
Standard deviations	275.74	.73	.56	2.80			
						R^2= .07	
						Adjusted R^2= .06	
						R= .27***	

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.
B : unstandardized regression coefficients;
β: the standardized regression coefficients;
$sr^{2:}$ the semi-partial correlations.

5.6 Summary: Chapter 5

Before summarizing this chapter, it is instructive to return to Figure 5.1 above, and how far those research questions have been answered are outlined in Figure 5.18 below. The main results of the present study indicate that the variance of naming and translation response latencies accounted for by lexical variables is far more than that explained by learner variables, meaning that what adult learners have experienced with each individual L2 word seems to be more significant than how proficient they are. Moreover, the presence of significant differences between picture naming and forward translation, both by participants and by items, gave support to the word association model. This result applied to both non-English majors and English majors, indicating that L2 words are lexically processed in adult L2 learners regardless of the amount of exposure they have received in input-limited contexts. Lastly, academic competence and a lexical item's semantic neighborhood density and age of acquisition best predicted the tendency toward word associations over and beyond other learner and lexical variables. Theoretical and practical implications of these findings will be presented in Chapter 6 and 7.

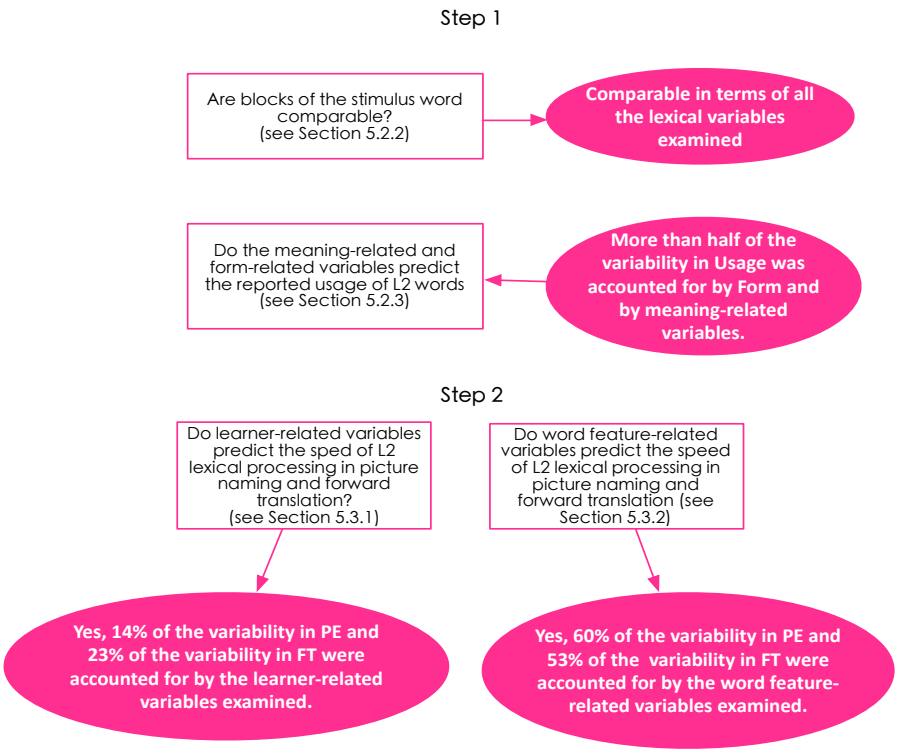


Figure 5. 18 Summary of Chapter 5: Research questions and results

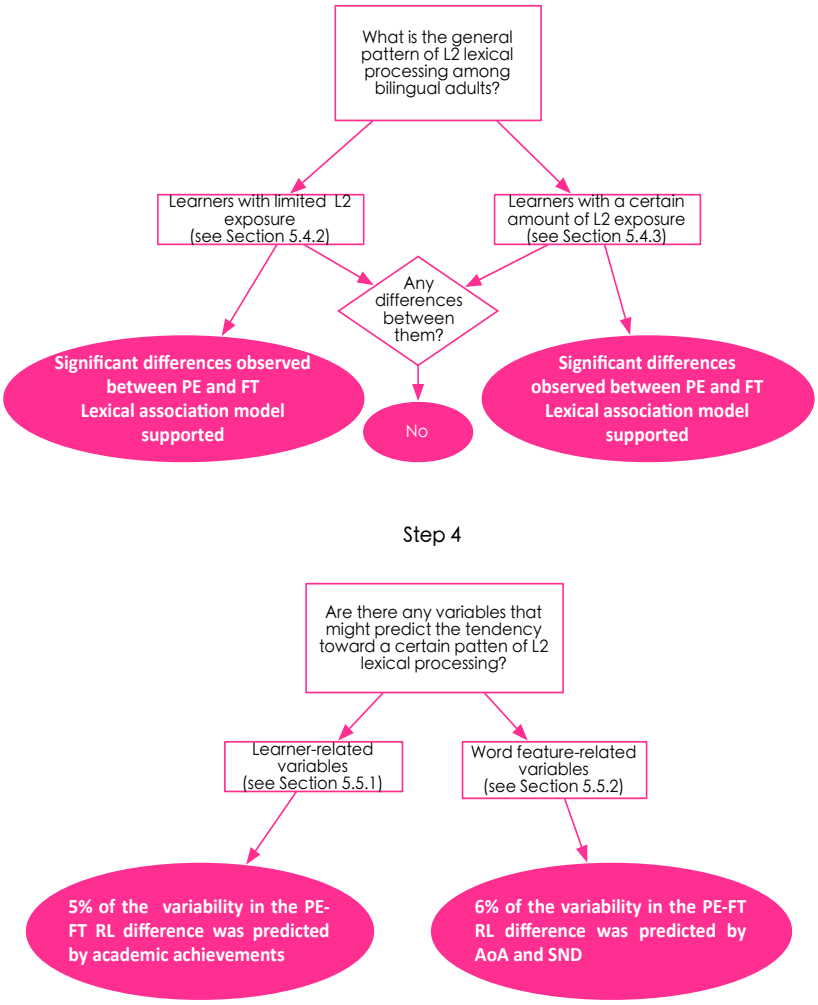


Figure 5.18 Summary of Chapter 5: Research questions and results (Continued)

Chapter 6 General Discussion

In this chapter, the main results of the present study are summarized, and then possible explanations to account for the observed learner and lexical effects are provided. In the following sections, several reasons are offered to explain why different patterns of bilingual lexical processing were observed with the child and adult participants, together with the proposal of a model specifying bilingual lexicons in development. The last section will present a schematic illustration of all possible determinants of bilingual lexical organization.

6.1 Summary of results

The main issues that the present study has addressed include: (1) the extent to which learner and lexical variables predict the efficiency of L2 lexical processing in production, (2) what mechanisms, in general, best characterize the structure of bilingual lexicons across age cohorts, and (3) what types of words are more likely to be lexically processed in what types of learners. Two groups of L2 learners (i.e. primary-age children and university-level adults) performed picture naming in English and then in Chinese, and then completed a Chinese-to-English translation task, and their response latencies and accuracies were recorded and analyzed. The main results obtained in this study can be summarized as follows:

(1) *Vocabulary size* and *L1 competence* significantly predicted the child participants' processing efficiency, whereas *vocabulary depth* and *the onset age of L2 learning* contributed significantly to the adult participants' processing efficiency.

(2) In the child group, the direct and total effects of *word typicality and age of*

acquisition on naming and translation latencies were found to be significant. By contrast, in the adult group, the direct effect of *Usage* on naming and translation latencies was found to be significant, and the total effects of *word typicality* and *Usage* on naming and translation latencies were found to be significant.

(3) In the child group, the *response-latency variance* explained by *learner variables* was not far less than that explained by *lexical variables*, whereas in the adult group, a much larger amount of the *variance* was explained by *lexical variables* relative to *learner variables*.

(4) The mechanisms underlying young children's L2 lexical processing were more likely to be characterized by *the concept mediation model*, whereas the mechanisms underlying adult learners' L2 lexical processing were more likely to be characterized by *the word association model*.

(5) Cross-linguistic routes likely underpin L2 lexical processing in adult L2 learners regardless of the amount of exposure that they had received in input-limited foreign language learning contexts.

(6) In the child group, *age of acquisition* significantly predicted the tendency toward a certain pattern of L2 lexical processing, whereas in the adult group, *age of acquisition* and *semantic neighborhood density* significantly predicted the tendency toward a certain pattern of L2 lexical processing.

To account for these results, possible explanations are provided below. How the following sections are structured is illustrated in Figure 6.1. It should be emphasized that any cross-age comparisons that were made below are located at the macro-level,

not the micro-level.

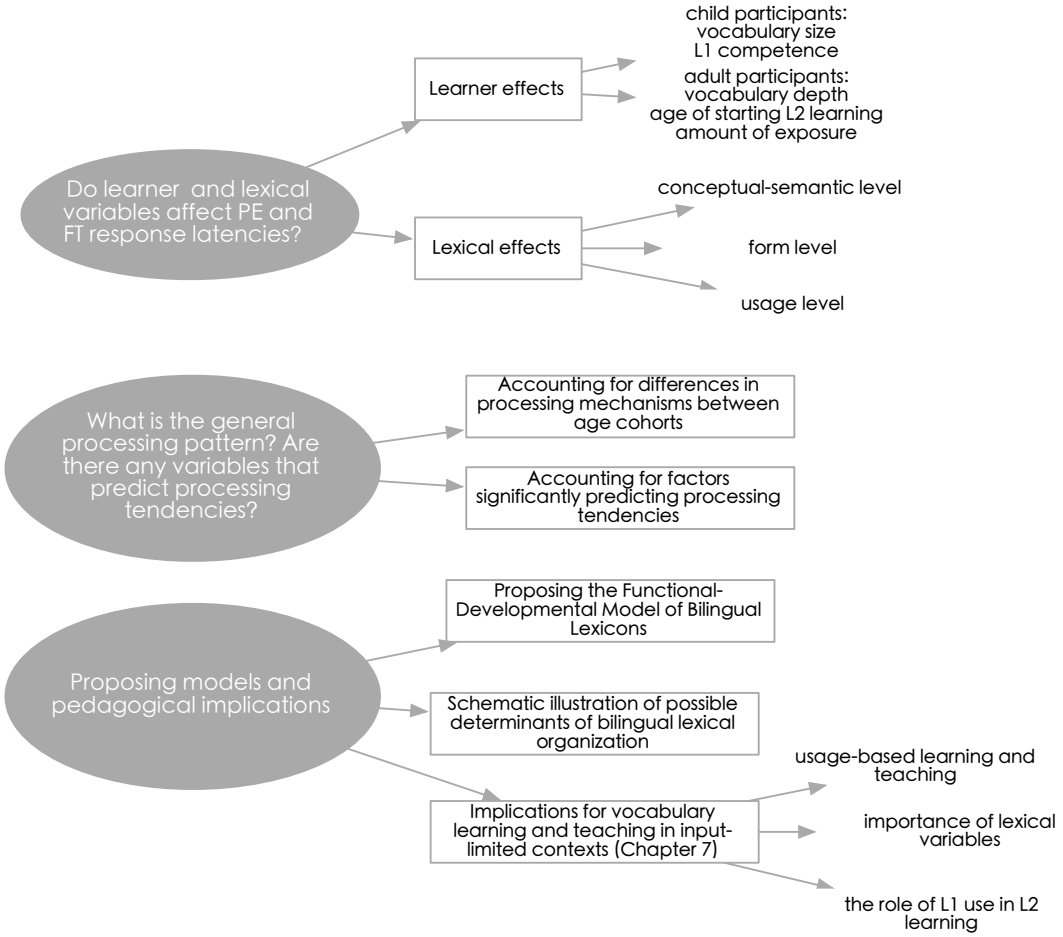


Figure 6. 1 Structure of the discussion chapter

6.2 Explaining the data: Learner and lexical characteristics

6.2.1 Learner effects

Vocabulary knowledge is not only about end products but also about the ability to access lexical items fluently. Vocabulary accessibility is considered important because it is not only what we know that matters in L2 learning but also what we can

do with our knowledge. As mentioned above, the fluency of L2 word production was assumed to be influenced by a series of learners' internal and external variables (Littlewood, 1984). Among all the variables examined in the present study, it has been shown that vocabulary size and L1 competence significantly predicted child L2 learners' processing efficiency, whereas vocabulary depth and the onset age of L2 learning significantly predicted adult L2 learners' processing speed. Only these significant predictors will be discussed in detail below.

Vocabulary knowledge grows in an incremental manner (Schmitt, 2000), thus its size varies with learners' previous experiences using and consolidating known words. The crucial role that vocabulary size played in the child participants' L2 production is compatible with a study by Edwards et al. (2004) where the size of lexical knowledge has been found to affect monolingual children's production accuracy and fluency in non-word repetition. Also, it should be pointed out that as children's vocabulary expands, they become better able to precisely distinguish between words at the phonemic level (Walley, 1993). It can thus be argued that child learners' phonological awareness may mediate some of the predictive effects that vocabulary size has on the fluency of lexical processing in production. Moreover, given that retrieving a learned word is declarative in nature (Ullman, 2004), then it stands to reason that the declarative memory system of child L2 learners, although not fully developed yet, may still play a crucial role in influencing the ease with which L2 words are accessed.

Additionally, as shown in item-specific analysis, the number of related senses had an indirect and significant effect on the child participants' processing speed via the

mediating variable of word usage. Given this result, learners' vocabulary depth should have a significant impact upon the recorded response latencies in by-participant analysis. However, this expected effect was not observed in the present study, possibly because this aspect of vocabulary knowledge is usually deemphasized by L2 teachers in the initial stage of vocabulary learning, and not surprisingly, the significant predictive effects of learners' vocabulary depth were not noted in the present study.

As regards the adult participants, vocabulary size was not a significant predictor of processing efficiency. There are a variety of possible reasons why the importance of vocabulary size appeared to be reduced as the participants' proficiency increased. First of all, this may be due to the nature of a nonlinear relationship between vocabulary size and oral production fluency in the sense that learners' productive performance may initially improve and then stabilize with their increasing vocabulary knowledge. It may also be a problem with the vocabulary assessments because the low-frequency words (e.g. at the 5000-word level) that were included in the adult participants' vocabulary tests might be irrelevant to the stimuli used in the reaction time-based experiments; however, this was not the case with the child participants where there was clearly more overlap between the assessment words and the stimulus words.

Together with vocabulary size, the child participants' L1 competence contributed significantly to predicting the speed of picture naming, a finding that can be at least explained by Cummins' *Interdependence Hypothesis* (1991) and Geva and Ryan's

(1993) *common Underlying Cognitive Processes Framework*. According to the former hypothesis, as introduced above, the cross-linguistic aspects of language proficiency are underpinned by a learner's common cognitive and personality attributes, and consequently the development of L1 literacy is likely to be in parallel with the acquisition of comparable skills in the L2. In addition, the present study demonstrated a significant positive correlation between the speed of L1 picture naming and that of L2 picture naming ($r = .29$, $p < .01$), a result that indicates the existence of a close relationship between an L2 speaker's L1 and L2 oral fluency. Within the latter framework (Geva & Ryan, 1993), the similarities in the basic operations of such cognitive components as working memory, short-term memory, executive control functions, and non-verbal intelligence within each individual are believed to explain L1-L2 close correlations in literacy skills and proficiency attainment. In a word, these underlying abilities are taken as a set of psychological givens that are hugely influential to the success and rate of language development whether in the L1 or in the L2.

It should be noted that the close relationship between L1 competence and L2 naming fluency could lend credence to the presence of lexical associations in the children's production of L2 words, but its existence does not deny the possibility that conceptual mediation might also be implicated in L2 lexical processing. Additionally, the failure to identify any significant effects of L1 competence on translation latencies may lie in the specific processing mechanisms and demands involved in performing forward translation.

As regards adult L2 learners, their depth of lexical knowledge and onset age of L2 learning were found to significantly predict naming and translation latencies. In the present study, a significant positive correlation was noted between the participants' vocabulary depth and breadth ($r = .31$, $p < .01$), indicating that the more L2 words an adult learner knows, the more likely it is that they achieve a greater depth of knowledge of these words. Even so, interestingly, it was vocabulary depth rather than vocabulary size that contributed significantly to predicting the recorded response latencies, an observation that could be accounted for by one of the following reasons.

First, subcomponents of lexical knowledge such as a word's syntax, morphology, word derivation, and association (Nation, 1990) are believed to assist in accessing a target lexical item. As such, those adult learners who have deeper vocabularies are better able to capitalize on these linguistic resources and to produce L2 words more rapidly than do those with shallower vocabularies. Second, given that spoken production goes through the stages of conceptualization, formulation, and articulation (Levelt, 1989), the word retrieval process involved in naming and translation may hinge heavily on the depth of lexical knowledge and the internal structure of semantic organization, which have been relatively well established in adults relative to children. As such, not surprisingly, vocabulary depth served to explain more variance in naming and translation latencies relative to that explained by vocabulary size in the present study.

As regards the onset age of L2 learning, its significant predictive ability can be explained by the critical period hypothesis, which posits that one language or more

introduced after an optimal age is unlikely to be acquired and processed in the same way when introduced in early childhood (e.g. Birdsong, 2009; Newport, 1988, 1990). An alternative explanation concerns learners' native languages because the more established the L1 is at the outset of L2 learning, the more likely that it interferes with the fluency of L2 production (Guion, Flege, Liu, & Yeni-Komshian, 2000). Additionally, such an L1 use effect reflects a constant interaction between the L1 and the L2 systems; therefore, neither of the languages can be fully deactivated in producing either the L1 or the L2 words. This explanation is also consistent with the observation of the presence of lexical associations even among relatively fluent L2 learners in the current study. The last explanation pertains to the linear relationship between the order in which words are learned and the speed of L2 lexical processing. This is to say, the order of acquisition influences how much experience a learner has with individual words, an influence which can extend to determine the ease or difficulty with which they access L2 words at a later stage.

In addition to vocabulary depth and the onset age of L2 learning, the amount of exposure significantly predicted translation latencies but not picture naming latencies, a finding that reflects the important role that practice plays in moderating the fluency of L2 lexical processing in forward translation. Another possible explanation relates to the common use of translation equivalents as an instructional method in input-limited foreign language contexts. This also explains why lexical associations were activated when relatively fluent adult participants produced L2 words in the current study.

This section has discussed why several learner variables contributed significantly to predicting the speed of L2 lexical processing, and in what follows, the observed lexical effects will be discussed.

6.2.2 Lexical effects

This study is the first to provide an in-depth analysis of the complex relationship between a range of lexical characteristics and the fluency of spoken word production across age cohorts among foreign language learners. In the following sections, the observed lexical effects will be explained in relation to previous research on lexical access in spoken production.

6.2.2.1 Conceptual-semantic level

WORD TYPICALITY. It is widely held that mental lexicons are characterized by conceptually taxonomic relationships, with prototypical items having more properties of its category than those that are weakly included in a category (Hampton, 1993). Whether or not the conceptual structure serves as a significant predictor of L2 lexical processing has been examined in the present study.

In the child participants, word typicality contributed significantly to the speed of picture naming and translation in a direct manner, a result that is compatible with previous studies in which participants have displayed a preference for typical items over atypical items (e.g. Bjorklund & Thompson, 1983; Jerger & Damian, 2005). This result also confirms the fact that categorization is one of the cognitive abilities that children are able to develop, even as young as three or four months (Eimas & Quinn, 1994).

In the adult participants, word typicality significantly and indirectly affected the response latencies via L2 word usage, in contrast to what was observed with the child participants. The presence of this indirect effect signals the importance of usage in mediating how typicality exerts its influence and could account for why mixed results have been obtained in the literature (e.g. Larochelle & Pineau, 1994; Mäntyla, 1993). One possible explanation to account for such an age-related difference concerns the conceptual structure in development, a process that is gradually conditioned by personal experiences with individual words. As rightly pointed out by Murphy (2004, p. 1), concepts are 'mental glues' that hold together our past experiences and our present interactions with the world. This may be one of the reasons why the distribution of adult-rated typicality was 0.26 standard deviations wider than that of child-rated typicality even when the difference between their mean ratings was only 0.04.

Taken together, in both age groups, the conceptual structure underlying the L2 mental lexicon plays an important role in predicting the efficiency of lexical processing, indicating that many L2 words are accessible to the conceptual memory system regardless of a learner's age. Also, this observed significant effect confirmed the validity of using subjectively estimated typicality in explaining the variance of the participants' performance on spoken production. Moreover, the likelihood that typicality effects are regulated by word usage increases with learners' age. What this suggests is that the conceptual nodes linked to L2 words may be more sensitive to the variable of usage in adult learners with higher proficiency levels relative to in child

learners with very limited proficiency levels.

SEMANTIC NEIGHBORHOOD DENSITY AND NUMBER OF RELATED SENSES. Models of spoken word production describe the process by which meaning and form are activated in production. To learn more about this process, semantic characteristics of target words were varied so as to identify the extent to which these variables affected the ease or difficulty with which a name was produced by the participants.

The results showed that semantic neighborhood density served as a significant and indirect predictor of naming and translation speed in both age groups. Its absence as a direct effect on lexical processing may explain why significant semantic neighborhood effects have not been observed in previous studies (e.g. Bormann, 2011; Storkel & Adlof, 2009). Moreover, word usage was found to mediate the relationship between neighborhood density and processing speed, indicating that semantic networks develop and enlarge as learners' vocabulary usage increases. The inseparable relationship between word usage and neighborhood density may also be due to the fact that the participating teachers had presented semantic associates of the targeted vocabulary words in L2 instructed contexts as a means of facilitating the participants' intake of new vocabulary at an advanced stage of language learning.

In comparison, the SND effect is more notable with the adult participants than with the child participants, possibly due to age-related differences in the size of semantic networks and the strength of interlinked nodes. This is to say, adult L2 speakers have already well-developed semantic networks in addition to the well-practiced mechanisms underlying interactive activation within semantic

complements, and thus a response in spoken production can be seen as a result of regulating overall activated lexicons. As regards child L2 learners, although their semantic networks are still in development, it has been shown that most of their lexicons at this stage are not randomly organized but have been clustered around neighborhoods.

To sum up, clearly, semantic neighborhood density plays a substantial role in spoken word production, and to what extent and in what manner it influences the speed of L2 lexical processing is mediated by word usage, suggesting that the SND effect exists partly due to the intermediate third variable.

Another semantic variable, the number of related senses, specifies how many senses a single word encompasses, in contrast to the semantic networks formed by a host of associated words. No effect of NoRS was observed with the adult participants, whereas NoRS contributed significantly and indirectly to predicting the response latencies in the child participants. Although the literature has reported that NoRS has facilitative effects on word identification and recognition among monolingual adults, such an ambiguity advantage did not hold for the adult participants' spoken production in the present study. These results, taken together, suggest that NoRS effects do not occur in all cases but only when certain task conditions are met. Additionally, child learners may be better able to grasp related senses of individual words rather than to form a semantic network, while the opposite pattern may hold for adult learners.

To account for the absence of significant NoRS effects among the adult

participants, another possibility is that foreign language learners likely have a far better knowledge of the core meaning of each individual word relative to its peripheral meanings. This is because the most frequently used sense of a word is undoubtedly its core meaning, and accordingly its peripheral meanings may get deemphasized by adult learners when they have not encountered those meanings for a long time or the word is usually used with extremely limited variations of contexts. It thus comes as no surprise that as language learning experience accumulates, adults tend to get more accomplished at enlarging vocabulary size and strengthening semantic networks rather than at mastering multiple (sometimes subtle) senses of individual words.

In sum, NoRS has been found to play a more important role in child L2 learners' lexical processing compared to in adult learners' lexical processing but it should be noted that its significant effects observed with the child participants are actually in part mediated by word usage.

WORD CONCRETENESS. The view that bilingual lexical processing is sensitive to the specific characteristics that words possess is initially based on the observation of different response patterns between concrete and abstract words. In the present study, however, word concreteness did not influence as expected, since concrete words were not superior to abstract words in terms of L2 lexical processing efficiency in both age groups. It could be because certain conditions need to be met for this effect to be identified. As was mentioned in the literature review, concreteness effects have been found to be sensitive to word frequency, and its magnitude is larger for

high-frequency words relative to low frequency ones (De Groot, 1992b). As a result of the inclusion of target words that covered a wide range of frequencies in the present study, concreteness effects might be consequently mitigated.

Moreover, as regards the child participants, most of the targeted vocabulary words they had been learning at this stage were concrete or nearly concrete. For this reason, variations in the distribution of word concreteness were not large, and accordingly concreteness effects were not sufficiently robust to be identified. Another point to consider relates to the teaching strategies used in primary-level foreign language contexts where language instructors hinge heavily on a variety of visual aids (e.g. images, pictures, and realia) to teach new words, as a result of which child learners' image systems are likely to be activated regardless of word types and then concreteness effects may be partialled out.

A further explanation relates to stimulus characteristics, that is, the target words were all picturable, which might make the concreteness effect less pronounced. It should be acknowledged that what was observed with word concreteness provides an example of the limitations inherent in the methodology adopted in the present study, in which some experimental variables cannot be manipulated directly.

In brief, the failure to observe significant concrete effects implies that there may be a more complex mechanism that sustains the operation of word concreteness in language production, a research area that deserves further investigation. Another point worth mentioning is that both child and adult participants' ratings of concreteness correlated well with semantic neighborhood density and word length, which indicated

support for concurrent validity of using subjectively-estimated concreteness.

6.2.2.2 Form level

The latent variable Form, underpinned by the number of letters, the number of phonemes, and the number of syllables, was not significant in predicting either children's or adults' response latencies. This result is considered in light of several previous studies that have investigated at least one of these word-length variables (see Table 2.6 above).

In a study by De Groot et al. (2002), the number of phonemes and syllables significantly predicted the performance of lexical decision among university-level Dutch-English bilingual learners but this effect was not observed with the number of letters. Additionally, De Groot (1992b) also found insignificant effects of NoL on the speed of normal translation among bilingual adults. The research by D'Amico et al. (2001) showed that the number of phonemes and syllables did not contribute significantly to picture naming latencies among native Italian children and adults. These inconsistencies do not mean that word length cannot affect any level of processing in spoken production but may indicate that NoL, NoP, and NoS have at most a weak effect on response latencies. Another possibility is that phonological encoding and its subsequent articulatory process are not sufficiently sensitive to word length.

As mentioned in the literature review, the presence of word length effects suggests sequential retrieval of phonological segments and articulatory programs, whereas the absence of word length effects indicates parallel retrieval of those

speaking components. The findings obtained in the present study would seem to support the parallel pattern. But given inconsistent results obtained in previous studies, it is clearly too early to draw a sound conclusion based on the available evidence, since more empirical investigations are needed to adjudicate between these theoretical alternatives.

6.2.2.3 Usage level

The latent variable Usage is defined by a combination of subjective frequency and age of acquisition in the adult group, whereas Usage is represented only by age of acquisition in the child group.

SUBJECTIVE WORD FREQUENCY. In general, frequency effects are indicative of learners' experiences encountering, acquiring and practicing words. As regards the adult participants, the present study replicates previous research results, showing that those words that are produced faster are used more frequently (e.g. Balota et al., 2004; Barry et al., 1997). Drawing on the *Episodic Model* mentioned in the literature review, this result confirms that the more frequently learners use an L2 word, the stronger its memory trace is consolidated, and then the more fluently the word will be accessed in spoken production. Also, the presence of frequency effects reveals that L2 spoken production is subject to the power law of practice, thus the fluency of accessing lexical items enhances with the accumulation of practice.

Moreover, the subjective usage frequency rated by the adult participants was found to correlate significantly with age of acquisition ($r = -.052$, $p < .01$). Such a high correlation validated the appropriateness of using questionnaires to obtain the

adult participants' subjective estimates. In the meantime, the reasonable model fit gives support to the view that subjective frequency and age of acquisition played somewhat independent determinant roles in language production. This finding is compatible with the view that subjective word frequency and age of acquisition share a common underlying construct, but at the same time, they have distinct properties of their own (Brysbaert & Ghyselinck, 2006). A combination of these two usage measures therefore provides complementary but not redundant information about how word usage exerts its influence on the speed of spoken production.

It should be pointed out that, in the child group, subjective frequency was found to be beyond the participants' cognitive operating system to rate accurately and reliably, which is understandable given their relatively low proficiency and relatively limited experience with English. Children's inability to rate this lexical variable was also demonstrated by the lack of a strong correlation between participant-rated frequencies and age of acquisition, and by its weak contribution made to predicting the variance in the observed outcome. This may well also explain why there is a severe paucity of research that has examined this lexical variable in relation to child learners.

To sum up, in the adult group, subjectively rated frequencies were correlated well with age of acquisition, and contributed significantly to predicting the variance in naming and translation latencies; whereas in the child group, judgments of subjective usage frequency extended far beyond the child participants' ability.

AGE OF ACQUISITION. As an alternative measure of word usage, age of acquisition

contributed significantly to the child participants' response latencies, with a total effect of .42 and .24 in picture naming and in translation respectively. This result is consistent with the literature in which significant AoA effects have been observed with monolingual children (D'Amico et al., 2001). In the adult group, the superiority of early-acquired over late-acquired words for processing was observed, a result that is compatible with several previous studies (e.g. Bonin et al., 2002; Cuetos et al., 1999; Izura & Ellis, 2004).

One explanation for the presence of AoA effects relates to Brown and Watson's (1987) *Phonological Retrieval Hypothesis* that early-acquired words are usually stored in a holistic manner, whereas late-acquired words are stored in segments, which likely incurs a cost in word retrieval during spoken production. Another possible explanation concerns the semantic characteristics manifested in words: early-acquired concepts have richer interconnectedness with other concepts than is the case for late-acquired words (Steyvers & Tenenbaum, 2005), such a semantic difference to which AoA effects can be attributed.

Moreover, both age groups showed a processing advantage for early-acquired words than for late-acquired counterparts. This observation is in agreement with the view that AoA effects are actually determined by the relative order in which lexical items are learned rather than being restricted to the influence of any critical periods (Hirsh et al., 2003).

The last point worthy of mention is that word usage, as represented by a combined variable of subjective frequency and age of acquisition in the adult group,

and by age of acquisition alone in the child group, was found to significantly mediate the effects of SND and NoRS on the speed of naming and translation, a result which suggests that usage-related effects are partly located prior to the articulation level (Brysbaert, Wijnendaele, et al., 2000). It should be stressed that the existence of such mediating effects does not mean that SND and NoRS effects are reduced to AoA effects, in that the direct contribution made by these two semantic variables to the participants' response latencies was also observed in this study although they were not as strong as their indirect effects.

In short, thanks to what is embedded in early-acquired words, this type of lexical knowledge can be accessed and retrieved more rapidly than that of late-acquired words in spoken production. Clearly, word usage can be seen as a significant determinant of L2 processing speed but it is by no means the only determinant since significant predictive effects have also been identified with other variables in the present study.

6.2.3 Summary: Learner variables vs. lexical variables

The variance explained by the lexical variables was larger than that explained by the learner variables, and such a difference was more notable with the adult participants. What these results suggest is that the mechanisms underlying L2 lexical processing are increasingly influenced or even determined by our experiences with individual words. This is to say, what learners have experienced with words is likely to individualize how they process the words at a later stage and then appears to be more important than which proficiency levels learners have achieved.

The following section will explain the results in relation to L2 lexical processing patterns among child L2 learners first and then among adult L2 learners.

6.3 Explaining the data: Mechanisms underlying L2 lexical processing

6.3.1 Child L2 learners

As mentioned in the literature review, there is an ongoing debate as to whether concept mediation or lexical association underlies early L2 learners' lexical processing. According to the *Developmental Hypothesis* (Kroll & De Groot, 1997), low-proficiency young learners would more likely access L2 target words through lexical routes rather than through conceptual routes, with empirical evidence mostly derived from the performance of adult beginning learners (e.g. Ferré, Sánchez-Casas, & Guasch, 2006; Kroll, Michael, Tokowicz, & Dufour, 2002; Talamas et al., 1999). An alternative view is that the conceptual store can be readily accessible to emergent bilingual learners even though their L2 vocabulary knowledge is limited. Empirical evidence supporting this view can be found in the studies by Chen and Leung (1989), Altarriba and Mathis (1997), and Comesaña and colleagues (2009; 2012).

The present study did not find significant differences in children's response latencies between picture naming and forward translation, thus giving support to the conceptual mediation model. This result seems to run counter to what the developmental hypothesis postulates and to echo those previous studies specifically targeted at child L2 learners. In what follows, candidate explanations for why conceptual mediation could be present in the child participants will be presented,

followed by possible accounts which might explain why these two processing patterns would more likely complement each other rather than being mutually exclusive.

The first explanation relates to the degree of L1 entrenchment, which usually increases with L2 learners' age. Children's L1s have not yet been deeply entrenched, thereby enabling a relatively high neural plasticity with which they can sponge up L2 input, and rendering L2 lexical processing less parasitic on the developing L1 (Birdsong & Molis, 2001; Flege, Yeni-Komshian, & Liu, 1999). As such, L1-L2 lexical connections may not be sufficiently strong to be activated above a threshold level, and the lexical link is therefore held to be weaker than its conceptual counterpart.

Together with the influence of L1 entrenchment, the distance between L1 lexicons and L2 lexicons, in terms of both neural and typological respects, may be another reason why conceptual mediation is implicated in children's processing of L2 lexical items. As reviewed earlier, Hernandez, Li, and MacWhinney (2005) has argued that, compared to the two languages of late learners, those of early learners tend to be more separated at the level of local cortical processing maps for articulation and lexical forms. Where this occurs, an extra effort is needed to bridge the L1 and the L2 that are spatially separated. Moreover, the perceived typological distance between Chinese and English, or more broadly, between Sino-Tibetan languages and Indo-European languages, may be larger in child L2 learners than in adult L2 learners. The most compelling evidence for this argument is that the phenomenon of cross-linguistic transfer in child L2 acquisition is not as abundant as that in adult L2

acquisition. In this light, arguably, the larger the perceived typological distance, the more likely it is that the strength of the conceptual link outweighs that of its lexical counterpart.

Third, the process by which instructed L2 teaching and learning take place provides us another insight into why the child participants were able to access form-concept mappings. It is highly likely that the participating teacher had used those types of instructional strategies that could easily facilitate the establishment of conceptual links. Through observing three sessions of the English course (approximately 3 hours), the researcher found that the participating teacher explained most of the newly encountered words using pictures, and relied heavily on presentation-practice-production instructional cycles. Also, the focus of the English lessons was placed on the development of the students' communicative competence through a variety of speaking and listening activities.

Furthermore, analyses on the participants' textbook called *PEP Primary English Students' Book* (year 5) revealed that new words are usually introduced with their associated pictures or cartoon images with which meaningful contexts are usually set up for vocabulary learning. Thus, as elucidated by Chen and Leung (1989), pictures can serve as cues for the retrieval of corresponding L2 words. It is worth mentioning that the effects of teaching methods on the activation pattern of bilingual lexicons have been examined in a study by Comesaña et al. (2009), in which only those Spanish-Basque children who learned novel words via pictures showed a significant semantic-interference effect. All in all, the way in which instructors teach new words

in input-limited contexts is believed to have a significant impact upon how these words are accessed and processed by learners at a later stage.

Fourth, in addition to instructed input, there are many resources available outside the classroom for learning English such as movies, songs, and the Internet. Using these materials may exert far-reaching influence on how L2 words are processed and represented in the mind, even though primary-age Chinese students learn English mostly through classroom instruction.

The last thing to note is that the present data seems inconsistent with a recent study by Poarch, Van Hell, and Kroll (in press), in which picture naming was found to be significantly slower than forward translation (p = .03), indicating the presence of lexical routes in 5th-grade Dutch-English children. There are two main explanations to account for this inconsistency. First, in Poarch, Van Hell, and Kroll's study, the same participants performed picture naming and then completed forward translation on almost the same set of stimuli, in which case the processing advantage found in translation might be due to a recency effect. In the present study, the same participants were used for both conditions, but each participant responded to different blocks of stimulus words that were matched on various lexical characteristics. Second, the children in Poarch, Van Hell, and Kroll's study had been learning English for 8 months, whereas the students in the present study had been learning English for two years and a half. The different lengths of learning periods may be another reason why the L1-L2 route mediated the Dutch-English learners' lexical processing whereas the concept-L2 route mediated the Chinese-English participants' lexical processing. This

is because, as will be discussed later, the age at which an L2 word is learned plays a crucial role in determining which processing pattern learners are most likely to follow. Arguably, within only an 8-month learning period, children might not have been provided with ample opportunities to practice the target stimuli, and not surprisingly, the word association model was therefore supported in Poarch, Van Hell, and Kroll's study.

The mixed results obtained thus far have illustrated the dynamics of bilingual lexical processing, which is subject to word-specific encounters. Under this scenario, it stands to reason that both processing routes are likely to be present in each individual learner, and the relative weights of these routes vary depending on what learners have experienced with individual words, and what types of words are processed in performing what kinds of tasks. To address these issues, this study used lexical items as units of analysis in an attempt to investigate under what circumstances concept mediation or lexical association most likely occurs in spoken production. More precisely, this study used lexical variables as reference points against which to analyze how these variables predicted the response-latency difference between picture naming and forward translation. Age of acquisition was found to be the most powerful predictor in this analysis, indicating that those words that are learned at an earlier stage are more likely to be conceptually processed, and accordingly those words that are learned at a later stage are more likely to be lexically processed. Based on this finding, the researcher proposed the *Functional-Developmental Model* (FDM, see Figure 6.2 below) that characterizes

bilingual lexicons in development and specifies the role of word-specific characteristics in lexical processing. The FDM is similar to De Groot's Distributed Feature Model in the sense that both assume that the structure of the bilingual lexicon is specific to word encounters and under the influence of lexical characteristics. On the other hand, the FDM is different from De Groot's model, in that it takes into account age of acquisition, a variable that was found to best predict the processing tendency in the present study.

In the remainder of this section, this proposed model will be elaborated, with a focus placed on discussing the scope and generalizability of the FDM with specific reference to previous research on bilingual lexical processing in children.

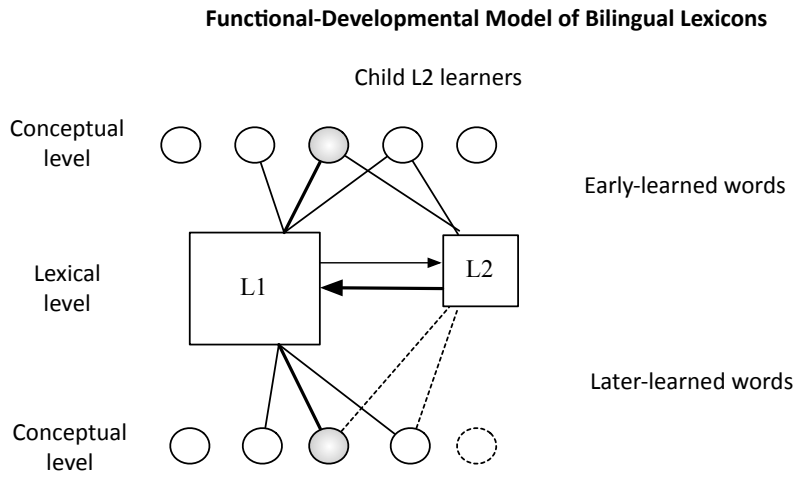


Figure 6. 2 Functional-Developmental Model of Bilingual Lexicons: Child L2 Learners

As shown in Figure 6.2, the FDM consists of two levels: the lexical level and the conceptual levels. At the lexical level, apparently, the L1 lexicon store is much larger than the L2 lexicon store because for most L2 learners, even those with advanced L2

proficiency, there are more lexical entries presented in the native than in the second language. It should be noted that, as rightly pointed out by Heredia (1997), after a long period of immersion in an L2 environment, learners' L2 may become a dominant language. Given this language shift, the labels L1 and L2 in the FDM are contextually defined by the relative dominance between the two languages rather than by the sequence of acquisition of them. Additionally, the L1 is diagrammed deeper relative to the L2 on the ground that the L1 repertoire tends to be more deeply entrenched than the L2 repertoire in a relative sense. The lexical link from the L2 to the L1 (represented by a thick line) is assumed to be stronger than that in the reverse direction (represented by a thin line).

At the conceptual levels, the L1 and L2 share common conceptual representations (represented by a dark node) irrespective of the age at which a particular lexical item is acquired. It should be noted that the degree to which L1 and L2 conceptual nodes overlap largely depends on the properties of a concept itself and learners' personal encounters with the concept. Once the target concept is activated, its semantically associated nodes also receive different degrees of activation by travelling along the links dispersed in the semantic field. Additionally, in most cases, more neighboring nodes are activated in processing the native language as compared to in processing the second language. This is because learners' more frequent use of the L1 is likely to have more neighboring nodes activated than does the L2 since the SND effects are basically mediated by word usage.

As shown in the FDM, conceptual links are the concept nodes through which

words and their corresponding concepts are connected, whereas lexical links are the associative links established in the network of L1 and L2 words. Both lexical and conceptual links are disproportionately active in bilingual memory but the relative weights assigned to each link vary with the age at which an L2 word is acquired. In processing early-acquired words, conceptual links are assumed to be stronger than lexical routes. Because of this, early-acquired words are usually mapped onto the shared concepts that have already been well developed within the L1 framework. However, it should be pointed out that, in such processes, lexical links are not completely switched off but maintain active to a lesser extent. By contrast, in processing late-acquired words, conceptual links are assumed to be stronger than lexical links, but a process in which lexical links do not disappear entirely.

The FDM is assumed to be consistent with the observation of asymmetrical translation reported in several previous studies (e.g. Kroll & Stewart, 1994; Poarch et al., in press; Sholl et al., 1995). To illustrate, in translating early-acquired words from the L1 to the L2, conceptual links and lexical links are both activated although the activation of the latter is not as strong as that of the former. In contrast, in performing backward translation, conceptual links are activated in a way similar to what occurs in forward translation but lexical links are much stronger in the L2-to-L1 direction than in the L1-to-L2 direction. Due to such a processing difference, backward translation appears to be faster than forward translation. The same explanation could also account for why translation asymmetry happens to late-acquired words. Additionally, the FDM holds explanatory power for understanding why semantic components are

actively involved in both directions of translation (e.g. Heij et al., 1996), and the semantic effects are larger in the forward direction relative to in the backward direction (e.g. De Groot et al., 1994).

The FDM seems to be in line with De Groot's *Distributed Feature Model* (1992b) in the sense that earlier-acquired words likely possess a high degree of concreteness so that the L1 and the L2 tend to have a high level of conceptual overlap; whereas later-acquired words are inclined to be more abstract and culturally-bounded thus the two languages tend to share fewer common features. Additionally, the factor age of acquisition does not exist independently of conceptual and semantic variables but actually serves as a behavioral outcome that reflects constant interactions between lexical characteristics and word usage.

In summary, the findings of the present study provide new insights into the issue of how child learners within input-limited instructional contexts process L2 words in spoken production. A response-latency comparison between picture naming and forward translation reveals that most of the L2 words are predominantly processed through conceptual links in children, and late-acquired words are more likely to be mediated by lexical routes compared to early-acquired words. Drawing on the cognitive characteristics specific to bilingual children, possible explanations have been advanced to account for these findings, together with the FDM proposed to characterize bilingual lexicons in child L2 learners.

6.3.2 Adult L2 learners

The presence of significant response-latency differences between picture naming and

forward translation indicates that L1 translation equivalents were activated when the adult participants produced the target words in the present study. This finding adds to the accumulating evidence that relatively proficient bilinguals may also employ the lexical route when accessing L2 words (e.g. Alvarez, Holcomb, & Grainger, 2003; Jiang, 2004; Menenit, 2006; Misra, Guo, Bobb, & Kroll, 2012; Thierry & Wu, 2007). However, it seems that this study's result goes against the RHM (Kroll & Stewart, 1994) in which the ability to bypass reliance on the lexical route is assumed to increase, and instead more reliance is supposed to be placed on the conceptual route as learners' L2 fluency enhances. In this case, it would be helpful to compare the present study against previous research using Potter et al.'s (1984) experimental paradigm, albeit in a small number, so as to identify why mixed results have been obtained and then to explore possible explanations for such inconsistencies.

Table 6. 1 Participant and stimulus information about studies using Potter et al.'s experimental paradigm

	Potter et al. (1984)	Kroll & Curley (1988)	Chen (1990)	The present study
Location	Boston	Massachusetts	Hong Kong	Beijing
Sample size	24	48 (two groups)	24	94
Age	College age	College age	College age	Mean = 18.80
Language experience	Cantonese-English bilinguals They had lived and studied in English-speaking countries for one or more years.	English native speakers or subjects had learned English and German simultaneously Native German speakers.	Cantonese-English bilinguals They had learned English for over 12 years.	Chinese-English bilinguals They had learned English for about 10 years in input-limited contexts.
Stimulus characteristics	96 items (members of 29 superordinate categories) Chinese names consisted of one or two characters.	6 lists of stimulus words from one or two semantic categories.	48 items of two types (pictures and characters) Chinese names consisted of one character.	Three blocks of 66 words Chinese names consisted of two characters.
Results/ Response latencies	Picture naming was slightly faster than forward translation (PN: 953 ms, FT: 1100 ms)	Picture naming was slightly faster than forward translation (PN: 1693 ms, FT: 1698 ms)	Picture naming was slightly faster than forward translation (PN: 1022 ms, FT: 1041 ms)	Picture naming was significantly slower than forward translation (PN: 1548 ms, FT: 1329 ms)

Note: FT: forward translation; PN: picture naming.

As shown in Table 6.1, the previous three studies (i.e. Chen, 1990; Kroll & Curley, 1988; Potter et al., 1984) have found no significant differences in response latencies between picture naming and translation, suggesting that relatively fluent bilinguals can access concepts directly without the mediation of L1 equivalents in producing L2 words. This result, however, was not replicated in the present study, such a discrepancy that can be accounted for by one of the following explanations.

The first explanation relates to the amount, type and intensity of L2 input that are believed to vary widely between language-rich immersion environments and input-limited instructed settings. As shown in a recent behavioral study by Linck, Kroll, and Sunderman (2009), American-English learners of intermediate Spanish

who had only classroom L2 experience produced significant semantic and lexical interference in performing translation recognition, whereas learners with similar proficiency levels who had been immersed in a study-abroad program were insensitive to the lexical form distractors but more sensitive to the semantic distractors. What this result suggests is that the learning context indeed modulates L2 lexical processing in such a way that learners are prone to inhibit the L1 while immersed in the L2 dominant environment. Even upon returning home to the L1 dominant environment, the participants remained insensitive to the lexical form, indicating that the learning context likely plays an important role in establishing and developing conceptual routes, and its importance minimally diminishes once the routes have been successfully added to the mental lexicon system.

When compared against the previous three studies where the participants had been immersed in the L2 dominant context, instructed mainly through the L2, or had learned the L1 and the L2 simultaneously, the present study involved typical foreign language learners (i.e. non-English majors) who had received much less abundant exposure to L2 input, such a contextual factor that might make the present participants rely more heavily on the lexical routes than the conceptual ones. But interestingly, those participants majoring in English who had received, on average, 40-hour exposure to the L2 per week also exhibited a tendency toward lexical association, indicating that not only the amount of input but also its type and quality play decisive roles in modulating the way in which L2 words are accessed and processed. This also reflects the fact that instructed input is essentially different from

natural input where human interactions serve communicative purposes rather than educational purposes under most circumstances.

The second explanation concerns the process by which L2 words are learned and consolidated. As Yang and Dai (2011) observed, a large number of Chinese university-level students learn English words in favor of memorizing vocabulary books and dictionaries where target words are usually presented with their translation equivalents. This type of rote learning has been widely considered by learners and teachers as an effective means of preparing for national standardized tests such as CET-4 and TEM-4. Given this, it could be argued that the extensive use of L1 equivalents in vocabulary learning and recycling further strengthens the weights of the lexical routes and accordingly weakens those of the conceptual routes.

Although the present data cannot provide direct evidence of, nor disprove, the causal relationship between learning strategies and mental representations of L2 words, the previous study by Comesaña et al. (2009) did show that more semantic nodes were activated when new words had been learned through pictures, as opposed to L1 equivalents (but also see Altarriba & Knickerbocker, 2011). As such, it is assumed that learning and teaching methods can profoundly influence the way adult learners process L2 words. This may also be one of the most important reasons why lexical associations were present in English majors who had even been exposed to the L2 for about 40 hours per week.

Third, whether or not the previous three studies have recruited sufficiently representative samples needs to be reconsidered given the inclusion of relatively small

sample sizes, which in effect might not accurately represent the population of interest as a whole. The present study, by contrast, has recruited a relatively large sample size with which statistical power of its results could increase accordingly. Additionally, caution should be taken when interpreting the results of the present study with reference to those of previous studies since they have considered different populations of learners.

The last explanation concerns stimulus characteristics or item difficulty, as indicative of learners' pre-experimental experiences with each individual word. Of particular significance are word usage frequency and age of acquisition because these variables have been found to directly influence the fluency of L2 lexical processing (e.g. De Groot, 1992b; Izura & Ellis, 2004), and to modulate the extent to which semantic and form variables have impact upon processing speed, as observed in the present study. However, without being provided with detailed information about each study's stimulus characteristics, it would be difficult to make direct comparisons between the present study and those previous ones. One exception has to do with Potter et al.'s study (1984), in which the authors briefly mentioned that its 96 items fell into 29 superordinate categories, in contrast to the present study where 198 items belonged to 10 superordinate categories. Arguably, the average frequency of the present stimuli might not be as high as that of Potter et al.'s study, a difference that could possibly make the impact of the conceptual routes less pronounced in the present study but more noticeable in Potter et al.'s study.

Most importantly, it appears that neither the word association model nor the

concept mediation model alone can capture the complexities and dynamics of bilingual lexical processing. Given this, similar to how the child data were treated and analyzed, the researcher used lexical variables as reference points against which to examine what types of words were more likely to be conceptually or lexically processed. The result showed that age of acquisition and semantic neighborhood density served as the best predictor of response-latency differences between picture naming and forward translation, indicating that those words that are learned at a later age and have sparser semantic neighbors are more likely to be processed through word associations. Based on this finding, the adult version of the Functional-Developmental Model was proposed, as shown in Figure 6.3 where both the lexical and conceptual routes coexist but their weights vary depending on the characteristics of lexical items. The following sections will discuss this model's generalizability and problem-solving ability within the literature of L2 lexical processing.

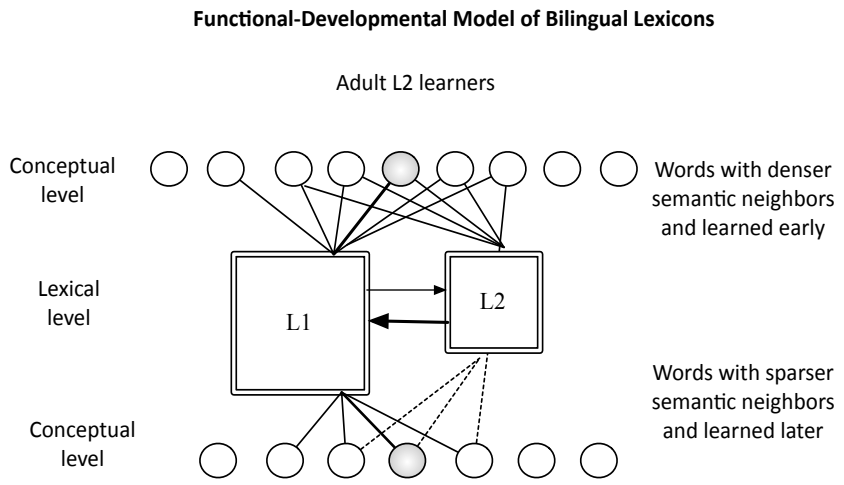


Figure 6. 3 Functional-Developmental Model of Bilingual Lexicons: Adult L2 Learners

Similar to what constitutes children's bilingual lexicons, the FDM involves both lexical and conceptual routes, and the activation level of each route varies with the age at which an L2 word is learned but also with the density of its semantic neighbors. To be precise, the L1 is assumed to play a more active role in processing late-learned and low-SND words relative to early-acquired and high-SND words, whereas the impact of lexical associations is likely to be reduced but not entirely absent when an L2 word has been learned less recently and has denser semantic neighborhoods. It should be noted that a comparison between Figure 6.2 and Figure 6.3 illustrates that there are at least two processing differences that exist between bilingual children and adults. First, compared to children's bilingual lexicons, adults' L1s are more deeply entrenched and the distance between the L1 and the L2 is hypothesized to be smaller. Additionally, as mentioned earlier, the brain plasticity decreases with age due to certain areas of the brain becoming more neutrally committed, a change that has been captured by the double lines around the squares of L1 and L2 lexicons in Figure 6.3. Second, from childhood to adulthood there is an increase in the number of semantic neighbors and in the connectivity of semantic networks, thus it comes as no surprise that more circles are shown at the conceptual levels and more conceptual-lexical connections are present in Figure 6.3 than in Figure 6.2.

For the same reason stated above in Section 6.3.1, the adult version of the FDM is assumed to be consistent with the observed asymmetry in translation latencies, that is, forward translation usually takes longer to complete than backward translation. Moreover, as mentioned earlier in the literature review, backward translation can be accomplished with relative ease of accessing the lexical route, whereas forward translation involves a higher degree of conceptual activation. The FDM is compatible with this observation in the sense that when processing early-acquired words, the

conceptual links involved in forward translation are believed to be stronger than those in backward translation. This is because the native language has more and stronger connections to the target concept and its neighboring ones. In contrast, the reverse pattern holds true for the lexical links: the forward direction is assumed to be weaker compared to its backward direction. A similar explanation applies to late-acquired words to account for why different degrees of lexical and conceptual activation are involved in forward and backward translation respectively. It therefore seems reasonable to conclude that both conceptual and lexical routes are involved in processing L2 lexicons in both directions of translation among adult L2 learners but comparatively the conceptual nodes are more actively involved in forward than backward translation whereas the lexical nodes are more actively involved in backward than forward translation.

To sum up, this study has demonstrated the presence of lexical routes in processing most of the L2 words among adult learners who have learned an L2 in input-limited contexts. But it should be emphasized that this does not mean that their conceptual activation is completely switched off in spoken production, at least because accessing earlier-acquired and higher-SND words has been found to engage conceptual activation in the present study. Based on these findings, the adult version of the FDM has been proposed to characterize bilingual lexicons specific to relatively fluent bilinguals.

6.4 Summary: Dynamics of bilingual lexical processing

In order to capture the dynamics of bilingual lexical processing, as far as possible, this study proposed a schematic model that aims to incorporate all possible variables that interact with and determine the way in which bilingual lexicons are represented in the

mind (see Figure 6.4). The model distinguishes between conceptual and intentional contents, processing mechanisms, and an output phase. Main factors that affect bilingual lexical processing are a complex set of learner and lexical characteristics, and the language task schemas that encourage or discourage cross-language competition. To be more specific, these influential variables can be further divided into the following: the experience that a learner has gained with individual words, the level of proficiency that a bilingual has achieved in his/her native and second languages, the age at which a particular word is acquired and the strategy used to acquire that word, the density of semantic neighborhood that a word possesses, the typological differences between a bilingual's two languages, the context in which L2 learning takes place, and the specific processing demands of a language task. All these variables work together to influence the mechanisms underlying L2 lexical processing, but in comparison, how learners have experienced individual words seems to be more important than the levels of proficiency that learners have achieved in their L2.

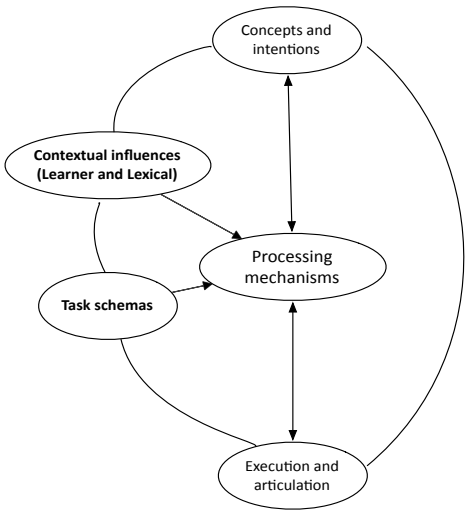


Figure 6. 4 Dynamics of bilingual lexical processing

Chapter 7 Conclusion

This chapter begins with a discussion on the limitations of the present study, and then suggests several directions for future research. After this, some pedagogical implications are proposed to inform classroom practice particularly within the context of foreign language learning and teaching. This chapter ends with some concluding remarks.

7.1 Limitations of the present study and recommendations for future research

7.1.1 Limitations of the present study

Although the present study provides some insights into the effects of learner and lexical variables on the fluency of processing L2 words and into the structure of bilingual lexicons across age group, it has several limitations that should be acknowledged.

First, this research is in part limited by its explanatory power due to the nature of the research design adopted. Given the use of a non-experimental design, it would be difficult to make unequivocal explanations of causality among the variables of interest. For example, it seems logical to reason that age of acquisition plays a causative role in determining the mechanisms underlying L2 lexical processing but it could be additionally possible that the effects of age of acquisition on the outcome variable are actually mediated by those variables that are not included in the present study. Moreover, not all the variance in the participants' performance can be explained by the variables included. Some of the unexplained variance could be attributed to

several factors that are beyond the scope of the present study, such as the degree of Chinese characters' transparency, the extent of cross-linguistic semantic similarity, and phonological patterns. Lastly, using multiple regression and structural equation modeling to analyze the data makes possible a much more comprehensive look at the relationships among the lexical variables, but this approach lacks a specific focus on certain variables through purposeful manipulation of experimental materials. In this case, it is not possible to make inferences about how a particular lexical variable affects the speed of processing on the basis of the data obtained in the present study.

Second, reaction time-based research can hardly provide explicit insight into language processing in real situations where a wide variety of linguistic, social-cultural, and cognitive factors are bound to affect language speakers' behaviors. Admittedly, using reaction times assist researchers in inferring the mechanisms underlying bilingual lexical processing but this type of experiments needs to be complemented with other approaches in order to fully understand what goes on when L2 learners produce words in the real world. Additionally, as pointed out by Jiang (2012), reaction time-based research is less direct or transparent compared to those studies drawing on error or accuracy data, but it should be acknowledged that many issues related to L2 mental representations and processes can hardly be addressed without having reaction time data.

As is the case with research design, control strategies likely enhance a study's internal validity but at the same time may lower its external validity. Given the design of this study, some limitations to external validity are present. Although this research

strived to gain a relatively diverse sample of child learners, it turned out that only one group of primary-age students was recruited. As such, the child sample might not represent a full continuum of L2 children, especially those who take intensive English programs at the elementary level in China. But this could also reflect a practicality that it is often easier for researchers to gain access to universities than primary schools to carry out psycholinguistic experiments due to the fact that in primary schools students have to be pulled out of classrooms one by one and this may cause interruptions to the established educational routines, whereas at universities, students are able to perform experiments in any time slots arranged with a researcher in advance, thus educational interruptions can be kept to a minimum.

Another possible threat to external validity concerns the types of stimulus materials used in the present study. Although the stimuli were selected based on the criteria that they varied widely in terms of the lexical characteristics under examination, it should be acknowledged that the selected words were all picturable in the present experiment, which might make the concreteness effect less pronounced. Moreover, the child participants were believed to have a very limited number of vocabulary words after receiving only about 100 hours of classroom exposure to the target language. Also, most of the words that they had been learning at this stage were concrete or nearly concrete. For these reasons, the number of words that the children were able to produce was actually limited, thus what has been observed with the child participants can be applicable only to the very earliest stages of L2 word learning.

Although the participants' language proficiency has been measured by

standardized assessments with adequate validity and reliability, some limitations should be noted. To begin with, assessments on four language skills were not administered to the adult participants since operationalizing such tests would be costly and time-consuming. More importantly, it was no easy task to recruit participants who were willing to spend a considerable amount of time (on average, 3-4 hours) taking standardized tests in the laboratory. To attenuate, if not overcome, this shortcoming, two types of vocabulary tests were administered instead to the adult participants in the present study, which is understandable given that vocabulary breadth and depth would be most relevant to the outcome variable. Moreover, the child participants had never taken the Cambridge Young Learners English (YLE) tests prior to joining this research project, in which case the students might experience difficulty with the test format that they were not familiar with. In order to address this problem, they had been given a chance to practice the YLE tests before the real test was administered, although eventually not every participant might feel comfortable taking the YLE tests.

Fourth, the present study has included a range of lexical variables but it seems implausible to cover all lexical characteristics that each individual word possesses for at least one of the following reasons: (1) it was far beyond the child participants' cognitive operations to rate some lexical characteristics accurately and consistently; (2) some lexical variables were not compatible with the specified structural or path models hypothesized in the present study; and (3) some lexical variables held extremely limited explanatory power to the model above and beyond their competing counterparts.

Generalizability of the present results is another limitation. Since the present study recruited beginner learners (age 10-11 years) and intermediate learners (aged 17-20 years), the collected data may not be representative of adolescents and advanced learners. As mentioned above, the present study consisted of picturable words; therefore, future effort is needed to examine whether the same result holds for non-picturable words. It seems impossible to include every lexical variable of L2 words in one study. Even so, it should be acknowledged that the results obtained in this study could only be generalizable to the lexical variables of interest. Any inference made beyond the lexical variables tested here should be taken with caution. Additionally, the native language of the present participants falls into the Sino-Tibetan language family. It thus seems difficult to generalize the present results to those who have an Indo-European language as the first.

This section has discussed the limitations of the present study. In what follows, the focus will be on possible directions for future research.

7.1.2 Recommendations for future research

Some recommendations for future research are outlined as follows.

First, much more needs to be learned about the lexical variables that contribute significantly to the speed of processing Sino-Tibetan languages. Most of the previous studies have focused on Indo-European languages, particularly English, French, and Dutch (e.g. Alario et al., 2004; Bonin et al., 2002; Cuetos et al., 2006; Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; De Groot, 1992b; Larochelle & Pineau, 1994; Severens et al., 2005), whereas there has been very little research targeted at

Sino-Tibetan languages (e.g. Chinese, Lolo-Burmese, and Tibetic). Chinese, for example, has strikingly different features compared to Indo-European languages, since it has a high degree of compounding and lexical tone, and its writing system is logographic rather than orthographic. Examining these under-researched linguistic systems is considered necessary because it would allow us to gain a better understanding of the nature of lexical characteristics by addressing the issue of whether a specific lexical effect is language-dependent or universal across languages.

Second, given that language production (top-down) and comprehension (bottom-up) are underpinned by two distinctively different processing mechanisms, it is deemed necessary to examine how a variety of lexical variables affect L2 comprehension so as to address whether or not and in what manner lexical effects vary with different task types. The language tasks that typically involve comprehension are lexical decision and semantic categorization although these two tasks are different in the sense that the former can be done without semantic knowledge (Joordens & Becker, 1997), whereas performing the latter relies more on semantic cues. Although there exist several research studies examining lexical effects in word recognition, a particular effort needs to be made to investigate whether or not word usage mediates the relationship between semantic and form variables, and learners' decision times in a way similar to what occurs in L2 spoken production. Moreover, eye tracking is another method that can be used to study the time course of learners' lexical processing through recording continuous eye movements (e.g. first fixation, single fixation duration, gaze duration, and total reading time). Eye tracking

is particularly well suited to examining how a lexical variable exerts its influence when learners read a sentence or text, and such information is not available in using word-by-word based experiments. As such, another recommendation for future study is to use eye tracking to gain a better understanding of the interaction between lexical variables and learners' reading performance in an attempt to develop evidence-based interventions to assist learners in the comprehension of texts.

Third, future researchers would be well advised to move beyond cross-sectional comparisons between age cohorts to examine developmental trends in a longitudinal fashion. The cross-sectional design has been widely used to explore processing similarities and differences across age groups and proficiency levels at the same point in time but any conclusions with regard to the course of developmental changes may run the risk of being confounded with cohort effects. More importantly, cross-sectional comparisons can hardly provide answers to such questions as 'When does a lexical item become more conceptually mediated?' or 'Will this learner process the same set of words more conceptually after three-year exposure to an L2?' To address issues like these, longitudinal designs are preferred over cross-sectional designs although it usually takes more time and resources to track changes longitudinally. Furthermore, the FDM proposed in the present study aims to accommodate different types of bilingual learners, since almost every word can be defined as either early acquired or late acquired. Even so, it would be beneficial for future research to explore how well the FDM fits with other types of bilinguals, particularly those who learn an L2 within language-rich contexts. Such information

would allow us to better assess the overall generalizability of the FDM, and modifications, if needed, can be made to it for the purpose of accommodating different types of bilingualism, as far as possible.

The fourth recommendation is that future researchers can focus exclusively on examining words that represent objects, actions, number, space or social partnership, because each of these five separable systems has its own deep roots in human phylogeny and ontogeny (Spelke & Kinzler, 2007). Also, it would be interesting to investigate how different types of words interact with each other, and how these words can eventually be integrated into the mental lexicon system in a structured way. In addition, relatively little is known about how compound words, idioms and other types of fixed expressions are represented in the mind, thus future effort may also be made to delineate the precise manner in which these lexical chunks are accessed and processed among bilingual learners.

Last but not least, it would be necessary to further examine how different teaching methods affect the structure of bilingual lexicons and the mechanisms underlying bilingual lexical processing, as such knowledge will hold important implications for pedagogical practices in classrooms. Also, it remains to identify whether the establishment and activation of lexical routes facilitate or inhibit targeted vocabulary gains and accessibility, and how these routes impact on learners' performance on four language skills. To put it in another way, future research needs to address whether concept mediation is always superior to lexical association in terms of the rate and route of L2 lexical acquisition and retention, and if not, in what

circumstances one processing pattern leads to greater vocabulary gains than the other. All in all, the gap between what has been observed in the laboratory and what occurs in the classroom needs to be bridged so as to inform and transform educators' practices. Also, bilingual lexical processing needs to be examined as an end state but equally important as a process.

7.2 Pedagogical implications

This section attempts to explore possible ways in which the present study provides insight into current conceptualizations of vocabulary learning and teaching within instructed L2 contexts.

First, given that word usage was found to mediate the extent to which conceptual, semantic, and form variables come into play in spoken production, it would be advisable for language practitioners to teach vocabulary in a usage-based fashion. The central idea behind the usage-based approach is that a learner's language grows out of specific instances in which language comprehension and production take place (Kemmer & Barlow, 2000). This is to say, language usage operates based on a linguistic system's structures and functions but also shapes personal understanding of the linguistic system through cumulative acts of mental processing involved in different kinds of language activities.

It should be acknowledged that the present study results could not provide direct implications for the practice of usage-based in L2 classroom. Even through, it is considered necessary to introduce its main characteristics, among others.

(a) *The primary purpose of vocabulary instruction is to develop learners'*

communicative competence with which they are able to use words fluently and appropriately. As defined by Hymes (1972), communicative competence encompasses not only knowledge of the rules of grammar but also the ability to use a language. To be precise, communicative competence is built upon four cornerstones: grammatical (formally feasible), psycholinguistic (implementationally feasible), sociocultural (contextually appropriate) and behavioral (actually incurring) knowledge (Hymes, 1972, p. 281).

(b) *Context plays a crucial role in the operation of a linguistic system* (Kemmer & Barlow, 2000). This is because words are usually used to comprehend and express meaning that is embedded within specific contexts, and beyond that contextual elements constantly interact with mental representations of words. In this light, the various contexts in which a lexical item typically occurs should be considered as an integral part of vocabulary instruction and acquisition.

(c) *A gradual experience-driven learning process is held to be central to vocabulary development.* Since the usage-based approach stresses the importance of understanding and producing a language, learners should be provided with ample opportunities to gain hands-on experiences practicing the targeted vocabulary words in meaningful contexts where illocutionary acts are the goals to be achieved through performing locutionary acts (Robinson & Ellis, 2008). To this end, comprehension and production are considered integral, not periphery, to the development of learners' lexical knowledge.

(d) *High-frequent uses promote learning and entrenchment* (i.e. cognitive

routinization) (for a review, see Ellis, 2012). As shown in the present study, high-frequency words were processed more rapidly compared to low-frequency words in picture naming and forward translation. Frequency has been unanimously recognized as a shaping force in linguistic representations and processing, which is particularly evidenced by language users' high sensitivity to recurrent patterns and units. As such, recycling lexical items through usage-based practices should be encouraged so as to make the target input more salient and outstanding to learners, and eventually to facilitate their vocabulary gains.

The above-listed four tenets can be applied to guide instructional activities in the following ways:

(a) *There should be a balance in vocabulary instruction across four strands — meaning-focused input, meaning-focused output, language-focused learning, and fluency development* (Nation, 2007). The communicative approach to vocabulary teaching emphasizes the importance of using authentic or natural materials and the implementation of problem-solving activities (e.g. information gap exercises). But it should be noted, as pointed out by Schmitt (2000), that engaging learners in communicative interactions is not a sufficient condition to assure adequate vocabulary knowledge, therefore it is strongly recommended that targeted vocabulary words should be selected in a principled manner and previously learned words should be continuously recycled in classrooms. For example, we may consider lexical characteristics in determining the sequence with which vocabulary words are learned, since learners' language performance was found to be sensitive to lexical

characteristics in the present study. Moreover, effective vocabulary teaching, especially within time-restricted learning contexts, should embrace not only meaning-focused instruction but also form-focused elements, especially when considering the results of several pervious studies showing that explicit instruction increases the likelihood that learners better acquire and retain the meaning of new words at least in the short term (e.g. Laufer, 2006; Laufer & Girsai, 2008; Sonbul & Schmitt, 2010; Tian & Macaro, 2012).

(b) *Context-dependent vocabulary instruction assists learners in gaining an in-depth understanding of a word but this teaching method needs to be undertaken in a principled manner.* First of all, language practitioners are well advised to make every effort to ensure that learners have a minimum level of knowledge to adequately understand the context in which a target word is embedded; otherwise it would be extremely difficult for learners to infer the target meaning correctly. In addition, learners should be given intensive instructions on how to accurately derive and infer what a word means from its embedded context, because having such strategic knowledge would better enable learners to turn an implicit learning process to an explicit one, and then to better memorize the meaning of a target word. Lastly, it should be pointed out that the combined use of implicit and explicit teaching strategies might lead to the greatest overall vocabulary gains. As suggested by Ramachandran and Rahim (2004), in explaining a new L2 word, it would be helpful for instructors to provide its L1 equivalent first, followed by its English synonym before presenting the target word in context. Also, as shown in the present study, the

word association model characterized the adult learners' lexical processing, indicating that the L1 likely plays a role in L2 vocabulary learning. Taken together, context itself is an invaluable resource for enriching learners' understanding of individual words provided that this teaching approach can be appropriately exploited.

(c) *Ample opportunities should be provided for learners to accumulate participatory experiences with individual words.* Both comprehension-based and production-based instructions are essential to language acquisition in the sense that the former provides learners with a lexical repertoire to capitalize on, whereas the latter leads learners to process known words more deeply. Given the availability of a wide variety of teaching activities, it would be advisable for language practitioners to assess the extent to which learners' motivational-cognitive loads, in terms of *need*, *search*, and *evaluation* respects, are involved in performing different types of tasks (i.e. *Involvement Load Hypothesis*, Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001). According to the hypothesis, the higher the level of involvement load, the more likely it is to promote vocabulary acquisition. What this suggests is that instruction should be devised to maximize the opportunities for learners to get enough experiences with targeted vocabulary words through engaging them in cognitively appropriate activities.

(d) *Usage frequency should be carefully considered in determining what and how words to be taught.* The present study showed that usage frequency served as a significant predictor of the efficiency of lexical processing. In this regard, frequency bands are central to determining which words are worthy of instruction at different

stages of learning processes. Instructors and course book designers are therefore advised to arrange vocabulary words based on their usage frequencies. Additionally, how well learners are expected to master a target word and how teachers are expected to teach the word should somewhat depend on its frequency level.

Usage-based vocabulary instruction contributes to narrowing the gap between what an L2 learner has achieved in classrooms and how the L2 is actually used in the real world. However, as Einstein observed, 'Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted.' Likewise, the usage-based approach can hardly be considered as a panacea for all learning situations, and indeed it needs to be implemented in a principled manner as suggested above.

A second pedagogical implication is based on the importance of lexical characteristics in spoken production, as observed in the present study. More attention should be placed on lexical characteristics in developing course books and curriculum standards; however, unfortunately, these features have not yet received sufficient attention. This may be one of the reasons why so many course books have selected targeted vocabulary words in an unsystematic manner. As reported in a study by Konstantakis and Alexiou (2012), some of the EFL course book series that are currently used in Greece appear to be thinly loaded with too many high-frequency words, which could potentially slow down students' progress due to the availability of a narrow range of vocabulary. In contrast, some other course books appear to fully or heavily loaded with too many low-frequency words, resulting in an escalation of

demands that could potentially overburden students' cognitive capacities. In order to attenuate, if not eliminate, these problems, language practitioners, course book writers, and curriculum designers are encouraged to weigh up the opportunities and cognitive loads that learning each individual word may bring about. One approach to achieve this is to use lexical characteristics as reference points against which to select appropriate target words.

One lexical characteristic that deserves our close consideration pertains to word typicality since this variable has been found to significantly predict the efficiency of L2 lexical processing for both age groups in the present study. As mentioned earlier in the literature review, a prototypical item that bears more similarities to its central members tends to be learned and retrieved more rapidly than does a non-prototypical item. This is particularly true for children who have shown sensitivity to the manipulation of word typicality in a direct manner. As such, language educators would be well advised to consider the internal structure of concepts to determine the sequence in which words are taught and learned in instructed L2 contexts, especially in the very earliest stages of vocabulary learning.

Another lexical variable worthy of consideration is usage frequency, since this variable served as the best predictor over and above others in the present study. It is conventionally held that those words with the highest frequency of usage should be taught and learned first. There are two very different ways to determine a word's frequency: one is to count the number of times the word occurs in a large volume of written and/or spoken materials, and the other is to ask participants to rate how

frequently they have used the word over a fixed period of time. In comparison, the former has been used more widely than the latter in selecting and arranging words at the present; however, the objective approach may not work equally well for all types of language learners. A case in point is that the most frequently used words in a monolingual corpus may not be the same as those used by bilingual speakers. As the results shown in the present study, objective frequency counts based on monolingual resources are not as powerful as subjective frequency estimates in predicting the child and adult L2 participants' productive performance. In such a case, it would be advisable for language educators to assess the difficulty level of words with reference to the frequency databases specifically targeted at foreign language learners although de facto not many databases of this type are available nowadays.

In contrast, frequency estimates generated by learners themselves appeared to explain a large amount of variance in the performance of language production in the present study. In this regard, it would be advisable for classroom instructors to record and to track how frequently a word is actually used by students in real practice. This information would then help curriculum designers and course book writers to decide whether the learning materials have provided both range and depth of lexical items necessary for learners to make progress and whether learners have acquired those words required to achieve a specific proficiency level. In short, teachers are encouraged to play a much more active role in influencing what vocabulary to be taught, and their voices may eventually lead to change or revise current syllables and policies.

The last pedagogical implication relates to the roles that learners' L1s are likely to play in L2 vocabulary learning and teaching. As assumed in the FDM, L2 lexical processing is partially underpinned by cross-linguistic associations, indicating that the L1 regularly influences L2 lexical processing although its influence varies with different learning situations and different types of learners. This view is particularly true when vocabulary learning is seen as a process rather than as an end state. It thus seems inappropriate to regard teachers' L1 use as a source of negative interference that inhibits the rate and route of L2 vocabulary acquisition. Indeed, it has been consistently found that L2 learners benefit from being provided with teachers' L1 use more than with L2-only explanations, particularly in terms of short-term learning (e.g. Acha, 2009; Laufer & Girsai, 2008; Ramachandran & Rahim, 2004; Zhao & Macaro, in press), and this superiority appears to be more obvious with young learners (e.g. Lee & Macaro, 2013). It should be pointed out that emphasizing the value of teachers' L1 use does not mean to play down communicative-based input but rather to suggest that a balanced language course should consist of not only meaning-focused input but also form-focused instruction.

Given the above, it would be helpful for language educators to be familiar with the process by which L2 words are accessed in comprehension and production. In teacher training courses, the issue of L1 use should be thoroughly discussed, which will make teachers aware of the advantages and disadvantages of using the L1 in L2 classes. Additionally, it would be advisable for policy makes and syllabus designers to seriously consider the role of medium of instruction in L2 vocabulary learning and

teaching, and to provide evidence-based guidelines for teachers to follow in practice.

To sum up, this section has suggested possible implications for current vocabulary teaching practices in input-limited contexts. Based on the findings obtained in the present study, the researcher advocates the implementation of an usage-based approach to L2 vocabulary teaching and learning but this pedagogy needs to be balanced by the inclusion of form-focused instruction and to give consideration to the roles that lexical characteristics likely play in influencing the ease or difficulty with which targeted vocabulary words can be acquired and retained.

7.3 Conclusion

The time taken to produce the name of a picture and to translate L1 words into L2 is sensitive to a variety of lexical characteristics. This research provides converging and new evidence on these lexical effects, and extends previous studies by considering age as a factor. On the methodological side, structural equation modeling was used to investigate the complex interrelationships among the lexical variables concerned, and to examine their effects on the recorded response latencies, a recommended analytical strategy that was seldom used for this type of experiment. By using this particular analytical strategy, this study has observed the presence of the indirect effects of word meaning and form on the speed of L2 lexical processing via word usage. That is to say, usage likely mediates how meaning and form affect the speed of spoken production, and such effects are more noticeable with adult L2 learners than with child L2 learners. Therefore, a note of caution may be raised when interpreting the results of previous studies where the mediating effects of word usage have not been

adequately addressed.

Over the past decades, much attention has been paid to examining the mechanisms and processes of L2 lexical access. Although no definite conclusion has yet been reached thus far, it is tentatively suggested that production of most L2 words engages different degrees of conceptual and lexical activation. The hard problems that researchers face nowadays are whether or not these two routes are competitive, cooperative or competitively cooperative per se, as well as in what circumstances one route is more active than the other. Another important issue that deserves our careful consideration is whether the co-activation of both routes is manifested only in each individual participant or can be extended to each individual lexical item. Based on the findings obtained in the present study, at the very least, it can be argued that taking lexical items as points of departure will give us additional insight into how L2 lexical items are processed and represented in the mind, the results of which cannot be obtained using by-participant analyses alone.

In closing, three main conclusions can be drawn from the present study. First, different aspects of vocabulary knowledge were found to predict the child and adult participants' response latencies respectively. Child learners who have a larger vocabulary size and adult learners who have richer vocabularies are able to access L2 words faster than those who do not. Second, a large amount of variance in the fluency of spoken production was explained by the lexical variables included. Atypical and low-frequency words are usually produced more slowly than are typical and high-frequency words in child and adult L2 learners. Semantic neighborhood density,

the number of related senses, and word typicality significantly and indirectly predicted naming and translation times via word usage in some experiments of the present study. This leads to the implication that only through word usage does the influence of some lexical variables come into play, echoing the view that 'Only in the stream of thought and life do words have meanings' (Wittgenstein, 1967, p.31). Third, the present study has shown that the concept mediation model is not restricted only to more-fluent L2 learners, and also the word association model is not confined only to less-fluent L2 learners. Additionally, the earlier an L2 word is learned, the greater the likelihood that the word will be conceptually accessed and processed at a later stage. These observations further demonstrate the dynamics and complexities involved in delineating the processes and mechanisms of bilingual lexical processing.

References

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40*(1), 23-31.

Aitchison, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon* (4th ed.). Oxford and New York: Wiley-Blackwell.

Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers, 31*(3), 531-552.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, 36*(1), 140-155.

Altarriba, J., & Knickerbocker, H. (2011). Acquiring second language vocabulary through the use of images and words. In T. P. & M. K. (Eds.), *Applying Priming Methods to L2 Learning, Teaching and Research: Insights from Psycholinguistics* (pp. 21-48). Amsterdam: John Benjamins Publishing Company.

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36*(4), 550-568.

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

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Arbuckle, J. (2009). *Amos 18 User's Guide*. Armonk, NY: SPSS Incorporated.

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2010). *Introduction to Research in Education* (8th ed.). Belmont, CA: Cengage Learning.

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition, 12*(01), 83-95.

Baars, B. J., & Gage, N. M. (2010). *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience* (2nd ed.). San Diego: Academic Press.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition, 57*(2), 193-216.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly, 36*(3), 421-458.

Bailey, A. (2005). Test Review: Cambridge Young Learners English (YLE) Tests. *Language Testing, 22*(2), 242-252.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340-357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639-647.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology, 50*(3), 560-585.

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.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., . . . Pléh, C. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10*(2), 344-380.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78-117.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research, 24*(1), 57-65.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience, 17*(6), 905-917.

Birdsong, D. (2009). Age and the end state of second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *The New Handbook of Second Language Acquisition* (pp. 401-424). Amsterdam: Elsevier.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44*(2), 235-249.

Bjorklund, D. F. (2000). *Children's Thinking: Developmental Function and Individual Differences* (3rd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Bjorklund, D. F., & Thompson, B. E. (1983). Category typicality effects in children's memory performance: Qualitative and quantitative differences in the processing of category information. *Journal of Experimental Child Psychology, 35*(2), 329-344.

Bloomfield, L. (1933). *Language*. New York: Henry Holt.

Blunch, N. (2012). *Introduction to Structural Equation Modeling Using IBM SPSS Statistics and AMOS*. Thousand Oaks, CA: Sage.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology, 93*(1), 89-114.

Bormann, T. (2011). The role of lexical-semantic neighborhood in object naming: Implications for models of lexical access. *Frontiers in Psychology, 127*(2).

Breakwell, G. M., Hammond, S., & Fife-Schaw, C. (2000). *Research Methods in Psychology*. Thousand Oaks, CA: Sage.

Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15*(3), 208-216.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 325-337.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.

Brysbaert, M. (2007). The language-as-fixed-effect fallacy: Some simple SPSS solutions to a complex problem. *London: Royal Holloway, University of London*.

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.

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition, 13*(7-8), 992-1011.

Brysbaert, M., Lange, M., & Wijnendaele, I. V. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology, 12*(1), 65-85.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 1-8.

Brysbaert, M., Wijnendaele, I. V., & Deyne, S. D. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica, 104*(2), 215-226.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8 (3)*, 531-544.

Byrne, B. M. (2009). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (2nd ed.). New York, NY: Taylor and Francis Group.

Byrne, B. M. (2011). *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. New York: Routledge.

Byrne, D. (2002). *Interpreting Quantitative Data*. London: Sage.

Cambridge ESOL n.d.: *Cambridge young learners English tests.* Cambridge: University of Cambridge ESOL Examinations. Retrieved March 1, 2013 from http://www.teachers.cambridgeesol.org/ts/exams/younglearnersandforschools.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-experimental Designs for Research*. Boston, MA: Houghton Mifflin Company.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177-208.

Caramazza, A., & Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology/Revue canadienne de psychologie, 34*(1), 77-81.

Caramazza, A., & Costa, A. (2001). Set size and repetition in the picture–word interference paradigm: Implications for models of naming. *Cognition, 80*(3), 291-298.

Carroll, D. W. (2008). *Psychology of Language* (5th ed.). Belmont, CA: Cengage Learning/Wadsworth.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology, 25*(1), 85-95.

Cattell, J. M. K. (1886). The time it takes to see and name objects. *Mind, 11*(41), 63-65.

Center for the studies of Chinese as a Second Language at Beijing Language and Culture University, Language corpus system of modern Chinese study. Retrieved January 15, 2013 from http://www.dwhyyjzx.com/cgi-bin/yuliao.

Chen, H. C. (1990). Lexical processing in a non-native language: Effects of language proficiency and learning strategy. *Memory & Cognition, 18*(3), 279-288.

Chen, H. C., & Leung, Y. S. (1989). Patterns of lexical processing in a non-native language. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*(2), 316-325.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335-359.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6th ed.). Oxon: Routledge.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497-505.

Comesaña, M., Perea, M., Piñeiro, A., & Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Psychology, 104*(1), 22-33.

Comesaña, M., Soares, A. P., Sánchez‐Casas, R., & Lima, C. (2012). Lexical and semantic representations in the acquisition of L2 cognate and non‐cognate words: Evidence from two learning methods in children. *British Journal of Psychology, 103*(3), 378-392.

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition, 125*(3), 452-465.

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism, 10*(2), 137-152.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *The Quarterly Journal of Experimental Psychology, 60*(8), 1072-1082.

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. d. Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 308-325). New York: Oxford University Press.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition, 2*(03), 231-244.

Costa, A., Heij, W. L., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition, 9*(02), 137-151.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41*(3), 365-397.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*(4), 491-511.

Cuetos, F., Alvarez, B., González-Nosti, M., Méot, A., & Bonin, P. (2006). Determinants of lexical access in speech production: Role of word frequency and age of acquisition. *Memory & Cognition, 34*(5), 999-1010.

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, 31*(4), 650-658.

Cummins, J. (1991). Interdependence of first-and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language Processing in Bilingual Children* (pp. 70-89). New York: Cambridge University Press.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990? *Cognition, 10*(1), 65-70.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of experimental child psychology, 65*(2), 171-237.

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development, 2*(1), 71-105.

Dalgish, G. M. (1991). Computer-assisted error analysis and courseware design: Applications for ESL in the Swedish context. *CALICO Journal, 9*(2), 39-56.

Daniel, J. (2011). *Sampling Essentials: Practical Guidelines for Making Sampling Choices*. London: Sage.

Dattalo, P. (2008). *Determining Sample Size: Balancing Power, Precision, and Practicality*. New York: Oxford University Press.

Dattalo, P. (2010). *Strategies to Approximate Random Sampling and Assignment*. New York: Oxford University Press.

De Groot, A. M. B. (1992a). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning* (pp. 389-412). Amsterdam: Elsevier.

De Groot, A. M. B. (1992b). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1001-1018.

De Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 27-51). Amsterdam: John Benjamins.

De Groot, A. M. B., Borgwaldt, S., Bos, M., & Van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47*(1), 91-124.

De Groot, A. M. B., Dannenburg, L., & Vanhell, J. G. (1994). Forward and backward word translation by blinguals. *Journal of Memory and Language, 33*(5), 600-629.

De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50*(1), 1-56.

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.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292-307.

DeKeyser, R., & Larson-Hall, J. (2005). What does the Critical Period really mean. In J. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 88-108). New York: Oxford University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(04), 499-533.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283-321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*(2), 124-142.

Dent, K., Johnston, R. A., & Humphreys, G. W. (2008). Age of acquisition and word frequency effects in picture naming: A dual-task investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(2), 282-301.

Desrochers, A., Liceras, J. M., Fernández-Fuertes, R., & Thompson, G. L. (2010). Subjective frequency norms for 330 Spanish simple and compound words. *Behavior Research Methods, 42*(1), 109-117.

Desrochers, A., & Thompson, G. L. (2009). Subjective frequency and imageability ratings for 3,600 French nouns. *Behavior Research Methods, 41*(2), 546-557.

Dijkstra, T., & Van Heuven, W. J. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist Connectionist Approaches to Human Cognition* (225 ed., pp. 189-225). Mahwah, NJ: Lawrence Erlbaum.

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Two words, one meaning: Evidence of automatic co-activation of translation equivalents. *Frontiers in Psychology, 2*(188), 1-20.

Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language Learning, 47*(1), 173-210.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition, 23*(2), 166-180.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology, 57*(2), 98-107.

Durda, K., & Buchanan, L. (2006). WordMine2. Retrieved December 1, 2012. from http://www.wordmine2.org

Durgunoğlu, A. Y., & Roediger, H. L. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language, 26*(4), 377-391.

Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: Evidence from the semantic Simon paradigm. *Psychonomic Bulletin & Review, 15*(5), 961-966.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421-436.

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic‐level categories in young infants. *Child Development, 65*(3), 903-917.

Ellis, N. (2012). Frequency-based accounts of second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 193-210). New York: Routledge.

Evans, V. (2010). Language and cognition: the view from cognitive linguistics. In V. Cook & B. Bassetti (Eds.), *Language and Bilingual Cognition* (pp. 67-107). New York: Psychology Press.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120*(4), 339-357.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

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.

Field, A. (2013). *Discovering Statistics Using SPSS* (4th ed.). London: Sage.

Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1075-1089.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51*(1), 1-22.

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition, 9*(02), 153-166.

Fivush, R. (1989). Scripts and categories: Interrelationships in development. In U. Neisser (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization* (pp. 234–254). Cambridge: Cambridge University Press.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Boston: The MIT Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 627-635.

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.

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. Kroll & A. M. B. D. Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 251-267). New York: Oxford University Press.

Francis, W. S., Tokowicz, N., & Kroll, J. F. (2014). The consequences of language proficiency and difficulty of lexical access for translation performance and priming. *Memory & Cognition, 42*(1), 27-40.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*(1), 27-52.

Gardner, H. (1974). Metaphors and modalities: How children project polar adjectives onto diverse domains. *Child Development, 45*(1), 84-91.

Garrard, P., Patterson, K., Watson, P. C., & Hodges, J. R. (1998). Category specific semantic loss in dementia of Alzheimer's type. Functional-anatomical correlations from cross-sectional analyses. *Brain, 121*(4), 633-646.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Speech and Talk* (Vol. 1, pp. 177-220). London: Academic Press

Garson, G. D. (2012a). Path Analysis: Quantitative Methods. Retrieved December 1, 2013 from http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm.

Garson, G. D. (2012b). Structural Equation Modeling: Quantitative Methods. Retrieved December 1, 2013 from http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm.

Geeraerts, D. (2006). Prototype theory. In D. Geeraerts (Ed.), *Cognitive Linguistics: Basic Readings* (pp. 141-165). Berlin/New York: Mouton de Gruyter.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. (2002). Motion Events in Language and Cognition. *Cognition, 83*(1), 49-79.

Gerhand, S., & Barry, C. (1999a). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition, 27*(4), 592-602.

Gerhand, S., & Barry, C. (1999b). Age-of-acquisition and frequency effects in speeded word naming. *Cognition, 73*(2), 27-36.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*(2), 256-281.

Geva, E., & Ryan, E. B. (1993). Linguistic and cognitive correlates of academic skills in first and second languages. *Language learning, 43*(1), 5-42.

Glaser, W. R. (1992). Picture naming. *Cognition, 42*(1), 61-105.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251-279.

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.

Goodwin, C. J. (2009). *Research in Psychology: Methods and Design*. Hoboken, NJ: John Wiley & Sons.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language, 24*(6), 631-645.

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

Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the Behavioral Sciences*. CA: Thomson.

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.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition, 1*(02), 67-81.

Greenberg, M. S., & Bjorklund, D. F. (1981). Category typicality in free recall: Effects of feature overlap or differential category encoding? *Journal of Experimental Psychology: Human Learning and Memory, 7*(2), 145-147.

Greidanus, T., & Nienhuis, L. (2002). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal, 85*(4), 567-577.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics: Speech Acts* (pp. 41-58). New York: Academic Press.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21*(02), 205-228.

Hacking, I. (1981). *Scientific Revolutions*. Oxford: Oxford University Press.

Hair, J. F. J., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2005). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Hampton, J. (1993). Prototype models of concept representation. In I. V. Mechelen, J. Hampton, R. Michalski & P. Theuns (Eds.), *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp. 67-95). London: Academic Press.

Harrington, D. (2008). *Confirmatory Factor Analysis*. New York: Oxford University Press.

Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing, 23*(1), 73-98.

Hartas, D. (2010). *Educational Research and Inquiry: Qualitative and Quantitative Approaches*. London: Continuum International Publishing Group.

Haugen, E. (1953). *The Norwegian Language in America: A Study in Bilingual Behaviour*. Philadelphia: University of Pennsylvania Press.

Heij, W. L., Hooglander, A., Kerling, R., & Van Der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language, 35*(5), 648-665.

Heredia, R. R. (1997). Bilingual memory and hierarchical models: A case for language dominance. *Current Directions in Psychological Science, 6*(2), 34-39.

Hermans, D. (2000). *Word production in a foreign language.* (Unpublished doctoral dissertation), University of Nijmegen, Nijmegen, The Netherlands.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9*(5), 220-225.

Hindmarsh, R. (1980). *Cambridge English Lexicon*. Cambridge: Cambridge University Press.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1331-1356.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93*(4), 411-428.

Hirsh, K. W., Morrison, C. M., Gaset, S., & Carnicer, E. (2003). Age of acquisition and speech production in L2. *Bilingualism: Language and Cognition, 6*(2), 117-128.

Hollis, M. (1994). *Philosophy of Social Science*. Cambridge: Cambridge University 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.

Howell, D. C. (2011). *Statistical Methods for Psychology*. Belmont, Calif:: Wadsworth Cengage Learning.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*(2), 151-195.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.

Hymes, D. H. (1972). On Communicative Competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected Readings* (pp. 269-293). Baltimore, USA: Penguin Education.

IELTS (2010). *IELTS Annual Review*, UCLES/British Council, IDP Education Australia, Cambridge.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology, 2*(255), 1-16.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica, 127*(2), 277-288.

Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language, 50*(2), 165-181.

Jackendoff, R. (2011). Conceptual semantics. In C. Maienborn, K. V. Heusinger & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 1, pp. 688-709). Berlin: De Gruyter Mouton.

Jerger, S., & Damian, M. F. (2005). What's in a name? Typicality and relatedness effects in children. *Journal of Experimental Child Psychology, 92*(1), 46-75.

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

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47-77.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal, 88*(3), 416-432.

Jiang, N. (2012). *Conducting Reaction Time Research in Second Language Studies*. New York: Routledge.

Jin, Y. S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills, 70*(3), 1139-1154.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120*(1), 113.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition, 13*(7/8), 789-845.

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1083.

Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7. Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods*. Mooresville, IN: Scientific Software.

Kan, P. F., & Kohnert, K. (2005). Preschoolers Learning Hmong and EnglishLexical-Semantic Skills in L1 and L2. *Journal of Speech, Language, and Hearing Research, 48*(2), 372-383.

Kawamoto, A. H., Farrar, W. T., & Kello, C. T. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1233-1247.

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage-based Models of Language* (pp. pp. vii–xxviii). Stanford, CA: CSLI Publications.

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language, 49*(4), 484-499.

Kirsner, K. (1994). Implicit processes in second language learning. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 283-312). San Diego, CA: Academic Press.

Kirsner, K., Lalor, E., & Hird, K. (1993). The bilingual lexicon: Exercise, meaning and morphology. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 215-248). Amsterdam: Benjamins.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language, 81*(1), 205-223.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics, 20*(1), 1-24.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). New York: Guilford Press.

Kolers, P. A. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior, 2*(4), 291-300.

Konstantakis, N., & Alexiou, T. (2012). Vocabulary in Greek young learners' English as a foreign language course books. *The Language Learning Journal, 40*(1), 35-45.

Kormos, J. (2006). *Speech Production and Second Language Acquisition*. Mahwah, NJ: Routledge.

Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 53-81). Amsterdam: John Benjamins.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*(3), 416-430.

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*(02), 119-135.

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: John Wiley & Sons.

Kroll, J. F., & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. d. Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18*(2), 137-171.

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.

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*(03), 373-381.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. d. Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 289-307). New York: Oxford University Press

Lambert, W. E., Havelka, J., & Crosby, C. (1958). The influence of language-acquisition contexts on bilingualism. *The Journal of Abnormal and Social Psychology, 56*(2), 239-244.

Lambert, W. E., Havelka, J., & Gardner, R. C. (1959). Linguistic manifestations of bilingualism. *The American Journal of Psychology, 72*(1), 77-82.

Larochelle, S., & Pineau, H. (1994). Determinants of response times in the semantic verification task. *Journal of Memory and Language, 33*(6), 796-823.

Laudan, L. (1978). *Progress and its Problems: Towards a Theory of Scientific Growth*. Berkeley: University of California Press.

Laufer, B. (2006). Comparing focus on form and focus on formS in second-language vocabulary learning. *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 63*(1), 149-166.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied*

Linguistics, 29(4), 694-716.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26.

Lee, J. H., & Macaro, E. (2013). Investigating age in the use of L1 or english - only instruction: Vocabulary acquisition by Korean EFL learners. *The Modern Language Journal, 97*(4), 887-901.

Leech, N., Barrett, K., & Morgan, G. A. (2007). *SPSS for Intermediate Statistics: Use and Interpretation* (3rd ed.). Mahwah, NJ: Erlbaum.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: John Wiley & Sons.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1995). The ability to speak: from intentions to spoken words. *European Review, 3*(1), 13-23.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 83-122). Oxford: Oxford University Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*(1), 122-142.

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychology and Aging, 17*(4), 677-689.

Levinson, D. (1998). *Ethnic Groups Worldwide: A Ready Reference Handbook*. Santa Barbara, CA: Greenwood Publishing Group.

Li, W. (2000). Dimensions of bilingualism. In W. Li (Ed.), *The Bilingualism Reader* (pp. 3-25). London: Routledge.

Lichacz, F. M., Herdman, C. M., Lefevre, J. A., & Baird, B. (1999). Polysemy effects in word naming. *Canadian Journal of Experimental Psychology, 53*(2), 189-193.

Linck, J. A., Kroll, J. F., & Sunderman, G. (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.

Littlewood, W. (1984). *Foreign and Second Language Learning: Language Acquisition Research and its Implications for the Classroom*. Cambridge: Cambridge University Press.

Lockhart, R. S. (2000). Methods of Memory Research. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 44-58). New York: Oxford University Press.

Luecht, R. M. (2003). Multistage complexity in language proficiency assessment: A framework for aligning theoretical perspectives, test development, and psychometrics. *Foreign Language Annals, 36*(4), 527-535.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge: Cambridge University Press.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review, 89*(5), 483-506.

MacKay, D. G. (1987). *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills*. New York: Springer Verlag.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39*(3), 384-389.

Macnamara, J. (1969). How can one measure the extent of a person's bilingual proficiency?. In L. Kelly (Ed.), *Description and Measurement of Bilingualism* (pp. 80-99). Toronto: Univesrity of Toronto Press.

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

Major, R. (1987). The natural phonology of second language acquisition. In A. James & J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 207-224). Dordrect, The Netherlands: Foris.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40*(2), 230-262.

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language, 49*(1), 20-42.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition, 10*(1), 69-75.

Mäntyla, T. (1993). Priming effects in prospective memory. *Memory, 1*(3), 203-218.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics, Series B, 36*(2), 115-128.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97-116.

Marsh, H. W. (1987). The hierarchical structure of self‐concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement, 24*(1), 17-39.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*(4), 462-472.

McFalls, E. L., Schwanenflugel, P. J., & Stahl, S. A. (1996). Influence of word meaning on the acquisition of a reading vocabulary in second-grade children. *Reading and Writing, 8*(3), 235-250.

McGrath, J. (1982). Dilemmatics: The study of research choices and dilemmas. In J. McGrath (Ed.), *Judgment Calls in Research* (pp. 69-80). Beverly Hills: Sage.

McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language, 29*(1), 43-65.

Meara, P. M. (1992). *EFL Vocabulary Tests*. Swansea: Centre for Applied Language Studies.

Meara, P. M. (2005). *X_Lex: The Swansea Vocahulaiy Levels Test (version 2.05)* Swansea: Lognostics.

Menenit, L. (2006). L2-L1 word association in bilinguals: Direct evidence. *Nijmegen CNS, 1*(1), 17-24.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society, 7*(3), 283-284.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25-40.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language, 48*(1), 131-147.

Ministry of Education of China. (2007). College English Curriculum Requirements. Beijing: Foreign Language Teaching and Research Press.

Ministry of Education of China. (2012). University Entrance Examination Syllabus. Beijing: Higher Education Press.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language, 67*(1).

Moon, J., & Jiang, N. (2012). Non-selective lexical access in different-script bilinguals. *Bilingualism: Language and Cognition, 15*(01), 173-180.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50*(3), 528-559.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 21(1), 116-133.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology, 91*(2), 167-180.

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20*(6), 705-714.

Morrison, K. R. (2009). *Causation in Educational Research*. London: Routledge.

Muñoz, C. (2006). The Effects of Age on Foreign Language Learning: The BAF Project. In C. Munoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 1-40). Clevendon: Multilingual Matters.

Murphy, G. L. (2004). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Mutter, S. A., & Hashtroudi, S. (1987). Cognitive effort and the word frequency effect in recognition and lexical decision. *The American Journal of Psychology, 100*(1), 93-116.

Myers, A., & Hansen, C. (2011). *Experimental Psychology* (7th ed.). Belmont, CA: Wadsworth Publishing Company.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.

Nation, I. S. P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching, 1*(1), 2-13.

Nation, I. S. P. (Ed.). (1984). *Vocabulary Lists: Words, Affixes and Stems*. English University of Wellington, New Zealand: English Language Institute.

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review, 99*(2), 322-348.

Nelson, K. (1996). *Language in Cognitive Development: The Emergence of the Mediated Mind*. New York: Cambridge University Press.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences, 10*(1), 147-172.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*(1), 11-28.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Oancea, A., & Furlong, J. (2007). Expressions of excellence and the assessment of applied and practice‐based research. *Research Papers in Education, 22*(2), 119-137.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*(4), 273-281.

Olobatuyi, M. E. (2006). *A User's Guide to Path Analysis*. Lanham, MD: University Press of America.

Osgood, C. E., & Sebeok, T. A. (1954). Psycholinguistics: a survey of theory and research problems. *The Journal of Abnormal and Social Psychology, 49*(4, Part 2), 1-203.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue canadienne de psychologie, 34*, 390-401.

Paivio, A., & O'Neill, B. J. (1970). Visual recognition thresholds and dimensions of word meaning. *Perception & Psychophysics, 8*(5), 273-275.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. d. Groot & J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 331-354). Mahwah, NJ: Lawrence Erlbaum Associates.

Paradis, M. (2000). Cerebral representation of bilingual concepts. *Bilingualism: Language and Cognition, 3*(01), 22-24.

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam/Philadelphia: John Benjamins.

Parks, R., Ray, J., & Bland, S. (1998). Wordsmyth English dictionary—Thesaurus. Retrieved 2013, October 26 from http://www.wordsmyth.net/.

Pavlenko, A. (1999). New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition, 2*(03), 209-230.

Pavlenko, A. (2008). Structural and conceptual equivalence in the acquisition and use of emotion words in a second language. *The Mental Lexicon, 3*(1), 91-120.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 125-160). Bristol, UK: Multilingual Matters.

Pavlenko, A., & Driagina, V. (2007). Russian emotion vocabulary in American learners' narratives. *The Modern Language Journal, 91*(2), 213-234.

Perea, M., Duñabeitia, 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.

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). The neural consequences of semantic richness - When more comes to mind, less activation is observed. *Psychological Science, 18*(5), 401-406.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review, 15*(1), 161-167.

Pimsleur, P. (1967). A memory schedule. *Modern Language Journal*, *51*(2), 73-75.

Poarch, G. J., Van Hell, J. G., & Kroll, J. F. (in press). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism: Language and Cognition*, 1-15.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.

Potter, M. C., So, K. F., Eckardt, B. V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23-38.

Poulisse, N. (1999). *Slips of the Tongue: Speech Errors in First and Second Language Production*. Amsterdam/Philadelphia: John Benjamins.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*(1), 36-57.

Pring, R. (2005). *Philosophy of Education: Aims, Theory, Common Sense and Research*. New York: Continuum International Publishing Group.

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.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review/La revue canadienne des langues vivantes, 56*(2), 282-308.

Ramachandran, S. D., & Rahim, H. A. (2004). Meaning recall and retention: The impact of the translation method on elementary level learners' vocabulary learning. *RELC Journal, 35*(2), 161-178.

Raupach, M. (1984). Formulae in second language production. In H. W. Dechert, D. Möhle & M. Raupach (Eds.), *Second Language Productions* (pp. 114-137). Tübingen: Gunter Narr.

Ray, W. J. (2012). *Methods toward a Science of Behavior and Experience* (10th ed.). Belmont, California: Wadsworth.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355-371.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing* (pp. 209-227). Amsterdam: John Benjamins

Robinson, P., & Ellis, N. C. (2008). Conclusions: Cognitive Linguistics, second language acquisition and L2 instruction—Issues for research. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 489-545). New York/London: Routledge.

Robson, C. (2002). *Real World Research*. Malden: Blackwell.

Rodd, J. M. (2004). The effect of semantic ambiguity on reading aloud: A twist in the tale. *Psychonomic Bulletin & Review, 11*(3), 440-445.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: effects of semantic relatedness. *Memory & Cognition, 40*(7), 1095-1108.

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*(2), 245-266.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89-104.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1), 107-142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*(3), 249-284.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition, 69*(2), 219-230.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language*. New York: Academic Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192-233.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural Bases of Typicality Effects. *Journal of Experimental Psychology: Human Perception and Performance, 2*(4), 491-502.

Sachs, O. S., & Coley, J. D. (2006). Envy and jealousy in Russian and English: Labeling and conceptualization of emotions by monolinguals and bilinguals. In A. Pavlenko (Ed.), *Bilingual Minds: Emotional Experience, Expression, and Representation* (pp. 209-231). Clevedon, UK: Multilingual Matters.

Schacter, D. L. (2007). Memory: Delineating the core. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts* (pp. 23-27). New York: Oxford University Press.

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*(6), 1270-1281.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*(4),

971-977.

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.

Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology, 24*(5), 591-601.

Schwanenflugel, P. J., & Akin, C. E. (1994). Developmental trends in lexical decisions for abstract and concrete words. *Reading Research Quarterly, 29*(3), 251-264.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*(5), 499-520.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 82-102.

Searle, J. R. (1969). *Speech Acts: An essay in the Philosophy of Language* (Vol. 626). Cambridge: Cambridge University Press.

Searle, J. R. (2010). *Making the Social World*. New York: Oxford University Press.

Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia, 20*(6), 615-627.

Severens, E., Lommel, S. V., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*(2), 159-187.

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.

Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory & Cognition, 31*(1), 100-113.

Snodgrass, J. G. (1993). Translating versus picture naming. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 83-114). Amsterdam/Philadelphia: John Benjamins.

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *English Language Teaching Journal, 64*(3), 253-260.

Southgate, V., & Meints, K. (2000). Typicality, naming, and category membership in young children. *Cognitive Linguistics, 11*(1/2), 5-16.

Sparrow, S. S., & Davis, S. M. (2000). Recent advances in the assessment of intelligence and cognition. *Journal of Child Psychology and Psychiatry, 41*(1), 117-131.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science, 10*(1), 89-96.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience, 4*(3), 232-243.

Squire, L. R., Rolls, E. T., Johnson, M. K., & Buckner, R. L. (2007). Memory systems. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts*. New York: Oxoford University Press.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, 93*(24), 13515-13522.

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the Psychology of Language* (Vol. 1, pp. 143-186). Hove, England: Lawrence Erlbaum Associates.

Stevens, J. (2009). *Applied Multivariate Statistics for the Social Sciences*. New York: Taylor and Francis Group.

Steyvers, M., & Tenenbaum, J. B. (2005). The large‐scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science, 29*(1), 41-78.

Storkel, H. L., & Adlof, S. M. (2009a). Adult and child semantic neighbors of the Kroll and Potter (1984) nonobjects. *Journal of Speech, Language, and Hearing Research, 52*(2), 289-305.

Storkel, H. L., & Adlof, S. M. (2009b). The effect of semantic set size on word learning by preschool children. *Journal of Speech, Language, and Hearing Research, 52*(2), 306-320.

Sunderman, G., & 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.

Syllabus for College English Test (2006). Shanghai: Shanghai Language Education Press.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics* (6th ed.). Boston: Pearson Education.

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.

Tehan, G., & Tolan, G. A. (2007). Word length effects in long-term memory. *Journal of Memory and Language, 56*(1), 35-48.

Terrace, H. S. (1985). In the beginning was the 'name'. *American Psychologist, 40*(9), 1011-1028.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*(30), 12530-12535.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university level students. *Language Teaching Research, 16*(3), 367-391.

Titone, R. (1972). *Le Bilinguisme Pécoce*. Paris: Dessart.

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.

Treiman, R., Clifton, C., Meyer, A. S., & Wurm, L. H. (2003). Language comprehension and production. In A. F. Healy & R. W. Procotor (Eds.), *Experimental Psychology* (Vol. 4). New York: Wiley.

Trochim, W. M., & Donnelly, J. P. (2008). *Research Methods Knowledge Base* (3rd ed.). Mason, Ohio: Atomic Dog/Cengage Learning.

Trofimovich, P., Martin-Chang, S., & Levesque, K. (2012). Age effects in L2 learning: comparing child and adult learners' performance on tests of implicit and explicit memory. In J. Altarriba & L. Isurin (Eds.), *Memory, Language, and Bilingualism: Theoretical and Applied Approaches* (pp. 161-187). Cambridge: Cambridge University Press.

Tse, C. S., & Altarriba, J. (2009). The word concreteness effect occurs for positive, but not negative, emotion words in immediate serial recall. *British Journal of Psychology, 100*(1), 91-109.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381-402). New York: Academic Press.

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition, 26*(6), 1282-1291.

Tyler, L. K., Moss, H., Durrant-Peatfield, M., & Levy, J. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language, 75*(2), 195-231.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2*(10), 717-726.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1), 231-270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition* (pp. 141-178). Washington, DC: Gergetown University Press.

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*(3), 193-211.

Van Loon-Vervoorn, W. A., Van Ham, P., & Van der Koppen, M. (1988). The importance of age of acquisition for imageability in word processing. In M. Dennis, J. Engelkamp & J. T. E. Richardson (Eds.), *Cognitive and Neuropsychological Approaches to Mental Imagery* (pp. 99-107). Dordrecht: Martinus Nijhoff.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip

of Italian tongues. *Psychological Science, 8*(4), 314-317.

Vigliocco, G., & Hartsuiker, R. J. (2005). Maximal input and feedback in production and comprehension. In A. Cutler (Ed.), *Twenty-first Century Psycholinguistics: Four Cornerstones* (pp. 209-228). Mahwah, NJ: Lawrence Erlbaum Associates.

Walley, A. C. (1993). The Role of Vocabulary Development in Children's Spoken Word Recognition and Segmentation Ability. *Developmental Review, 13*(3), 286-350.

Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus*. Chichester: John Wiley & Sons.

Wang, X., & Forster, K. I. (2010). Masked translation priming with semantic categorization: Testing the Sense Model. *Bilingualism: Language and Cognition, 13*(03), 327-340.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*(3), 829-853.

Wechsler, D. (1999). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Harcourt Assessment.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology, 50*(2), 439-456.

Weinreich, U. (1953). *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.

Wiese, R. (1984). Language production in foreign and native languages: Same or different. In W. Dechert, D. Mohle & M. Raupach (Eds.), *Second Language Productions* (pp. 11-25). Tubingen: Gunter Narr.

Winch, P. (2003). *The Idea of a Social Science and its Relation to Philosophy* (2nd ed.). London: Routledge.

Wu, M. (2012). An investigation of Chinese undergraduate English majors' graduation theses from 2007 to 2009. *English Linguistics Research, 1*(1), 97-103.

Xiao, L. (2007). Assessing the roles of breadth and depth of vocabulary knowledge in second language proficiency. *Foreign Language Teaching and Research, 39*(5), 352-559.

Yang, W., & Dai, W. (2011). Rote memorization of vocabulary and vocabulary development. *English Language Teaching, 4*(4), 61-64.

Yates, M. (2012). The influence of semantic neighbours on visual word recognition. *Journal of Research in Reading, 35*(2), 215-226.

Yates, M., Locker, L., & Simpson, G. B. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Memory & Cognition, 31*(6), 856-866.

Zdrazilova, L., & Pexman, P. M. (2013). Grasping the invisible: Semantic processing of abstract words. *Psychonomic Bulletin & Review, 20*, 1312-1318.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*(1), 1-29.

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition, 32*(1), 31-38.

Zhao, T., & Macaro, E. (in press). What works better for the learning of concrete and abstract words: teachers' L1 use or L2-only explanations? *International Journal of Applied Lingusitics*.

Zhou, Y. , Gao, Y.H., & Zang, Q. (2010). The development of English learning motivatino among Chinese undergraduates in Junior and Senior years. *English Education in China*, 29 (3), 1-10.

Zughoul, M. R. (1991). Lexical choice: Towards writing problematic word lists. *International Review of Applied Linguistics, 29*(1), 45-60.

Appendices

Appendix 3.1 Participation information sheet for teachers

DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0) 1865 274024 Fax: +44 (0) 1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk
Director Professor Ernesto Macaro



PARTICIPATION INFORMATION SHEET FOR TEACHERS

Dear head teachers/directors/teachers,

I, Ting Zhao, am currently reading for a DPhil in Education at the University of Oxford. My research is aimed at examining how Chinese students process different types of English words. As you know, vocabulary learning is a very important component of language learning. But many Chinese students find vocabulary learning difficult. Therefore, by investigating the possible mechanisms underlying vocabulary learning, we are hoping to provide insights into Chinese teachers' vocabulary teaching practice. Your participation in this study will help us to find a better way of teaching English words.

If you decide to participate we would ask you to present this project to your students and to give them information leaflets and consent forms. We would really appreciate if we are allowed to approach the students during school hours in term 2 and term 3. When we come to visit your school, we will work with each student individually in a quiet area of the school on some word tasks on a computer. As for the students, they will respond to different pictures and words as they appear on a computer screen. After this, we will ask them a few general questions about their English learning experience, and then they will be given some tasks about English vocabulary. It will take about 60 minutes for each participant to complete all the tasks. We would like to express our sincere thanks for your help with this project.

Your participation is completely voluntary. The data collected in this research will remain confidential and will be used only for the purpose of this research. Results for each participant and the identity of the school will be kept strictly confidential. Participants will be identified only by code numbers, and all information and results will be kept in a secure file to which only the researcher has access over the course of study. All electronic data will be kept on a secure computer designated for this study. All data will be destroyed after the research is completed. You and your students would be free to withdraw from this study at anytime without any penalty. You are encouraged to ask any questions that you may have about this study. However, answers that could influence the results of this study will be deferred to

the end of this study. You have the right to inquire about the results of this study. This study has received ethics clearance through the University of Oxford's ethical approval process for research involving human participants.

If you agree to take part in this study, please complete and sign the attached consent form and return it to me as soon as possible.
If you have any questions, please feel free to contact me at ting.zhao@wolfson.ox.ac.uk. If you are not satisfied with any aspects of this research, you can also contact my supervisor Dr. Victoria Murphy at victoria.murphy@education.ox.ac.uk. If you remain unhappy and wish to make a formal complaint, please contact the Chair of DREC, Dr. Lars Malmberg at lars-erik.malmerg@education.ox.ac.uk.

Yours sincerely,



Appendix 3.2 Participation information sheet for students

DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0) 1865 274024 Fax: +44 (0) 1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk
Director Professor Ernesto Macaro



PARTICIPATION INFORMATION SHEET FOR STUDENTS

Dear students,

I, Ting Zhao, am currently reading for a DPhil in Education at the University of Oxford. My research is aimed at examining how Chinese students process English words. You are being invited to take part in this research project. Your participation will help us to better understand the mechanisms underlying English vocabulary learning.

If you agree to be in this study, we would like you to complete some computer-based tasks in a quiet area near your classroom. You will be asked to respond to words and pictures as they appear on a computer screen. You will be given the option to take a break when needed. After this, you will be asked to rate word features and to do several small tasks about English vocabulary. Then, we will ask you a few general questions about your English learning experience. All questions are confidential, and you are free to miss out any questions you do not want to answer. All the tasks will take about 60 minutes to complete.

It is unlikely that you will experience any risks or discomforts in this study. Your grade will not be affected in any way. Your participation is completely voluntary. The data collected in this research will be used only for the purpose of this research. Results for each participant will be kept strictly confidential. You will be identified by a code number, and all the information and results will be kept in a secure file to which only the researcher has access over the course of study. All electronic data will be kept on a secure computer designated for this study. All data will be destroyed after the research is completed. You would be free to withdraw from this study at anytime without any penalty. You are encouraged to ask any questions that you may have about this study. However, answers that could influence the results of this study will be deferred to the end of this study. You have the right to inquire about the results of this study. You will receive a small gift after you have completed all the tasks.

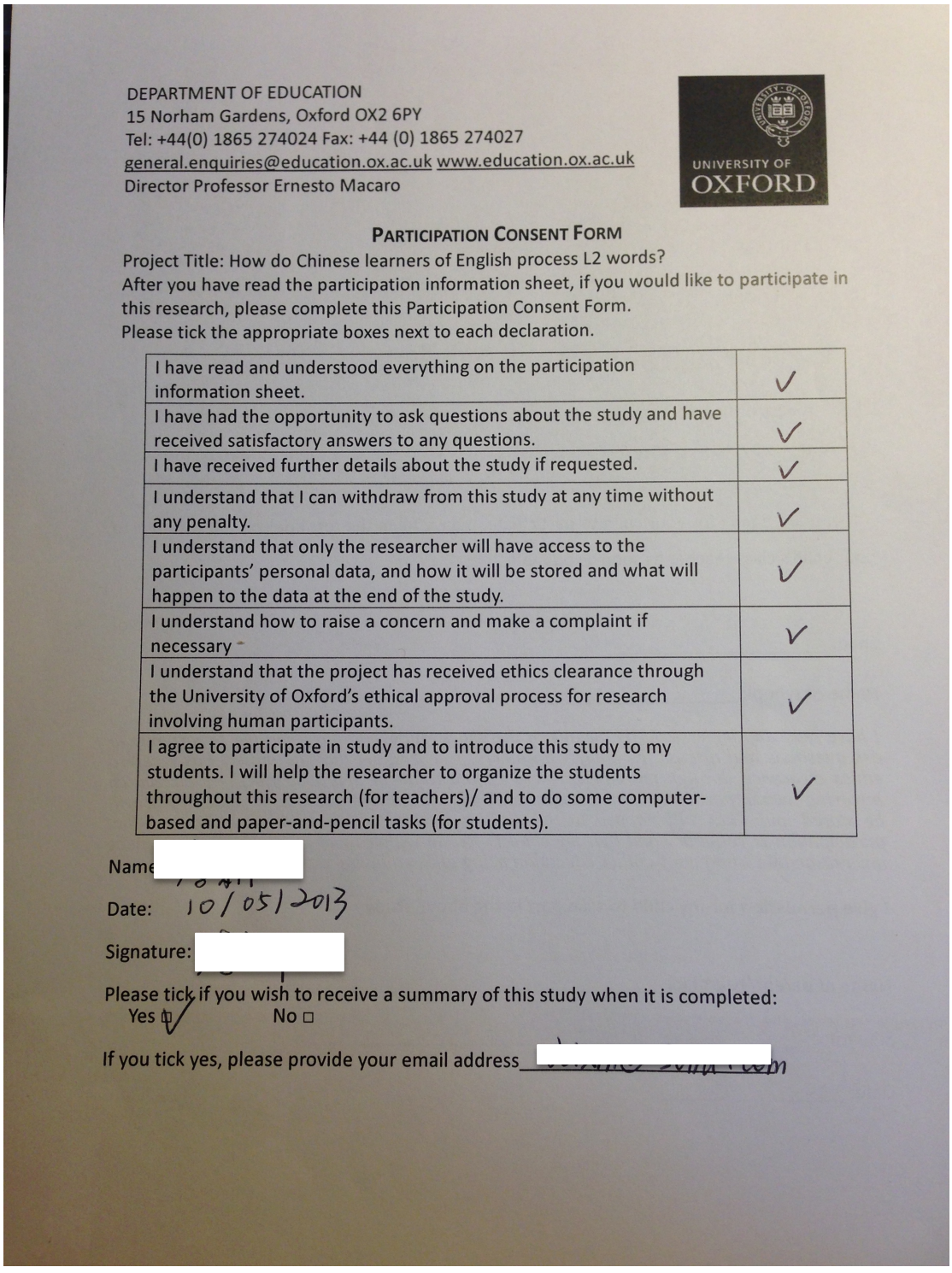
This study has received ethics clearance through the University of Oxford's ethical approval process for research involving human participants. If you agree to take part in this study, please complete and sign the attached Consent Form and return it to your teacher as soon as possible.

If you have any questions, please feel free to contact me at ting.zhao@wolfson.ox.ac.uk. If you are not satisfied with any aspects of this research, you can also contact my supervisor Dr. Victoria Murphy at victoria.murphy@education.ox.ac.uk. If you remain unhappy and wish to make a formal complaint, please contact the Chair of DREC, Dr. Lars Malmberg at lars-erik.malmerg@education.ox.ac.uk.

Yours sincerely,



Appendix 3.3 Consent form for teachers and students



DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0) 1865 274024 Fax: +44 (0) 1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk
Director Professor Ernesto Macaro

UNIVERSITY OF OXFORD

PARTICIPATION CONSENT FORM

Project Title: How do Chinese learners of English process L2 words?
After you have read the participation information sheet, if you would like to participate in this research, please complete this Participation Consent Form.
Please tick the appropriate boxes next to each declaration.

I have read and understood everything on the participation information sheet.	✓
I have had the opportunity to ask questions about the study and have received satisfactory answers to any questions.	✓
I have received further details about the study if requested.	✓
I understand that I can withdraw from this study at any time without any penalty.	✓
I understand that only the researcher will have access to the participants' personal data, and how it will be stored and what will happen to the data at the end of the study.	✓
I understand how to raise a concern and make a complaint if necessary	✓
I understand that the project has received ethics clearance through the University of Oxford's ethical approval process for research involving human participants.	✓
I agree to participate in study and to introduce this study to my students. I will help the researcher to organize the students throughout this research (for teachers)/ and to do some computer-based and paper-and-pencil tasks (for students).	✓

Name [redacted]

Date: 10/05/2013

Signature: [redacted]

Please tick if you wish to receive a summary of this study when it is completed:
Yes ☑ No ☐

If you tick yes, please provide your email address [redacted]

Appendix 3.4 Participation information sheet for parents

DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0) 1865 274024 Fax: +44 (0) 1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk
Director Professor Ernesto Macaro

研究项目说明函

亲爱的家长们,

您们好！本人，赵婷，目前正在牛津大学攻读应用语言学博士学位。博士论文是关于中国孩子如何加工不同类型的英语词汇的研究。您孩子所在的学校已表示愿意参与到此次调研活动中。在此我们想邀请您的孩子参与到此次调研中。他们的参与将帮助研究者更好地了解词汇习得的过程和机制。

在调研期间，研究人员将定期拜访您孩子的的学校。参与者需要独立完成电脑生成的任务，对电脑屏幕上出现的图片进行命名，或者把出现的中文单词译成英文单词。另外，他们需要完成两个有关词汇的测试，并填写与英语词汇相关的调查问卷。完成所有的任务需大概一小时。

这项研究不会给您和您的孩子带来任何不适或危险。您和您孩子的参与是完全基于自愿的原则。参与者可以随时终止研究关系，这并不会给参与者带来任何的负面影响。我们将确保数据的保密性。研究者只能通过编号识别每位参与者。研究完成后所有的数据将被销毁。您和您的孩子可以随时向研究者提问。但如果您的疑问会影响我们的实验结果，我们将在研究项目结束后再向您作答。您有权利了解此次实验的结果。此次研究项目已得到牛津大学伦理委员会的批准。

我们已附上回执单。请将填好的回执单尽快交回班主任处。非常感谢您的支持和合作！

如果您还有任何疑问，请和我联系：

电子邮件 ting.zhao@wolfson.ox.ac.uk　　电话：13581586369

或者同我的导师联系：Dr Victoria Murphy　victoria.murphy@education.ox.ac.uk.

感谢您的参与和支持！



英国牛津大学教育系在读博士

Appendix 3.5 Consent form for parents

DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0) 1865 274024 Fax: +44 (0) 1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk
Director Professor Ernesto Macaro

UNIVERSITY OF OXFORD

中国英语学习者如何加工第二语言的词汇？

- 我们已与您孩子所在的学校取得联系。学校已表示愿意参与到此次关于英语学习的调研活动中。
- 在调研期间，研究人员将定期拜访您孩子的的学校。参与者需要完成电脑生成的任务，并填写与英语词汇相关的调查问卷。
- 如果您同意您的孩子参与到此次活动中，请将填好的回执单尽快交回班主任处。

非常感谢您的支持和合作！

学生姓名: _____

班级: 5.2 初2

我已仔细阅读过此研究项目的简介，并获知此次研究项目已得到牛津大学伦理委员会的批准。我已被告知研究者将如何处理和储存此次研究收集到的数据。我已了解到是否参与此次项目是完全基于自愿的原则。参与者可以随时终止研究关系，这并不会给参与者带来任何的负面影响。

我同意我的孩子参与到此次关于英语学习的调研活动中。

家长姓名: _____

签字: _____

Appendix 3.6 Checklist of experimental procedures

ADULT VERSION

Code number: _____

No.	Experimental Procedure	Materials		Time
1	Picture naming and forward translation tasks: ABC (English-translation-Chinese)/ BCA (translation-Chinese-English) / CAB (Chinese-English-translation)	MacBook/Mouse	YES/NO	25 mins
		Audacity	YES/NO	
		SuperLab	YES/NO	
2	Subjective frequency ratings, word concreteness ratings and word typicality rating task	Paper-and-pencil task	YES/NO	40 mins
3	Vocabulary size task	IBM Computer	YES/NO	5 mins
		X_Lex v2.05	YES/NO	
4	Word associates test	IPad	YES/NO	5 mins
		Quicktap Survey		
5	Language background questionnaire	IPad	YES/NO	2-3 mins
		Quicktap Survey		

Notes:

_____

_____

_____

CHILD VERSION

Code number: _____

No.	Experimental Procedure	Materials		Time
1	Vocabulary depth task	Ipad		5 mins
		Quicktap Survey		
2	Picture naming and forward translation tasks: ABC (English-translation-Chinese)/ BCA (translation-Chinese-English) / CAB (Chinese-English-translation)	MacBook/Mouse	YES/NO	25 mins
		Audacity	YES/NO	
		SuperLab	YES/NO	
3	Vocabulary size task	IBM Computer	YES/NO	5 mins
		X_Lex v2.05	YES/NO	

Notes:

_____

_____

_____

Appendix 3.7 Rating task format: word concreteness (adult participants)

RESEARCH PROJECT:

WORD FEATURES AND L2 LEXICAL PROCESSING

WORD CONCRETENESS RATING

BOOKLET

PERSONAL INFORMATION (OPTIONAL)

Gender: Male / Female

Department: _____

Name/Initials: _____

Date: _____

INSTRUCTIONS
指 导 语

You are about to rate approximately 200 English words according to how concrete they are. You may rate them according to the degree of their involvement of five senses (i.e. by seeing, by hearing, by tasting, by smelling and by touching).

下面有大约 200 个英文单词。请对以下单词的具象程度进行评分。你可根据每个单词所引起的五种感官（视觉、听觉、味觉、嗅觉和触觉）的程度进行评分。

If you think a word can elicit a HIGH degree of imgeability, and STRONG auditory, olfactory, gustatory, and tactile senses, please circle/tick the rating **7** in the appropriate place. If you think a word does not elicit any degrees of imgeability, and auditory, olfactory, gustatory, and tactile senses, please circle/tick the rating **1** in the appropriate place.

若一个单词能呈现出明显的视觉意象，或能引起强烈的听觉，味觉，嗅觉和触觉，请在相应的位置填写 7。如果一个单词没有呈现出任何视觉意象，或无法引起听觉，味觉，嗅觉和触觉，请在相应的位置填写 1。

Remember to read each word carefully before you rate it. If you come across a word you do not know, then do not rate it. Leave the space blank on the response sheet. However, try to rate as many words as you can.

在作出判断之前，请认真阅读每个单词。若遇到您不认识的单词，请不要对其进行评分，保持其为空白即可。但是，请尽量对所有的单词进行评分。

There are no right or wrong answers. Many thanks for your participation!

答案无对错之分，请根据您自己的判断作答。谢谢参与！

	ABSTRACT					CONCRETE	
Ambulance	1	2	3	4	5	6	7
Ant	1	2	3	4	5	6	7
Apple	1	2	3	4	5	6	7
Arm	1	2	3	4	5	6	7
Arrow	1	2	3	4	5	6	7
Bag	1	2	3	4	5	6	7
Ball	1	2	3	4	5	6	7
Banana	1	2	3	4	5	6	7
Beach	1	2	3	4	5	6	7
Beans	1	2	3	4	5	6	7
Beard	1	2	3	4	5	6	7
Bee	1	2	3	4	5	6	7
Bell	1	2	3	4	5	6	7
Belt	1	2	3	4	5	6	7
Bird	1	2	3	4	5	6	7
Black	1	2	3	4	5	6	7
Blue	1	2	3	4	5	6	7
Boat	1	2	3	4	5	6	7
Bone	1	2	3	4	5	6	7
Brain	1	2	3	4	5	6	7

Appendix 3.8 Rating task format: word concreteness and subjective word frequency (child participants)

Word Rating Booklet

Name: _____ Class: _____

Age: _____ Gender: _____



例如：

词语	使用频率						消极				积极		眼睛	耳朵	鼻子	舌头	手
table	从不	偶尔	有时	经常	总是		1	2	3	4	5						

词语	使用频率						消极				积极		眼睛	耳朵	鼻子	舌头	手
peach	从不	偶尔	有时	经常	总是		1	2	3	4	5						
run	从不	偶尔	有时	经常	总是		1	2	3	4	5						
yellow	从不	偶尔	有时	经常	总是		1	2	3	4	5						
socks	从不	偶尔	有时	经常	总是		1	2	3	4	5						
elephant	从不	偶尔	有时	经常	总是		1	2	3	4	5						
sheep	从不	偶尔	有时	经常	总是		1	2	3	4	5						
eye	从不	偶尔	有时	经常	总是		1	2	3	4	5						
jump	从不	偶尔	有时	经常	总是		1	2	3	4	5						
sleep	从不	偶尔	有时	经常	总是		1	2	3	4	5						
three	从不	偶尔	有时	经常	总是		1	2	3	4	5						
词语	使用频率						消极				积极		眼睛	耳朵	鼻子	舌头	
red	从不	偶尔	有时	经常	总是		1	2	3	4	5						
leg	从不	偶尔	有时	经常	总是		1	2	3	4	5						
thirteen	从不	偶尔	有时	经常	总是		1	2	3	4	5						
finger	从不	偶尔	有时	经常	总是		1	2	3	4	5						
tomato	从不	偶尔	有时	经常	总是		1	2	3	4	5						
four	从不	偶尔	有时	经常	总是		1	2	3	4	5						
green	从不	偶尔	有时	经常	总是		1	2	3	4	5						
mouse	从不	偶尔	有时	经常	总是		1	2	3	4	5						
sweater	从不	偶尔	有时	经常	总是		1	2	3	4	5						
shirt	从不	偶尔	有时	经常	总是		1	2	3	4	5						
moon	从不	偶尔	有时	经常	总是		1	2	3	4	5						
watermelon	从不	偶尔	有时	经常	总是		1	2	3	4	5						
词语	使用频率						消极				积极		眼睛	耳朵	鼻子	舌头	手
teacher	从不	偶尔	有时	经常	总是		1	2	3	4	5						
six	从不	偶尔	有时	经常	总是		1	2	3	4	5						
baby	从不	偶尔	有时	经常	总是		1	2	3	4	5						
cat	从不	偶尔	有时	经常	总是		1	2	3	4	5						
curtain	从不	偶尔	有时	经常	总是		1	2	3	4	5						
desk	从不	偶尔	有时	经常	总是		1	2	3	4	5						
driver	从不	偶尔	有时	经常	总是		1	2	3	4	5						
flower	从不	偶尔	有时	经常	总是		1	2	3	4	5						
grass	从不	偶尔	有时	经常	总是		1	2	3	4	5						
jacket	从不	偶尔	有时	经常	总是		1	2	3	4	5						

Appendix 3.9 Rating task format: Word typicality (adult participants)

RESEARCH PROJECT: WORD FEATURES AND L2 LEXICAL PROCESSING

WORD TYPICALITY RATING

BOOKLET

PERSONAL INFORMATION (OPTIONAL)

Gender: Male / Female

Department: _____

Name/Initials: _____

Date: _____

INSTRUCTIONS

指 导 语

You are about to rate approximately 200 English words according to how typical the word represents its category.

下面有大约 200 个单词，请根据每个单词在其所属类别中的典型性程度进行评分。

If you think a word is a good example (i.e. typical) of its category, then rate the word 5. If you think a word is a bad example (i.e. atypical) of its category, then give the word a rating of 1.

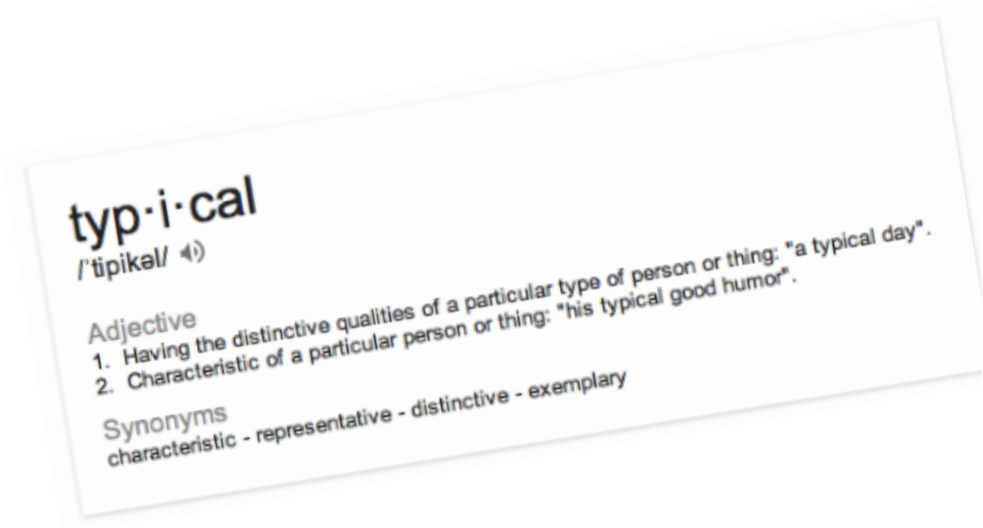
如果一个单词是其所属类别的好（典型）例子，请选 5。如果一个单词是其所属类别的坏（非典型）例子，请选 1。

Remember to read each word carefully before you rate it. If you come across a word you do not know, then do not rate it. Leave the space blank on the response sheet. However, try to rate as many words as you can.

在作出判断之前，请认真阅读每个单词。若遇到您不认识的单词，请不要对其进行评分，保持其为空白即可。但是，请尽量对所有的单词进行评分。

There are no right or wrong answers. Please write down the number in the corresponding column. Many thanks for your participation!

答案无对错之分，请根据您自己的判断作答。将单词所对应的数字写在方框内即可。谢谢参与！





Animals 动物类

1. Ant	2. Bee	3. Bird	4. Butterfly	5. Camel
6. Cow	7. Dolphin	8. Duck	9. Elephant	10. Fish
11. Fox	12. Frog	13. Kangaroo	14. Lion	15. Monkey
16. Mouse	17. Panda	18. Rabbit	19. Shark	20. Sheep
21. Tiger				

1 Atypical	2	3	4	5 Typical



Fruit and Vegetables 果蔬类

1. Apple	2. Banana	3. Beans	4. Cabbage	5. Carrot
6. Cherry	7. Corn	8. Eggplant	9. Grape	10. Mushroom
11. Onion	12. Orange	13. Peach	14. Peanut	15. Pear
16. Pepper	17. Pineapple	18. Potato	19. Strawberry	20. Tomato
21. Watermelon				

1 Atypical	2	3	4	5 Typical

Appendix 3.10 Rating task format: Word typicality (child participants)

Word Rating Booklet

Name: _____ Class: _____

Age: _____ Gender: _____





Animals 动物类

1. Donkey	2. Frog	3. Duck	4. Butterfly	5. Elephant
6. Tiger	7. Sheep	8. Lion	9. Monkey	10. Rabbit
11. Mouse	12. Panda			

1 Atypical	2	3	4	5 Typical



Fruit 水果类

1. Apple	2. Peach	3. Strawberry	4. Cabbage	5. Watermelon
6. Grape	7. Pear	8. Banana	9. Tomato	10. Potato
11. Orange	12. Bean			

1 Atypical	2	3	4	5 Typical



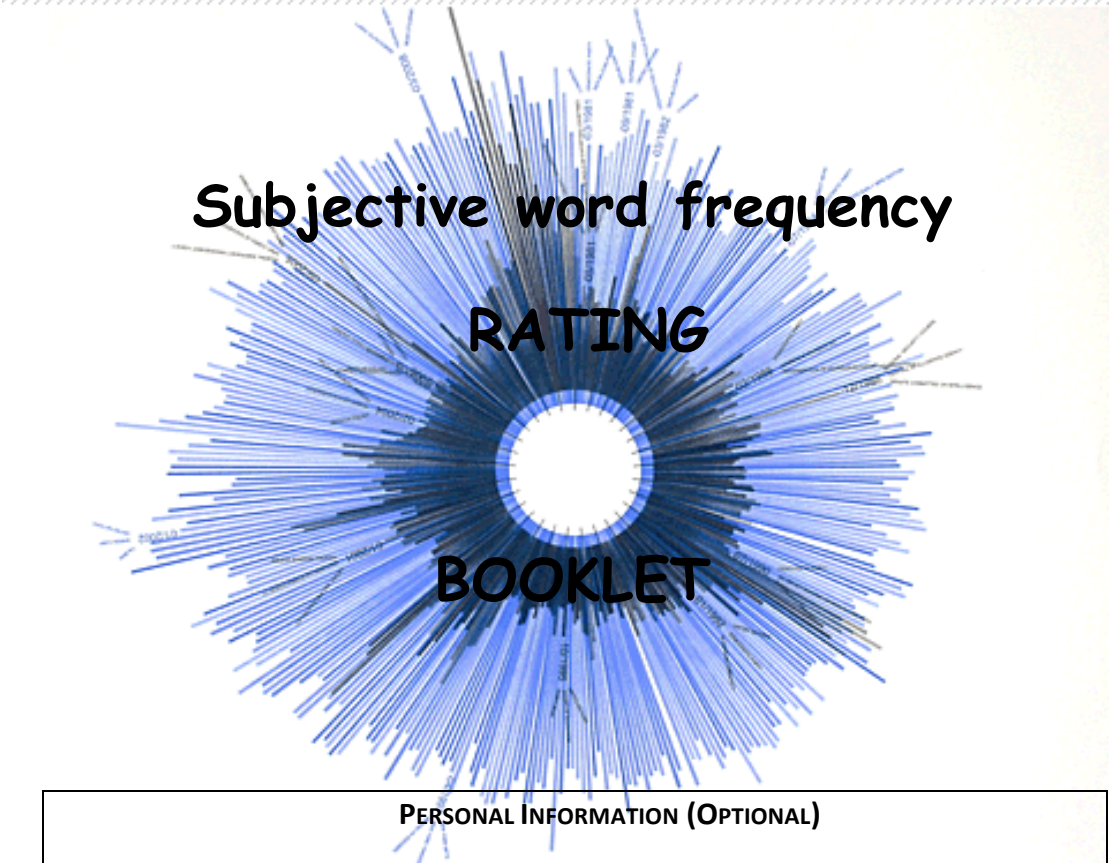
Clothing 衣服类

1. Shorts	2. Boots	3. Socks	4. Hat	5. Pants
6. Sweater	7. Shirt	8. Shoes	9. Skirt	

1 Atypical	2	3	4	5 Typical

Appendix 3.11 Rating task format: Subjective word frequency (adult participants)

RESEARCH PROJECT: WORD FEATURES AND L 2 LEXICAL PROCESSING



PERSONAL INFORMATION (OPTIONAL)

Gender: Male / Female

Department: _____

Name/Initials: _____

Date: _____

INSTRUCTIONS
指 导 语

You are about to rate approximately 200 English words according to how frequent you use the words in listening, speaking, writing, and reading.
下面有大约 200 个单词，请根据每个单词的在听、说、读、写中所使用的频率进行评分。

If you think you have always used a word, then rate the word 7. If you think you have never used a word, then give the word a rating of 1.
如果你总是使用一个单词，请选 7。如果你从不使用一个单词，请选 1。

Remember to read each word carefully before you rate it. If you come across a word you do not know, then do not rate it. Leave the space blank on the response sheet. However, try to rate as many words as you can.
在作出判断之前，请认真阅读每个单词。若遇到您不认识的单词，请不要对其进行评分，保持其为空白即可。但是，请尽量对所有的单词进行评分。

There are no right or wrong answers. Please write down the number in the corresponding column. Many thanks for your participation!
答案无对错之分，请根据您自己的判断作答。将单词所对应的数字写在方框内即可。谢谢参与！



	Never	Rarely	Occasionally	Some times	Often	Almost always	Always
Ambulance	1	2	3	4	5	6	7
Ant	1	2	3	4	5	6	7
Apple	1	2	3	4	5	6	7
Arm	1	2	3	4	5	6	7
Arrow	1	2	3	4	5	6	7
Bag	1	2	3	4	5	6	7
Ball	1	2	3	4	5	6	7
Banana	1	2	3	4	5	6	7
Beach	1	2	3	4	5	6	7
Beans	1	2	3	4	5	6	7
Beard	1	2	3	4	5	6	7
Bee	1	2	3	4	5	6	7
Bell	1	2	3	4	5	6	7
Belt	1	2	3	4	5	6	7
Bird	1	2	3	4	5	6	7
Black	1	2	3	4	5	6	7
Blue	1	2	3	4	5	6	7
Boat	1	2	3	4	5	6	7
Brain	1	2	3	4	5	6	7

Appendix 3.12 Word associates test (child participants)

每个单词下面有六个选项 a, b, c, d, e, f。请判断所提供的选项是否与目标词意义相关（相反）或者能与其互相搭配使用。

1. pencil
a. drawing
b. soup
c. clean
d. ruler
e. black
f. kitchen

2. eye
a. body
b. see
c. left
d. fan
e. like
f. fridge

3. panda
a. cinema
b. giant
c. shop
d. train
e. animal
f. China

4. vegetable
a. plants
b. photo
c. green
d. play
e. quiet
f. fresh

5. skirt
a. girl
b. grass
c. short
d. food
e. boy
f. clothing

6. river
a. nature
b. game
c. flow
d. fly
e. long
f. desk

7. spring
a. season
b. salty
c. early
d. bridge
e. read
f. warm

8. clean
a. kite
b. letter
c. fly
d. remove
e. room
f. wash

9. open
a. bottle
b. trash
c. free
d. smart
e. fruit
f. close

10. use
a. computer
b. apply
c. need
d. set
e. insect
f. sour

11. play
a. piano
b. path
c. floor
d. watch
e. joyful
f. exercise

12. buy
a. pay
b. jump
c. book
d. air
e. sell
f. summer

13. small
a. group
b. tiny
c. night
d. morning
e. tasty
f. size

14. heavy
a. rain
b. cold
c. light
d. pants
e. funny
f. weight

15. hungry
a. tall
b. mirror
c. feel
d. uncomfortable
e. yellow
f. full

16. right
a. suitable
b. blue
c. wrong
d. shelf
e. choice
f. sweet

17. old
a. house
b. age
c. bread
d. snow
e. young
f. pink

18. yummy
a. dinner
b. thin
c. delicious
d. moon
e. pleasant
f. fast

19. kind
a. helpful
b. person
c. cold
d. tired
e. winter
f. short

20. bike
a. move
b. milk
c. fork
d. sweep
e. traffic
f. ride

21. drive
a. direction
b. move
c. force
d. mouse
e. meet
f. brown

22. swim
a. sink
b. sports
c. fish
d. lamp
e. noodle
f. comic

23. easy
a. hard
b. jacket
c. job
d. juicy
e. cozy
f. garden

24. listen
a. music
b. cloud
c. park
d. ear
e. visit
f. tell

Appendix 3.13 Word associates test (adult participants)

词汇测试

Please choose three words from a-f. The chosen words must be associated with the test words in meaning or these words are often used together.

1. bell
a. device
b. cell
c. belly
d. attention
e. door
f. cut

2. dew
a. morning
b. sew
c. guideway
d. drop
e. water
f. construct

3. file
a. filet
b. store
c. away
d. mile
e. endure
f. stationery

4. fur
a. for
b. fury
c. skin
d. coat
e. mammal
f. anxious

5. graph
a. bar
b. connection
c. grab
d. telegraph
e. opera
f. outline

6. horn
a. hunt
b. car
c. press
d. glue
e. thorny
f. warn

7. ink
a. liquid
b. inn
c. drink
d. ignore
e. sign
f. black

8. liver
a. disease
b. organ
c. leaver
d. deliver
e. wreath
f. metabolic

9. lamp
a. furniture
b. shade
c. lamb
d. clamp
e. weapon
f. illuminant

10. pear
a. eat
b. tree
c. edible
d. pair
e. utter
f. pearl

11. fold
a. twofold
b. bold
c. melon
d. together
e. map
f. bend

12. enter
a. exist
b. renter
c. member
d. building
e. carpenter
f. craft

13. drag
a. effort
b. down
c. lag
d. dragon
e. odd
f. pull

14. leave
a. stop
b. live
c. cleaver
d. comfort
e. seat
f. away

15. rise
a. fall
b. smoke
c. upright
d. rice
e. prize
f. victim

16. fix
a. mix
b. suffix
c. repair
d. chair
e. secure
f. drug

17. press
a. squeeze
b. button
c. dress
d. force
e. express
f. present

18. rely
a. trust
b. relax
c. rarely
d. luggage
e. on
f. confidence

19. read
a. look
b. red
c. learn
d. staple
e. aloud
f. button

20. vote
a. presidential
b. collective
c. waiter
d. devote
e. wool
f. select

21. eager
a. keen
b. desire
c. eagle
d. meager
e. mend
f. volunteer

22. fair
a. bias
b. moderate
c. fear
d. affair
e. fan
f. competition

23. logic
a. lodge
b. dialogic
c. fund
d. reliable
e. follow
f. inference

24. neat
a. clever
b. room
c. orderly
d. knot
e. beneath
f. war

25. stable
a. economy
b. sound
c. table
d. worm
e. investable
f. firm

26. hard
a. mission
b. strict
c. complex
d. card
e. orchard
f. grape

27. mad
a. crazy
b. pad
c. made
d. cherry
e. mental
f. behavior

28. sad
a. blue
b. news
c. feeling
d. sand
e. ambassador
f. rose

29. ugly

a. heart
b. appearance
c. irritable
30. hug
d. smugly
e. enthusiasm

31. awful

a. terrible
b. manner
c. offensive
d. massy
e. earful
f. thirsty

Appendix 4.1 Characteristics of stimulus words: Child participants

Block 1

Stimulus	WC (average)	WC (total)	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
peach	2.21	11.03	3.09	2	4	5	3	1	2.97	9.5
run	1.6	8	3.94	205	22	3	3	1	4	10.5
yellow	1.47	7.33	3.8	37	3	6	4	2	2.92	9
socks	1.84	9.22	2.77	11	2	5	4	1	3.08	9.5
elephant	1.83	9.14	3.54	2	1	8	7	3	2.97	9
sheep	1.84	9.22	3.14	12	2	5	3	1	2.94	9.5
eye	1.61	8.06	4.06	6	7	3	1	1	4	9
jump	1.63	8.14	4.03	1	15	4	4	1	3.81	11.5
sleep	1.64	8.19	3.06	12	2	5	3	1	3.83	11.5
three	1.54	7.72	3.23	62	2	5	3	1	3.42	9
red	1.47	7.36	4.06	27	4	3	3	1	3.36	9
leg	1.76	8.81	3.77	3	4	3	3	1	3.33	9
thirteen	1.49	7.44	3.43	10	2	8	5	2	2.64	9.5
finger	1.78	8.92	3.43	3	4	6	5	2	3.17	9
tomato	2.24	11.22	3.43	13	2	6	6	3	3.33	10
four	1.49	7.44	3.66	34	2	4	3	1	3.33	9
green	1.48	7.42	3.91	32	6	5	4	1	3.36	9
mouse	1.88	9.42	2	2	4	5	3	1	2.78	9
shirt	1.76	8.78	3.77	1	1	5	3	1	3.14	10.5
watermelon	2.23	11.17	3.66	1	2	10	9	4	3.44	9.5
teacher	1.74	8.72	3.91	9	1	7	4	2	3.97	9.5
six	1.5	7.5	3.69	20	2	3	4	1	3	9
baby	1.84	9.22	2.91	2	11	4	4	1	2.83	10
frog	1.97	9.83	3.51	1	4	4	4	1	3.42	9
window	1.74	8.72	3.94	9	3	6	5	2	3.58	10
desk	1.8	9	3.34	2	3	4	4	1	3.67	9.5
driver	1.75	8.75	3.74	11	3.57	6	4	1	3.11	10
flower	1.99	9.94	3.71	7	4	6	4	2	3.5	11
grass	1.85	9.25	3.69	12	8	5	4	1	3.39	11
shorts	1.78	8.89	3.51	7	1	6	3	1	3.17	10.5
key	1.76	8.78	3.14	2	7	3	2	1	3.28	10.5
plate	1.74	8.72	2.97	2	7	5	4	1	2.92	10
tree	1.85	9.25	3.91	8	5	4	3	1	3.69	11
write	1.69	8.44	3.26	10	8	5	3	1	3.17	11.5
orange	1.92	9.58	3.37	5	3	6	5	2	3.28	9.5

Block 2

Stimulus	WC (average)	WC (total)	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
strawberry	1.99	9.94	3.29	1	2	10	5	1	3.88	9.5
two	1.59	7.97	3.97	254	2	3	2	1	4.57	9
mouth	1.82	9.08	3.63	5	5	5	3	1	4.58	9
doctor	1.76	8.78	3.86	123	2	6	5	2	4.4	9.5
apple	2.24	11.22	4.23	3	3	5	4	2	5.49	9.5
potato	2.11	10.53	3.6	3	3	6	6	3	4.19	10.5
purple	1.62	8.11	3.34	21	2	6	5	2	3.64	9
white	1.56	7.78	3.57	59	4	5	3	1	4.39	9
cloud	1.54	7.72	3.63	5	3	5	4	1	3.85	10.5
swim	1.83	9.17	3.5	1	6	4	4	1	4.32	11.5
foot	1.91	9.56	3.29	25	6	4	3	1	4.73	9
ear	1.83	9.17	3.63	3	4	3	2	1	4.22	9
banana	2.12	10.61	3.91	1	2	6	6	3	4.66	9.5
fourteen	1.56	7.81	3.8	10	2	8	6	2	4.24	9.5
shoes	1.69	8.47	2.94	13	4	5	3	1	4.86	10.5
blue	1.54	7.72	3.86	27	3	4	3	1	4.82	9
sing	1.69	8.44	3.37	1	2	6	4	2	4.44	11.5
snow	1.71	8.56	3.4	5	5	4	3	1	4.49	10.5
tiger	1.82	9.11	4.03	3	3	5	4	2	3.42	10.5
panda	1.83	9.14	4.31	1	2	5	5	1	4.23	9
five	1.6	8	3.77	26	2	4	3	1	4.91	9
skirt	1.74	8.69	3.86	5	3	5	4	1	5.65	10.5
chair	1.77	8.83	4.86	8	5	5	3	1	4.53	9.5
computer	1.86	9.31	3.86	4	2	8	7	3	4.33	10
lion	1.88	9.42	4.3	5	3	4	4	2	3.6	9
boots	1.74	8.69	3.86	16	9	5	3	1	4.86	10.5
drink	1.85	9.25	3.6	5	12	5	4	1	4.28	11.5
eat	1.92	9.61	3.8	7	7	3	2	1	4.48	11
fan	1.83	9.14	3.74	1	8	3	3	1	3.95	10
girl	1.73	8.67	3.57	279	3	4	3	1	4.2	9.5
leaf	1.87	9.33	2.46	2	7	4	3	1	3.85	11.5
seven	1.55	7.75	3.91	21	3	5	4	2	4.59	9
spoon	1.87	9.33	2.57	2	4	5	4	1	3.21	10
student	1.79	8.97	3.66	6	2	7	7	2	5.65	9.5
donkey	1.86	9.28	3.43	1	3	6	5	2	3.58	10.5

Block 3

Stimulus	WC (average)	WC (total)	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
arm	1.73	8.64	3.71	11	3	3	3	1	3.42	9
grape	2	10	3.26	1	3	5	4	1	3.17	9.5
sun	1.56	7.81	3.94	8	5	3	3	1	3.36	10.5
butterfly	1.75	8.75	2.69	1	3	9	7	3	3.47	11.5
farmer	1.66	8.31	3.63	1	1	6	5	2	3.17	10
pink	1.57	7.86	3.49	26	3	4	4	1	3.56	9
nineteen	1.58	7.89	3.8	6	2	8	6	2	2.81	9.5
monkey	1.74	8.69	3.17	1	2	6	5	2	3.36	9
pants	1.73	8.64	3.43	6	2	5	5	1	3.42	10.5
dance	1.72	8.58	3.4	5	5	5	4	1	2.86	11.5
fridge	1.71	8.56	4.43	2	1	6	4	1	2.69	10
nose	1.76	8.81	4.06	3	4	4	3	1	3.56	9
black	1.72	8.58	3.29	60	4	5	4	1	3.53	9
ten	1.56	7.81	4.23	8	2	3	3	1	3.61	9
brown	1.57	7.86	3.69	34	3	5	4	1	3.08	9
eight	1.59	7.97	3.83	30	2	5	2	1	3.31	9
clean	1.73	8.64	3.6	2	3	5	4	1	3.69	10
cabbage	2.11	10.53	3.54	10	2	7	5	2	3.53	11
pear	1.96	9.81	3.49	1	2	4	3	1	3.19	9.5
boy	1.73	8.64	3.54	326	3	3	2	1	3.56	9.5
bean	2.11	10.53	2.54	5	4	5	4	1	3.53	11
cup	1.83	9.14	4.12	7	7	3	3	1	3.68	10
nurse	1.65	8.25	3.49	4	3	5	3	1	3.67	10
floor	1.72	8.61	2.43	17	9	5	3	1	3.61	10
hand	1.76	8.81	3.97	72	17	4	4	1	3.61	9
sweater	1.76	8.81	3.51	6	2	7	5	2	3.53	10.5
mountain	1.69	8.44	3.66	13	2	8	6	2	3.19	11
nine	1.64	8.22	4	18	2	4	3	1	3.31	9
phone	1.76	8.78	3.77	3	4	5	3	1	3.47	10
duck	1.82	9.08	2.86	3	3	4	3	1	3.19	9
rabbit	1.77	8.83	3.69	5	2	6	5	2	3.5	9
read	1.59	7.97	3.74	1	14	4	3	1	3.83	11
river	1.83	9.14	3.77	15	2	5	4	2	3.22	11
sit	1.64	8.19	2.89	17	8	3	3	1	3.11	11.5
hat	1.75	8.75	2.91	4	2	3	3	1	3.28	10.5

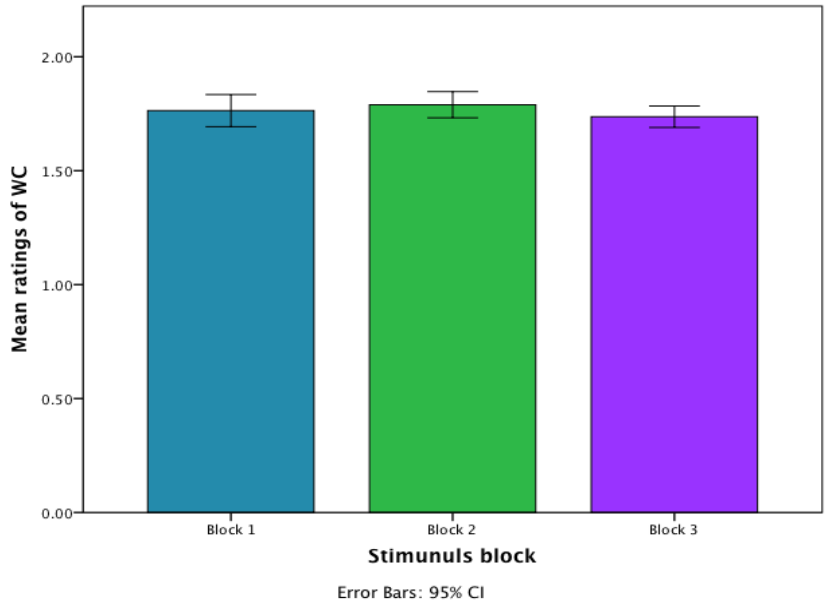
Appendix 4.2 Comparing word features across blocks: Child participants

Word concreteness (WC)

Since both the kurtosis and skewness values fell between +2.00 and -2.00 within each block, the departure from a normal distribution was considered non-significant (see Table 10 below). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .12 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of WC, F (2, 102) = .83, p = .44 η_p^2 = .02. Descriptive statistics of WC ratings across blocks are presented in Table 1. Results of the comparison between blocks in terms of the WC are illustrated in Figure 1.

Table 1 Descriptive statistics of WC in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	1.76	.20	1.47	2.24
Block 2	35	1.79	.17	1.54	2.24
Block 3	35	1.74	.14	1.56	2.11



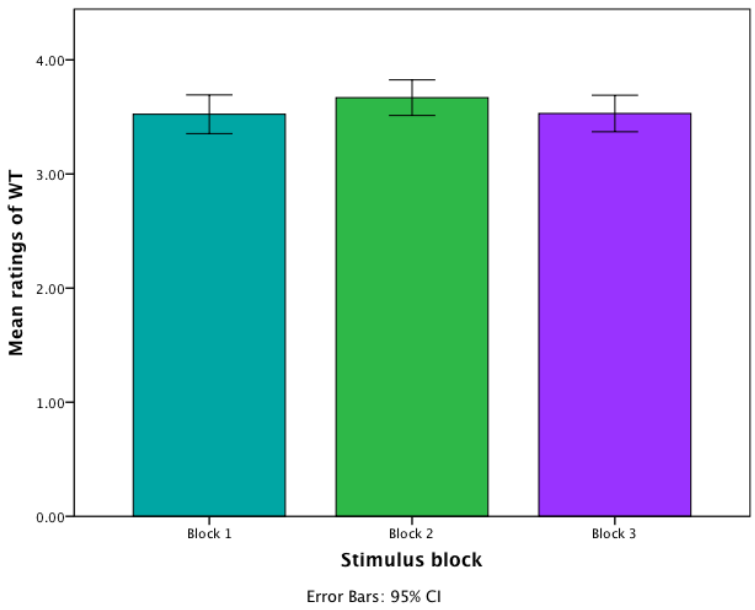
Note: Error bars represent the 95% CI.
Figure 1 Mean ratings of WC across three blocks

Word typicality (WT)

Both kurtosis and skewness values indicated that there was no serious departure from normality (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .78 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the WT, F (2, 102) = 1.06, p = .35, η_p^2 = .02. Descriptive statistics of WT ratings across blocks are presented in Table 2. Results of the comparison between blocks in terms of the WT are illustrated in Figure 2.

Table 2 Descriptive statistics of WT in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	3.52	.49	2.00	4.54
Block 2	35	3.67	.45	2.46	4.86
Block 3	35	3.53	.47	2.43	4.43



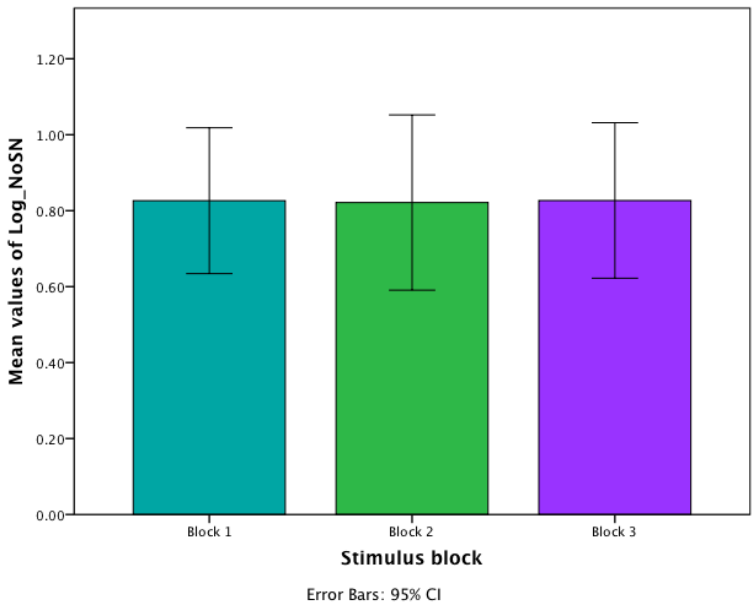
Note: Error bars represent the 95% CI.
Figure 2 Mean ratings of WT across three blocks

Log-transformed semantic neighborhood density (Log-SND)

The normality check suggested that the variable SND was not normally distributed in all blocks so a log-transformation was applied to this variable. After this, the distribution of the Log-SND was approximately normal within each block (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .66 (see Table 11 below for details). The ANOVA result indicated that there was no statistically significant difference between the blocks being compared in terms of the Log-SND, F (2, 102) = .001, p = .99, η_p^2 = .000. Descriptive statistics of the Log-SND across blocks are presented in Table 3. Results of the comparison between blocks in terms of the Log-SND are illustrated in Figure 3.

Table 3 Descriptive statistics of Log-SND in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	.85	.57	0.00	2.31
Block 2	35	.86	.67	0.00	2.45
Block 3	35	.83	.60	0.00	2.51



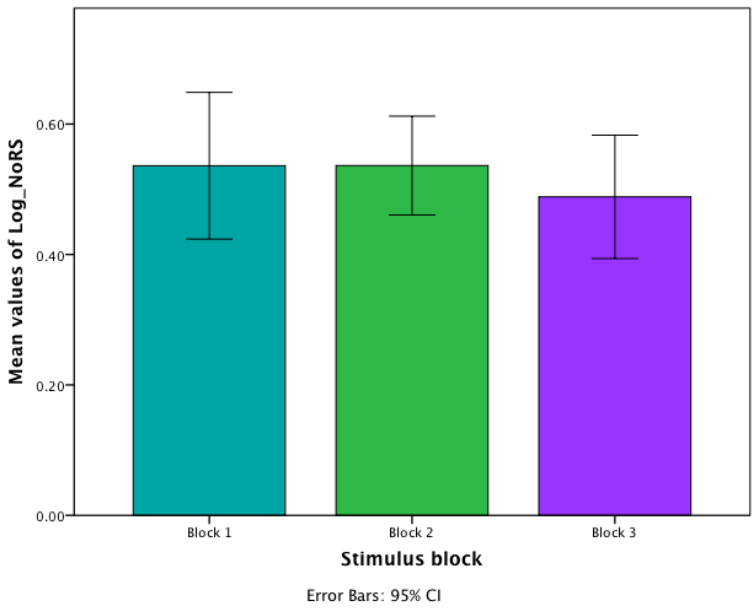
Note: Error bars represent the 95% CI.
Figure 3 Mean values of Log-SND across three blocks

Log-transformed number of related senses (Log-NoRS)

The distribution of NoRS was found to significantly violate the assumption of normality, in which case a logarithmic transformation was applied to correct for its non-normality. After the transformation, the distribution was approximately normal within each block (see Table 10 below for details). The Levene's test result suggested that the homogeneity of variance was not violated, p = .17 (see Table 11 below for details). The ANOVA result indicated that there was no statistically significant difference between blocks in terms of the Log-NoRS, F (2, 102) = .35, p = .71, η_p^2 = .01. Descriptive statistics of the Log-NoRS across blocks are presented in Table 4. Results of the comparison between blocks in terms of the Log-NoRS are illustrated in Figure 4.

Table 4 Descriptive statistics of Log-NoRS in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	.54	.34	0.00	1.34
Block 2	35	.55	.22	.30	1.08
Block 3	35	.49	.28	0.00	1.23



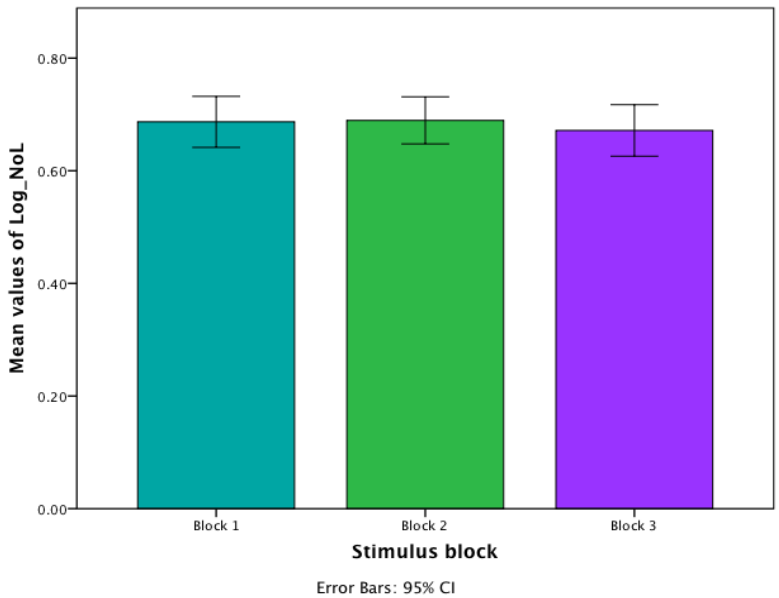
Note: Error bars represent the 95% CI.
Figure 4 Mean values of Log-NoRS across three blocks

Log-transformed number of letters (Log-NoL)

Skewness and kurtosis values showed that the variable NoL was not normally distributed within two blocks. Therefore, a logarithmic transformation was applied to correct for its non-normality (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .66 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between the blocks being compared in terms of Log-NoL, F (2, 195) = .20, p = .82, η_p^2 = .004. Descriptive statistics of the Log-NoL across blocks are presented in Table 5. Results of the comparison between blocks in terms of the NoL are illustrated in Figure 5.

Table 5 Descriptive statistics of Log-NoL in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	.68	.12	.48	.90
Block 2	35	.69	.12	.48	1.00
Block 3	35	.67	.13	.48	.95



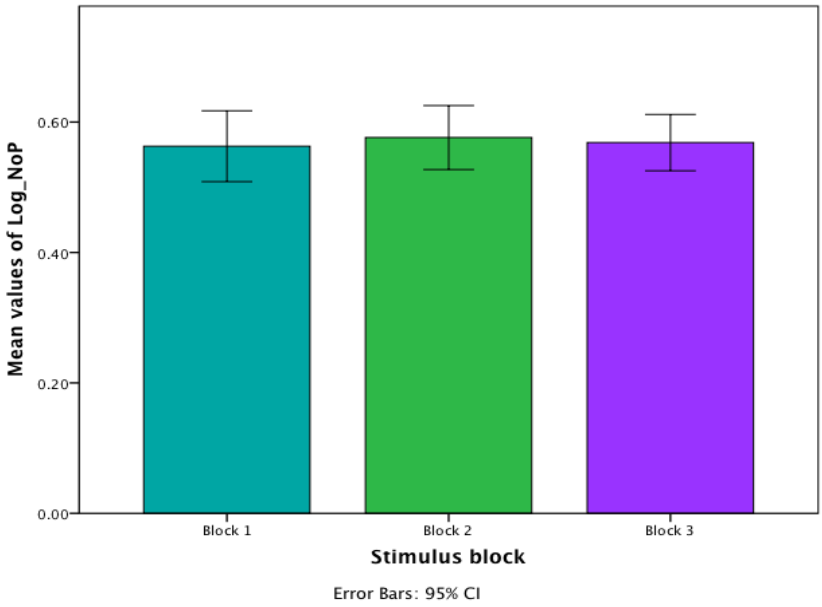
Note: Error bars represent the 95% CI.
Figure 5 Mean values of Log-NoL across three blocks

Log-transformed number of phonemes (Log-NoP)

Skewness and kurtosis values showed that the variable NoP was not normally distributed within one block. Therefore, a logarithmic transformation was applied to correct for its non-normality (see Table 10 below for details). The Levene's test result suggested that the homogeneity of variance was not violated, p = .86 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the Log-NoP, F (2, 102) = .08, p = .93, η_p^2 = .001. Descriptive statistics of the Log-NoP across blocks are presented in Table 6. Results of the comparison between blocks are illustrated in Figure 6.

Table 6 Descriptive statistics of Log-NoP in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	.56	.10	.30	.85
Block 2	35	.57	.14	.30	.85
Block 3	35	.57	.13	.30	.85



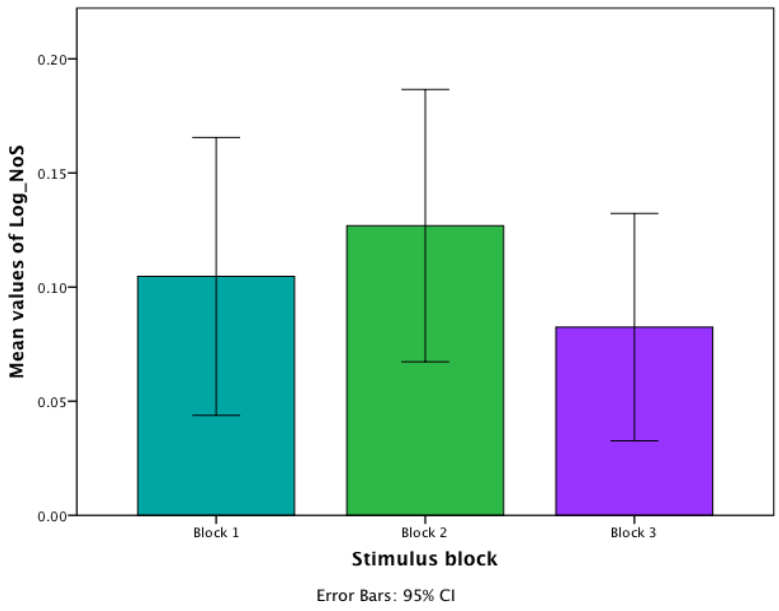
Note: Error bars represent the 95% CI.
Figure 6 Mean values of Log-NoP across three blocks

Log-transformed number of syllables (Log-NoS)

Skewness and kurtosis values revealed that the assumption of normality was not met, so a logarithmic transformation was applied to correct for its non-normality (see Table 10 below for details). The Levene's test result suggested that the homogeneity of variance was not violated, p = .12 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the Log-NoS, F (2, 102) = .63, p = .54, η_p^2 = .01. Descriptive statistics of the NoS across blocks are presented in Table 7. Results of the comparison between blocks in terms of the NoS are illustrated in Figure 7.

Table 7 Descriptive statistics of Log-NoS in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	.07	.14	0.00	.48
Block 2	35	.12	.17	0.00	.48
Block 3	35	.08	.14	0.00	.48



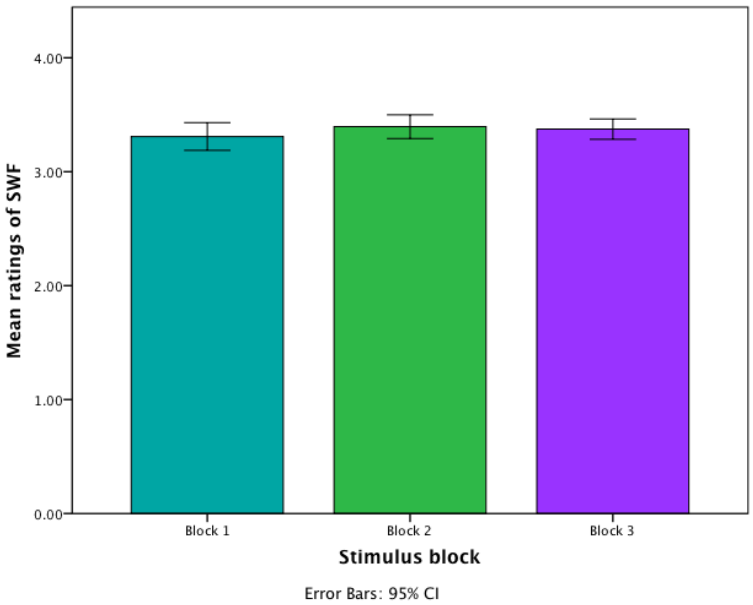
Note: Error bars represent the 95% CI.
Figure 7 Mean values of Log-NoS across three blocks

Subjective word frequency (SWF)

The distribution of SWF was found to be normal within each block (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .28 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the SWF, F (2, 102)= .75, p = .48, η_p^2 = .01. Descriptive statistics of the SWF across blocks are presented in Table 8. Results of the comparison between blocks are illustrated in Figure 8.

Table 8 Descriptive statistics of SWF in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	3.31	.35	2.64	4.00
Block 2	35	3.40	.30	2.94	4.22
Block 3	35	3.37	.26	2.69	3.83



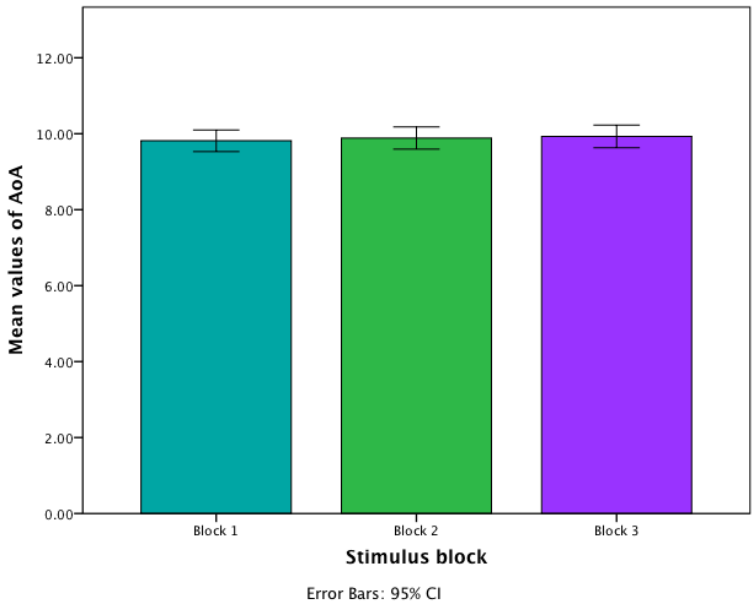
Note: Error bars represent the 95% CI.
Figure 8 Mean ratings of SWF across three blocks

Age of acquisition (AoA)

The normality check showed that the AoA was normally distributed (see Table 10 below for details). The Levene's test result showed that the homogeneity of variance was not violated, p = .94 (see Table 11 for details). The ANOVA result suggested that there was no statistically significant difference between the blocks being compared in terms of the AoA, F (2, 102)= .16, p = .85, η_p^2 = .003. Descriptive statistics of the AoA across blocks are presented in Table 9. Results of the comparison between blocks are illustrated in Figure 10.

Table 9 Descriptive statistics of AoA in each block

	n	Mean	SD.	Minimum	Maximum
Block 1	35	9.81	.83	9.00	11.50
Block 2	35	9.89	.85	9.00	11.50
Block 3	35	9.93	.87	9.00	11.50



Note: Error bars represent the 95% CI.
Figure 9 Mean ratings of AoA across three blocks

Table 10 Skewness and Kurtosis values

Variable		Skewness	Kurtosis
WC (average)	Block 1	.63	.44
	Block 2	.58	.61
	Block 3	1.22	1.71
WT	Block 1	-.54	1.66
	Block 2	-.39	2.05
	Block 3	-.62	.21
Log-SND	Block 1	.42	-.13
	Block 2	.74	.18
	Block 3	.60	.54
Log-NoRS	Block 1	.29	-.05
	Block 2	.63	-.48
	Block 3	.92	1.11
Log-NoL	Block 1	-.19	-.39
	Block 2	.30	.37
	Block 3	.12	-.51
Log-NoP	Block 1	.33	1.37
	Block 2	.09	-.15
	Block 3	.07	-.10
Log-NoS	Block 1	1.58	1.03
	Block 2	.82	-.96
	Block 3	1.34	.19
SWF	Block 1	.35	-.45
	Block 2	.72	.14
	Block 3	-.79	.42
AoA	Block 1	.76	-.62
	Block 2	.61	-.81
	Block 3	.41	-1.15

Table 11 Test of Homogeneity of Variances

Variable	Levene statistic	df1	df2	Sig.
WC	2.21	2	102	.12
WT	.25	2	102	.78
Log-SND	.42	2	102	.66
Log-NoRS	1.82	2	102	.17
Log-NoL	.42	2	102	.66
Log-NoP	.15	2	102	.86
Log-NoS	2.15	2	102	.12
SWF	1.29	2	102	.28
AoA	.06	2	102	.94

Appendix 4.3 Assumptions of SEM: Do Meaning and Form predict Usage among child L2 learners?

Preliminary analysis. In order to examine the tenability of assumptions required for the statistical analyses, descriptive statistics for all the variables were examined, including their means, standard deviations, and frequency histograms.

Sample size and missing data. As regards the sample size, according to Bentler and Chou's recommendation (1987), under normal distribution, a ratio of 5 cases or observations per free parameter would be necessary, with a higher ratio (10:1) recommended. The hypothesized model had 27 parameters to be estimated; thus, the rule '5 cases per parameter' indicated that a minimum sample size of 135 cases was needed to assess the hypothesized model. This requirement, however, was not met since the sample size was 104 (one multivariate outlier removed) in the present study. Given this, composite variables were constructed afterwards so as to reduce the complexity of the hypothesized model and to decrease the number of stimulus words required for a reliable model. There were no missing data identified in the data set.

Outliers. There are two types of outliers: univariate and multivariate outliers. A univariate outlier is any data point with an extreme or unusual value on a single variable, while multivariate outliers are cases with an unusual combination of scores on two or more variables (Tabachnick & Fidell, 2012). One way of testing univariate outliers is to convert raw scores to standardized scores and then to examine the maximum and minimum scores against conventional standards based on a sample size. According to Stevens (2009), conventionally, the z values for a sample size larger than 100 should be less than 4.00 and more than -4.00, and the percentage of observations beyond 3 standard deviations should be less than 1. Following these guidelines, there was no case falling outside the tolerance limits. To assess multivariate outliers, the Mahalanobis distance, a relative measure of data point's distance from the group centroid under the hypothesis of normality, was calculated (B. M. Byrne, 2009). One multivariate outlier (line 105) was identified and removed from

further analysis ($p < .001$).

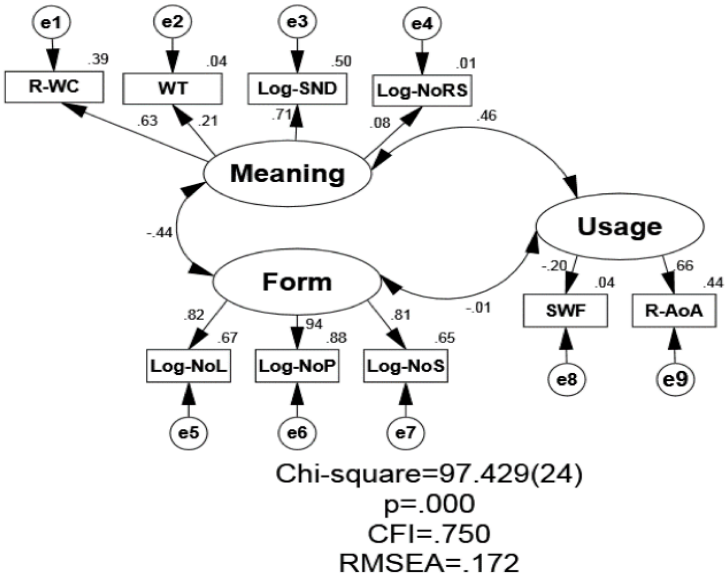
Normality and linearity. One of the important distribution requirements for conducting SEM is the assumption of normality, including univariate normality and multivariate normality. This requirement is crucial because, as DeCarlo (1997) pointed out, skewness tends to affect tests of means, and kurtosis tends to impact upon tests of variances and covariances. Univariate normality of the observed variables was assessed through examining skewness and kurtosis values. None of the raw and transformed values was found to be beyond the recommended limits mentioned earlier (+/-2.00). Moreover, because SEM is based on the analysis of covariance structure, we should be more concerned about evidence of kurtosis, in particular multivariate kurtosis (B. M. Byrne, 2011). According to the guidelines provided by B. M. Byrne (2011), the kurtosis value that is equal to or greater than 7 indicates severe departure from normality. A review of the kurtosis values suggested that no variable to be substantially kurtotic. The univariate normality is considered as a necessary but not sufficient condition for securing multivariate normal distributions in variables (B. M. Byrne, 2009). Multivariate normality is usually assessed by Mardia's (1970, 1974) normalized estimate of multivariate kurtosis, and the Mardia's statistic that exceeds 5.00 is believed to be indicative of non-normally distributed data (Bentler, 1995). If the assumption of multivariate normality is violated, bootstrapping will be used. The Bollen-Stine corrected *p* value rather than the usual maximum likelihood-based *p* value will be used to assess the overall model fit as well as the significance level of all regression coefficients. As to the current analysis, the z-statistic of 1.36 indicated that the assumption of multivariate normality was not violated.

Linearity was assessed using pairwise scatter plots, and it was found that all observed variables appeared to be linearly related.

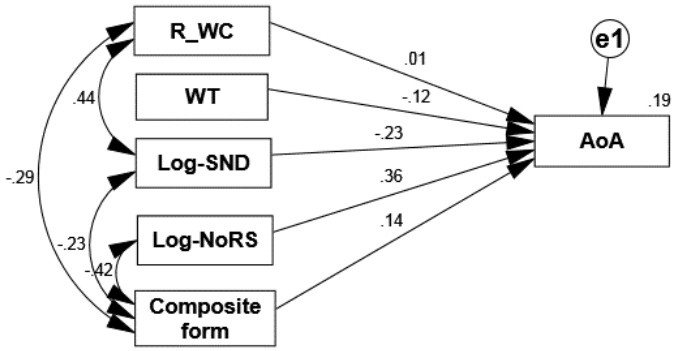
Multicollinearity and singularity. Multicollinearity is a high correlation between two or more independents (.90 and above), and singularity is a perfect or nearly

perfect correlation between two or more independents. According to Tabachnick and Fidell (2012), highly correlated independents can lead to a failure of achieving statistical significance of regression coefficients, and a failure of model converge on a solution. To examine for multicollinearity and singularity, a scan of the correlation matrix of the indicants showed that none of the correlation coefficients exceeded 0.90, indicating the absence of multicollinearity and linearity.

Appendix 4.4 Measurement model with standardized coefficients: Meaning, Form, and Usage of L2 words among child L2 learners (IBM SPSS AMOS output)



Appendix 4.5 Path model with standardized coefficients: Meaning, form and usage of L2 words among child L2 learners (IBM SPSS AMOS output)



Chi-square=5.665(6)
p=.462
CFI=1.000
RMSEA=.000

Appendix 4.6 Results of associations between nominal learner-related variables and the response latency of PE and that of FT among child L2 learners (n = 39)

A. PE and learner-related nominal variables

Variables	Lambda value	Approx. Sig.
Gender	.00	1.00
Living area	.03	.31

B. FT and learner-related nominal variables

Variables	Lambda value	Approx. Sig.
Gender	.00	1.00
Living area	.03	.31

Appendix 4.7 Assumptions of MR: Predictive effect of learner-related variables on PE and FT response latencies among child L2 learners (n = 39)

Sample size and missing data. For the present study, with a ratio of 10 observations to each predictor, which is a necessary but not desirable condition, the necessary sample size is 20 for predicting the PE and 30 for predicting the FT. The number of cases in the present study (n = 39) met the requirements of sample size. There were no missing data observed.

Normality, linearity, and homoscedasticity of residuals. Analysis of the scatterplot showed that all residuals closely formed a rectangular pattern, indicating that the above assumptions were met (see Figure 1 for examples).

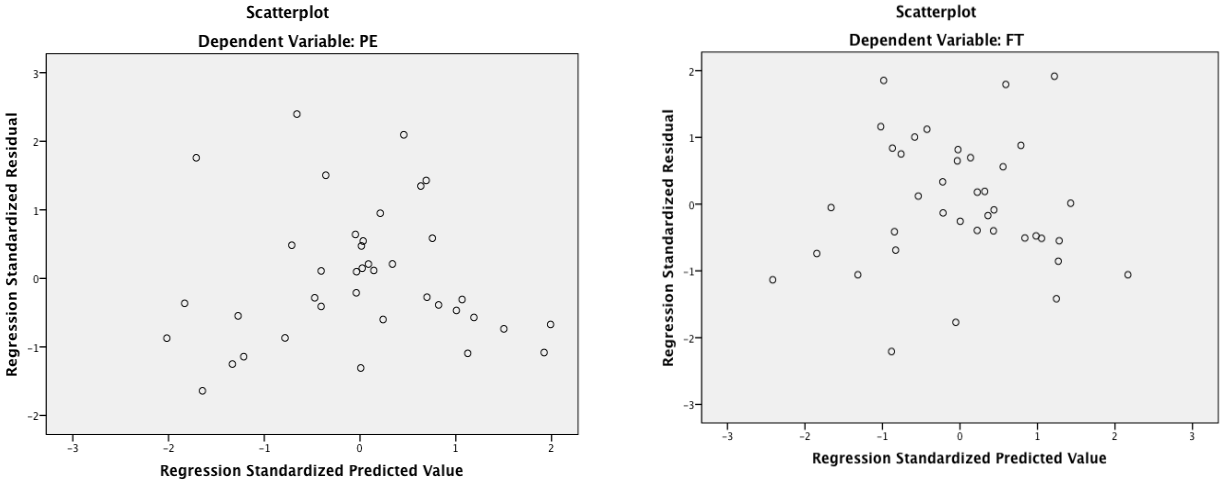


Figure 1 Residual scatterplot (Output from IBM SPSS REGRESSION.)

Outliers. Univariate outliers were checked using standardized scores, which showed that all values were within the limits (+/-3.00 SD). Multivariate outliers were checked by examining the Mahalanobis distance of each case to the centroid of all cases. Mahalanobis distance is distributed as a chi-square variable, with degrees of freedom equal to the number of IVs. The results showed that none of the cases exceeded the value of 13.82 (critical χ^2 at $\alpha = 0.001$ for 3 dfs) in the model predicting PE response latencies, and the value of 16.27 (critical χ^2 at $\alpha = 0.001$

for 3 dfs) in the model predicting FT response latencies. Therefore, no multivariate outliers were observed in the data set.

Multicollinearity and singularity. None of the tolerances approached zero, and all the variance inflation factor (VIF) values were not significantly far from 1 (see Table 1 below), so there was no inappropriate level of multicollinearity and singularity.

Table 1 Collinearity statistics (Tolerance and VIF)

Model predicting PE response latencies

Variables	Tolerance	VIF
Vocabulary size	.944	1.060
Chinese course scores	.944	1.060

Model predicting FT response latencies

Variables	Tolerance	VIF
Vocabulary size	.614	1.627
Reading proficiency level	.592	1.688
Listening proficiency level	.509	1.966

Appendix 4.8 Assumptions of SEM model: Predictive effect of word feature-related variables on PE and FT response latencies

Predictive effect of word feature-related variables on PE response latencies

Identifiability of the model. As a preliminary check of the identifiability of the model, the number of data points and parameters to be estimated was counted. With 7 variables, there were 7 (7+1)/2 = 28 data points. The hypothesized model contained 22 parameters to be estimated; therefore, the model was over-identified and tested with 6 dfs.

Sample size and missing data. According to the minimum sample size requirement (cf. Appendix 4.3), it was considered acceptable to conduct the path analysis with the obtained size of 104 lexical items as this size was very close to the minimum requirement although the desirable size was not reached. There were no missing data in the data set.

Normality and linearity. The distributions of all lexical variables as well as log-transformed variables were normally distributed. As for PE response latencies, its distribution was normal (skewness value = .13, and kurtosis value = -.42). Multivariate normality was checked by Mardia's statistic (value = .77), which was well below the threshold of 5.00; therefore, multivariate normality was met in the current analysis.

As Tabachnick and Fidell (2012) pointed out, it is not feasible to examine all pairwise scatterplots to assess linearity; so pairs of scatterplots were randomly examined using IBM SPSS GRAPHS. All observed pairs appeared to be linearly related, if at all.

Outliers. There were no univariate outliers for word feature-related variables, and there were no univariate outliers for PE response latencies (within +/-3.00 SD). Multivariate outliers were checked using the Mahalanobis distance, and there were no multivariate outliers observed.

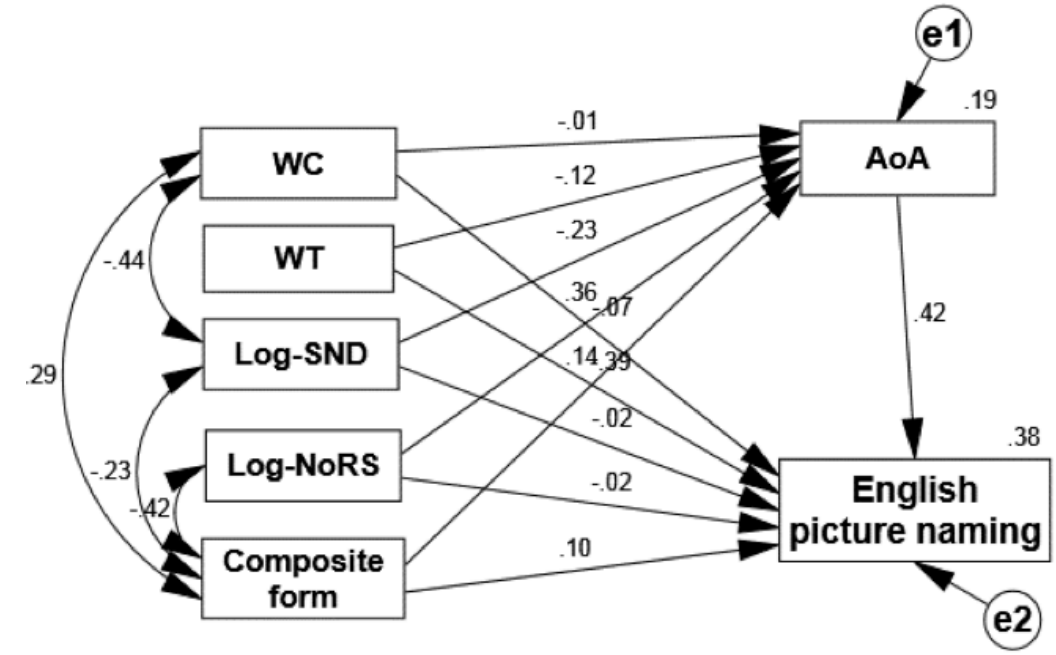
Multicollinearity and singularity. To examine for multicollinearity and

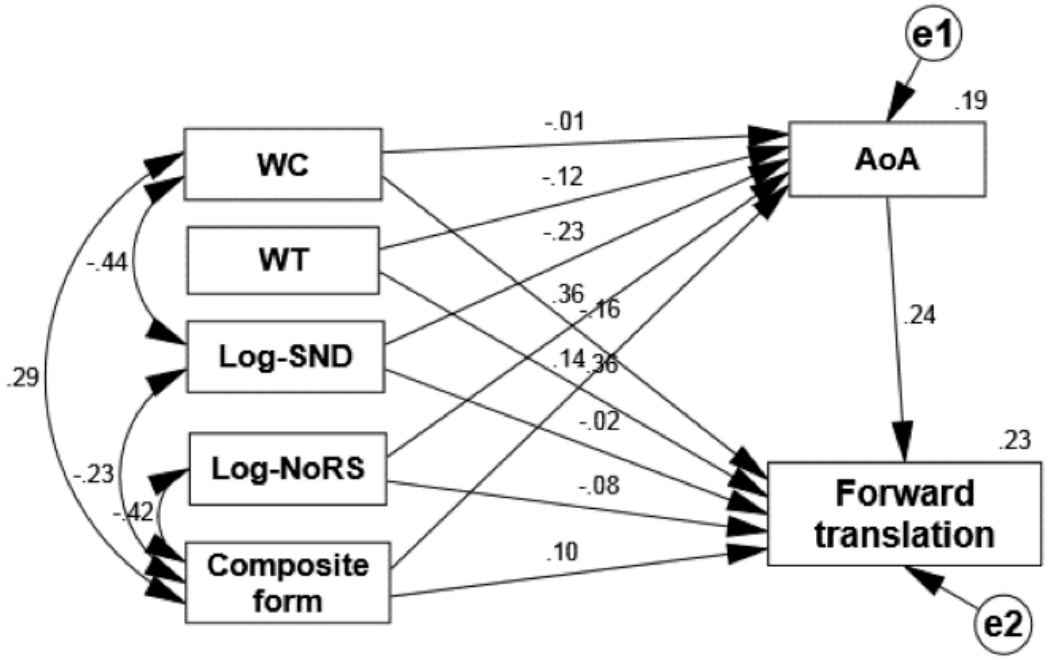
singularity, a check on the correlation matrix of indicants showed that none of the correlation coefficient exceeded 0.90 (cf. Appendix 4.3), indicating the absence of multicollinearity and linearity.

Predictive effect of word feature-related variables on FT response latencies

All the assumptions were checked in a similar way in which the PE model was evaluated. The results suggested that none of the assumptions was seriously violated in terms of sample size, missing data, normality, linearity, outliers, multicollinearity and singularity.

Appendix 4.9 Path models with standardized coefficients: Predictive effects of word feature-related variables on PE and FT response latencies (IBM SPSS AMOS output)





Appendix 4.10 Results of associations between nominal learner-related variables and PF-FT response latency differences among child L2 learners (n = 39)

Variables	Lambda value	Approx. Sig.
Gender	.00	1.00
Living areas	.03	.31

Appendix 4.11 Assumptions of MR: Predictive effects of learner-related and word feature-related variables on PE-FT RL differences among child L2 learners

LEARNERS-RELATED VARIABLES

Sample size and missing data. As for the current analysis, with a recommended ratio of 20 observations to each predictor, the necessary sample size was estimated to be 20 for assessing the effect of academic achievements on the PE-FT response latency difference. The number of cases in the present study (n = 39) was considered sufficient. There were no missing data.

Normality, linearity, and homoscedasticity of residuals. Univariate normality was checked using skewness and kurtosis values. The univariate normality of the predictor variable had been checked and confirmed in Section 4.3.1. As to the outcome variable, its skewness value was .46, and kurtosis value was .99, which suggested the normality of its distribution. Analysis of the scatterplot showed that all residuals closely formed a rectangular pattern, indicating that the assumption of linearity and homoscedasticity of residuals were not violated.

Outliers. Univariate outliers were checked using standardized scores, which showed that all values were within +/-2.00 SD. Multivariate outliers were checked by examining the Mahalanobis distance of each case to the centroid of all cases. The results showed that none of the cases exceeded the value of 10.83 (critical χ^2 at $\alpha = .001$ for 1 dfs). Therefore, there were no multivariate outliers identified in the data set.

Multicollinearity and singularity. The assumption of multicollinearity and singularity were checked in the same way as described in Section 4.3.1.

WORD FEATURE-RELATED VARIABLES

Sample size and missing data. As for the current analysis, with a recommended ratio of 20 observations to each predictor, the necessary sample size was estimated to be 20

for predicting the PE-FT response latency difference. The number of cases in the present study (n = 105) was considered sufficient. There were no missing data.

Normality, linearity, and homoscedasticity of residuals. Univariate normality was checked using skewness and kurtosis values. The univariate normality of predictor variables had been checked and confirmed in Section 4.3.1. As to the outcome variable, its skewness value was -.40, and kurtosis value was -.12, which suggested the normality of its distribution. Analysis of the scatterplot showed that all residuals closely formed a rectangular pattern, indicating that the assumption of linearity and homoscedasticity of residuals were not violated.

Outliers. Univariate outliers had been checked using standardized scores (cf. Section 5.3.3). Values of the outcome variable were within +/-3.00 SD limits. Multivariate outliers were checked by examining the Mahalanobis distance of each case to the centroid of all cases. The results showed that none of the cases exceeded the value of 10.83 (critical $\chi^2\ at\ \alpha = .001$ for 1 dfs). Therefore, there were no multivariate outliers in the data set.

Multicollinearity and singularity. The assumption of multicollinearity and singularity were checked in the same way as described in Section 4.3.1.

Appendix 5.1 Characteristics of stimulus words: Adult participants

Block 1

Stimulus	WC	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
arm	5.46	4.35	11	3	3	3	1	3.82	9
bag	5.46	2.74	2	6	3	3	1	4.69	9
balloon	5.45	2.28	2	8	7	5	2	3.12	17
banana	5.74	4.53	1	2	6	6	3	4.66	9
bell	4.11	2.3	3	3	4	3	1	3.12	14
black	3.05	4.26	60	4	5	4	1	4.09	9
brown	2.74	3.32	34	3	5	4	1	3.49	9
butterfly	5.4	2.69	1	3	9	7	3	3.37	11
car	5.6	4.56	7	3	3	3	1	4.77	9
cherry	5.12	3.59	3	5	6	4	2	2.95	17
clean	2.66	3.45	2	3	5	4	1	4.97	11
corn	4.98	2.59	9	4	4	4	1	2.97	12
cube	3.75	3.1	1	3	4	4	1	2.69	17
dentist	4.49	3.72	1	1	7	7	2	3.2	14
doctor	5.15	4.57	123	2	6	5	2	4.4	10
drive	3.66	3.11	28	9	5	4	1	4.54	10
duck	5.43	3.35	3	3	4	3	1	3.29	10
ear	5.49	4.39	3	4	3	2	1	4.22	9
eyeglasses	5.32	2.28	1	1	10	7	3	4.03	13
fight	3.02	2.98	6	5	5	3	1	3.97	11
finger	5.38	4.18	3	4	6	5	2	4.02	9
fish	5.29	3.12	1	2	4	3	1	4.26	11
hat	5.4	3.33	4	2	3	3	1	3.75	12
island	4.18	3.19	7	3	6	5	2	3.62	13
lightening	3.98	3.29	1	1	10	6	1	3.29	15
lion	5.57	4.07	5	3	4	4	2	3.6	12
monkey	5.55	4.53	1	2	6	5	2	3.68	9
mushroom	5.22	2.59	1	2	8	6	2	2.72	15
nail	5.09	2.14	1	2	4	3	1	3	17
nineteen	3.28	3.08	6	2	8	6	2	4.09	9
nose	5.42	4.27	3	4	4	3	1	4.17	9
nurse	4.98	4.07	4	3	5	3	1	3.52	9
onion	5.37	2.84	33	2	5	5	2	3.23	12
panda	5.71	4.13	1	2	5	5	1	4.23	9
pear	5.25	4.03	1	2	4	3	1	3.78	9
pink	3.06	4.04	26	3	4	4	1	3.43	9

Stimulus	WC	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
plane	5.35	4.4	2	3	5	4	1	3.95	9
read	3.08	4.15	1	14	4	3	1	5.37	10
run	4.65	4.44	205	22	3	3	1	4.85	11
sea	4.85	4.26	12	5	3	2	1	4.92	13
seventy	3.29	2.88	1	4	7	7	3	3.82	10
ship	5.02	4.05	5	2	4	3	1	3.75	12
shirt	5.09	4.48	1	1	5	3	1	4.42	10
shoes	5.55	3.44	13	4	5	3	1	4.86	10
shout	3.38	2.86	8	2	5	3	1	3.95	13
sixty	3.6	3.51	14	4	5	6	2	4.12	9
skirt	5.06	4.51	5	3	5	4	1	5.65	10
snow	4.65	4.26	5	5	4	3	1	4.49	10
soldier	5.2	3.72	1	2	7	5	2	3.52	11
square	3.58	3.87	1	7	6	8	1	3.37	14
stomach	4.91	3.22	1	4	7	6	2	4.06	16
stone	5	2.73	13	7	5	4	1	3.86	15
strawberry	5.51	4.4	1	2	10	5	1	3.88	9
swim	4.05	4.05	1	6	4	4	1	4.32	11
teacher	5.02	4.8	9	1	7	4	2	5.69	9
teeth	5.55	3.26	1	4	5	3	1	3.83	13
ten	3.14	4.2	8	2	3	3	1	4.77	9
tiger	5.57	4.28	3	3	5	4	2	3.42	13
truck	5.22	3.46	1	4	5	4	1	3.15	14
twenty	3.38	3.79	29	4	6	6	2	4.22	9
two	3.82	4.35	254	2	3	2	1	4.57	9
volcano	4.58	2.95	3	2	7	7	3	2.55	15
waiter	5.12	3.64	1	2	6	4	2	4.03	13
wallet	5.23	2.12	2	1	6	5	2	4.03	13
wave	3.83	2.96	3	7	4	3	1	3.15	14
zero	3.35	3.66	1	4	4	4	2	4.49	9

Block 2

Stimulus	WC	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
arrow	3.83	2.25	3	3	5	3	2	2.83	17
beans	5.18	2.48	5	4	5	4	1	3.2	10
bee	5.18	2.86	4	1	3	2	1	3.32	13
belt	4.86	2.63	2	5	4	4	1	3.14	16
button	5.11	2.22	1	4	6	5	2	3.21	14
camel	5.44	3.41	1	1	5	5	2	2.58	16
carriage	4.55	2.52	4	5	8	5	2	2.53	17
cave	4.8	2.63	2	2	4	3	1	2.79	15
clap	3.2	3.47	3	7	4	4	1	2.89	16
climb	3.03	3.8	5	5	5	4	1	4.14	13
cloud	4.52	3.9	5	3	5	4	1	3.85	11
coat	5.24	4.46	10	4	4	3	1	4.33	12
triangle	4.11	2.59	2	2	8	6	2	2.91	13
desert	3.83	2.93	3	3	6	5	2	3.12	14
earth	3.83	3.67	4	4	5	2	1	4.73	14
eggplant	4.97	2.78	2	1	8	7	1	2.71	9
eight	3.17	4.33	30	2	5	2	1	4.59	9
eighteen	2.91	3.43	12	2	8	4	2	4.23	9
eighty	3.56	3.14	9	4	6	3	2	3.85	9
elephant	5.26	4.39	2	1	8	7	3	3.41	9
elevator	5.23	2.99	1	3	8	7	4	3.88	14
eye	5.26	4.46	6	7	3	1	1	5.02	9
fall	2.83	2.84	45	10	4	3	1	4.23	17
flood	4.21	2.59	1	4	5	4	1	3.52	14
four	2.97	4.11	34	2	4	3	1	4.55	9
frog	5.33	2.76	1	4	4	4	1	2.91	17
grey	2.7	3.31	19	3	4	3	1	3.29	12
hundred	3.02	3.58	14	3	8	7	2	4.74	9
jump	3.09	4.41	1	15	4	4	1	4.35	10
knock	3.11	3.24	1	6	5	3	1	3.94	13
lawyer	4.68	4.2	2	1	6	4	2	3.48	16
leg	4.97	4.17	3	4	3	3	1	4.23	9
lung	5.02	3.08	7	1	4	3	1	3.18	17
moon	4.76	4.36	12	4	4	3	1	4.61	12
mouse	5.42	3.12	2	4	5	3	1	4.02	9
mouth	5.35	4.2	5	5	5	3	1	4.58	9
muscle	4.38	2.53	3	3	6	5	2	3.67	15

Stimulus	WC	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
neck	5.08	3.81	5	3	4	3	1	3.58	9
orange	5.38	4.48	5	3	6	5	2	4.42	9
peach	5.56	4.06	2	4	5	3	1	3.12	9
peanut	5.24	2.17	8	2	6	5	2	3.12	17
pineapple	5.29	3.55	1	2	9	7	3	3.55	15
policeman	5.12	4.16	1	1	9	7	3	4.18	11
postman	4.89	3.57	1	1	7	7	2	3.14	12
raincoat	5.35	2.77	1	1	8	6	1	3.18	12
red	2.3	4.8	27	4	3	3	1	4.79	9
reporter	4.76	4.27	2	3	5	7	3	3.53	12
shark	5.53	3.17	18	1	5	4	1	2.98	14
sheep	4.85	3.92	12	2	5	3	1	3.55	12
shoulder	4.65	3.86	5	4	8	5	2	4.03	17
singer	4.82	3.4	1	2	6	4	2	4.44	11
six	2.74	4.37	20	2	3	4	1	4.41	9
socks	5.44	3.1	11	2	5	4	1	3.7	9
star	4.05	3.95	3	6	4	4	1	4.62	13
student	5.32	3.12	6	2	7	7	2	5.65	9
submarine	4.91	2.79	1	2	9	8	3	2.83	16
subway	5.39	4.63	1	3	6	5	2	4.53	15
tomato	5.39	3.7	13	2	6	6	3	4.18	9
train	4.97	4.45	5	5	5	4	1	4.59	11
twelve	2.91	3.74	18	2	6	5	1	4.44	9
type	3.41	3.01	6	2	4	3	1	3.48	14
umbrella	5.41	1.84	1	2	8	7	3	3.92	13
watch	4.68	2.56	34	3	5	3	1	4.68	12
white	2.58	4.34	59	4	5	3	1	4.39	9
write	3.02	4.32	10	8	5	3	1	5.14	10
yellow	2.53	4.61	37	3	6	4	2	4.3	9

Block 3

Stimulus	WC	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
ant	5.36	2.53	1	1	3	3	1	3.58	8
apple	5.43	4.64	3	3	5	4	2	5.49	8
ball	4.93	3.05	2	5	4	3	1	4.73	12
beach	4.28	3.49	3	1	5	3	1	4.13	14
beard	4.99	2.24	11	2	5	4	1	3.78	17
bird	5.18	3.36	5	3	4	3	1	4.91	16
blue	2.72	4.74	27	3	4	3	1	4.82	8
boat	4.69	3.63	14	2	4	3	1	4.04	11
bone	4.67	2.82	4	4	4	3	1	3.55	17
brain	3.37	3.64	1	3	5	4	1	4.57	17
cabbage	5.16	3.28	10	2	7	5	2	3.63	13
carrot	5.19	3.01	2	3	6	5	2	3.57	12
chef	4.52	3.64	2	1	4	3	1	2.84	16
circle	3.03	4.09	1	5	6	5	2	4.4	13
cook	3.64	2.99	6	6	4	3	1	4.75	11
cow	5.1	3.77	4	2	3	2	1	3.66	13
cry	3.46	3.09	10	6	3	3	1	4.58	13
dance	3.3	4.22	5	5	5	4	1	4.63	11
dive	3.16	3.02	1	7	4	3	1	3.06	13
dolphin	5.19	3.05	3	2	7	6	2	3.28	13
eleven	2.64	3.85	13	2	6	6	3	4.36	8
farmer	4.7	3.38	1	1	6	5	2	3.97	10
five	3.18	4.47	26	2	4	3	1	4.91	8
foot	5.16	3.99	25	6	4	3	1	4.73	8
forest	4.45	4.6	10	2	6	6	2	4.63	13
fourteen	2.25	3.06	10	2	8	6	2	4.24	8
fox	5.42	3.53	4	4	3	4	1	3.06	13
gloves	5.21	2.98	3	3	6	5	1	3.28	12
grape	5.46	4.16	1	3	5	4	1	3.64	13
green	2.25	4.54	32	6	5	4	1	5.09	8
hair	5.28	3.2	20	4	4	3	1	4.63	12
heart	4	3.8	24	8	5	4	1	4.96	12
hunter	4.1	3.15	2	2	6	5	2	3.07	16
kangaroo	5.3	3.56	1	1	8	7	2	3.04	10
lake	4.7	4.49	6	1	4	3	1	4.45	11
line	2.97	3.49	16	14	4	3	1	4.64	13
miner	4.21	2.85	1	1	5	4	2	2.78	16

Stimulus	WC	WT	SND	NoRS	NoL	NoP	NoS	SWF	AoA
mountain	4.79	4.25	13	2	8	6	2	4.69	13
necklace	4.85	2.9	4	1	8	6	2	3.19	14
one	3	4.61	649	4	3	3	1	5.75	8
painter	4.34	3.38	4	2	7	5	2	3.39	13
pants	5.25	4.24	6	2	5	5	1	3.91	9
pepper	5.01	2.35	15	5	6	4	2	3	13
potato	5.27	3.53	3	3	6	6	3	4.19	10
purple	2.54	3.72	21	2	6	5	2	3.64	8
rabbit	5.19	3.93	5	2	6	5	2	3.85	8
rainbow	3.99	3.49	1	2	7	5	2	4.16	17
ring	4.63	2.64	2	9	4	3	1	4.13	14
river	4.94	4.61	15	2	5	4	2	4.76	8
rocket	4.88	2.12	1	3	6	5	2	3.82	14
sailor	4.45	2.79	1	3	6	4	2	3.24	17
satellite	4.93	2.02	16	4	9	7	3	3.03	16
shoot	2.69	3.11	1	6	5	3	1	3.87	14
shower	3.7	2.94	5	5	6	3	2	4.76	15
sixteen	2.73	3.39	10	2	7	7	2	4.36	8
skateboard	4.96	2.28	1	1	10	6	1	3.01	15
sleep	3.39	3.48	13	5	5	4	1	5.73	10
suit	5.3	4.12	15	5	4	3	1	3.78	14
sun	4.75	4.42	8	5	3	3	1	5.16	11
sweater	4.96	4.26	6	2	7	5	2	3.99	12
thirteen	2.51	3.1	10	2	8	5	2	4.13	8
three	3.01	4.51	62	2	5	3	1	5.31	8
tie	4.99	2.99	4	7	3	2	1	3.42	13
tongue	4.73	3.2	3	6	6	3	1	3.66	15
watermelon	5.31	4.28	1	2	10	9	4	4.37	8
wheelchair	5.39	1.86	1	1	10	7	1	3.21	17

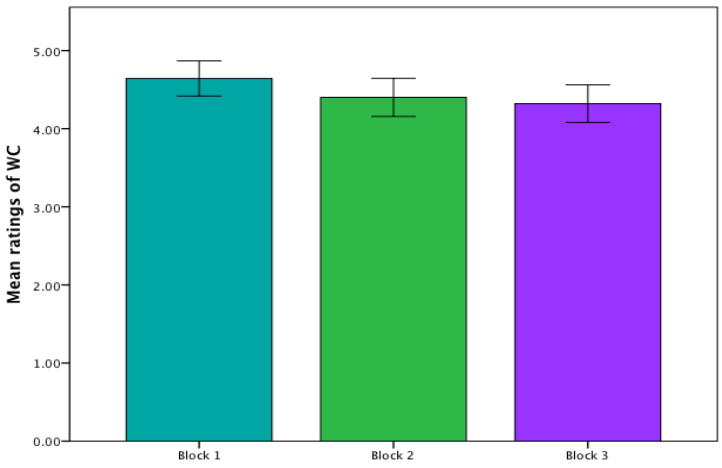
Appendix 5.2 Comparing word features across blocks: Adult participants

Word concreteness (WC)

Since both the kurtosis and skewness values fell between +2.00 and -2.00 within each block, the departure from a normal distribution was considered not very significant (see Table 10 below). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .67 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of WC, F (2, 195) = 2.00, p = .14, η_p^2 = .02. Descriptive statistics of WC ratings across blocks are presented in Table 1. Results of the comparison between blocks in terms of the WC are illustrated in Figure 1.

Table 1 Descriptive statistics of WC in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	2.66	5.74	4.64	.92
Block 2	66	2.30	5.56	4.40	.99
Block 3	66	2.25	5.46	4.32	.98



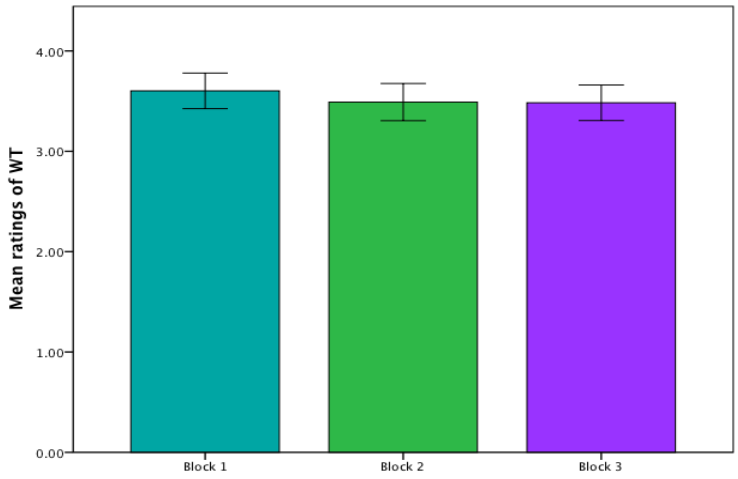
Note: Error bars represent the 95% CI.
Figure 1 Mean ratings of WC across three blocks

Word typicality (WT)

Both kurtosis and skewness values indicated that there was no serious departure from normality (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .63 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the WT, F (2, 195) = .55, p = .58, η_p^2 = .006. Descriptive statistics of WT ratings across blocks are presented in Table 2. Results of the comparison between blocks in terms of the WT are illustrated in Figure 2.

Table 2 Descriptive statistics of WT in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	2.12	4.80	3.60	.72
Block 2	66	1.84	4.80	3.49	.75
Block 3	66	1.86	4.74	3.48	.72



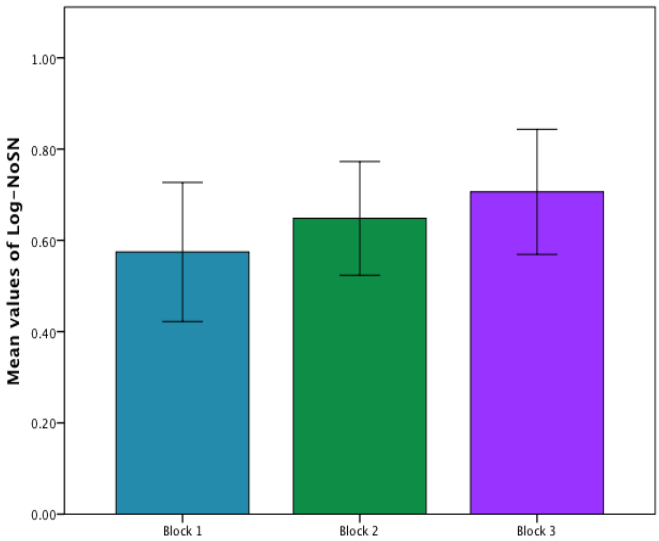
Note: Error bars represent the 95% CI.
Figure 2 Mean ratings of WT across three blocks

Log-transformed semantic neighborhood density (Log-SND)

The Log-SND values were derived from WordMine 2 (Durda & Buchanan, 2006). The normality check suggested that the variable SND was not normally distributed so a log-transformation was applied to this variable. After this, the distribution of the Log-SND was approximately normal within each block (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .49 (see Table 11 below for details). The ANOVA result indicated that there was no statistically significant difference between the blocks being compared in terms of the Log-SND, F (2, 195) = .85, p = .43, η_p^2 = .01. Descriptive statistics of the Log-SND across blocks are presented in Table 3. Results of the comparison between blocks in terms of the Log-SND are illustrated in Figure 3.

Table 3 Descriptive statistics of Log-SND in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	0	2.40	.58	.62
Block 2	66	0	1.77	.65	.51
Block 3	66	0	2.81	.71	.56



Note: Error bars

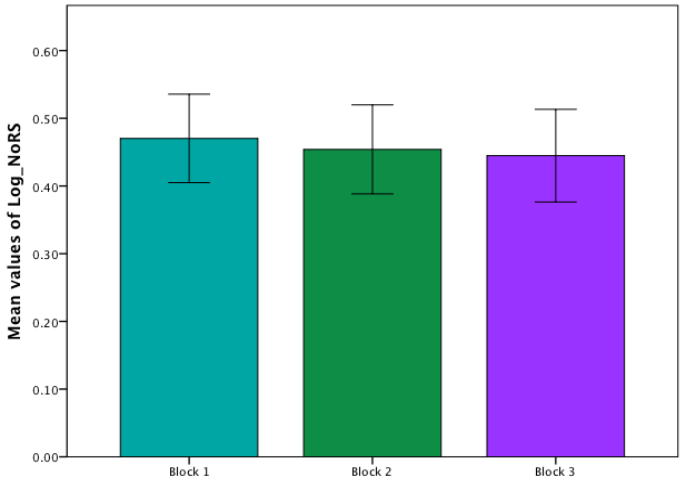
Figure 3 Mean values of Log-SND across three blocks

Log-transformed number of related senses (Log-NoRS)

The distribution of NoRS was found to significantly violate the assumption of normality, in which case a logarithmic transformation was applied to correct for its non-normality. After the transformation, the distribution was approximately normal within each block (see Table 10 below for details). The Levene's test result suggested that the homogeneity of variance was not violated, p = .48 (see Table 11 below for details). The ANOVA result indicated that there was no statistically significant difference between blocks in terms of the Log-NoRS, F (2, 195) = .43, p = .65, η_p^2 = .004. Descriptive statistics of the Log-NoRS across blocks are presented in Table 4. Results of the comparison between blocks in terms of the Log-NoRS are illustrated in Figure 4.

Table 4 Descriptive statistics of log-transformed number of related senses in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	0	1.34	.48	.26
Block 2	66	0	1.18	.45	.27
Block 3	66	0	1.15	.44	.28



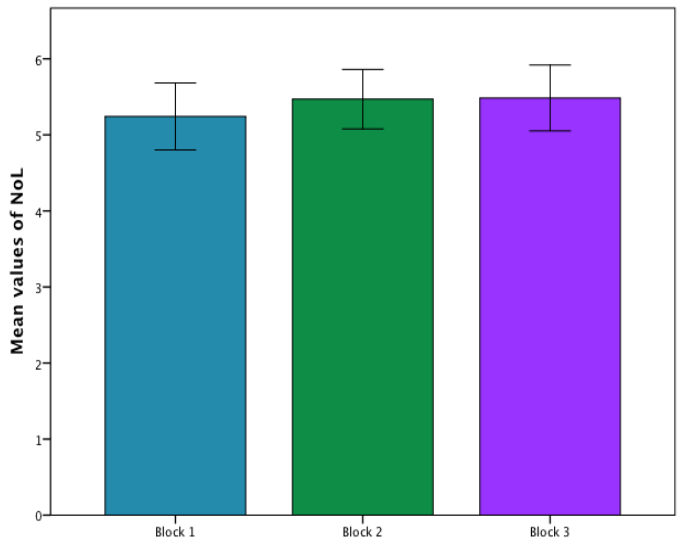
Note: Error bars represent the 95% CI.
Figure 4 Mean values of Log-NoRS across three blocks

Number of letters (NoL)

Skewness and kurtosis values showed that the variable NoL was normally distributed within each block (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .89 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between the blocks being compared in terms of NoL, F (2, 195) = .56, p = .57, η_p^2 = .01. Descriptive statistics of the NoL across blocks are presented in Table 5. Results of the comparison between blocks in terms of the NoL are illustrated in Figure 5.

Table 5 Descriptive statistics of NoL in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	3	10	5.23	1.73
Block 2	66	3	9	5.52	1.64
Block 3	66	3	10	5.48	1.76



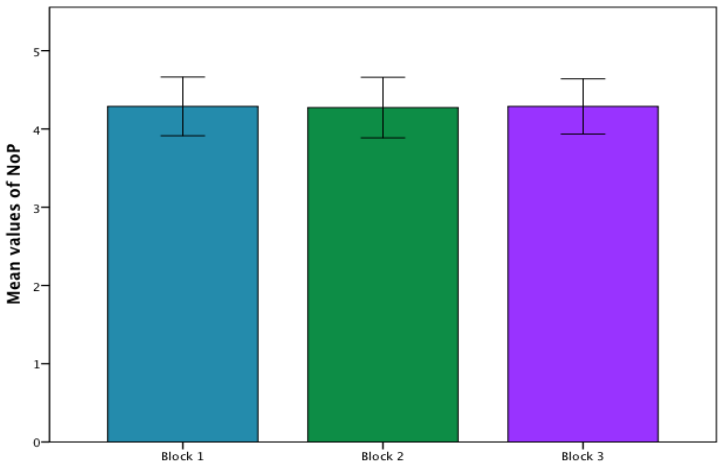
Note: Error bars represent the 95% CI.
Figure 5 Mean values of NoL across three blocks

Number of phonemes (NoP)

The normality check showed that skewness and kurtosis values were both less than 1, so the assumption of normality was met (see Table 10 below for details). The Levene's test result suggested that the homogeneity of variance was not violated, p = .57 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the NoP, F (2, 195) = .05, p = .95, η_p^2 = .000. Descriptive statistics of the NoP across blocks are presented in Table 6. Results of the comparison between blocks are illustrated in Figure 6.

Table 6 Descriptive statistics of number of phonemes in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	2	8	4.22	1.42
Block 2	66	1	8	4.30	1.59
Block 3	66	2	9	4.29	1.43



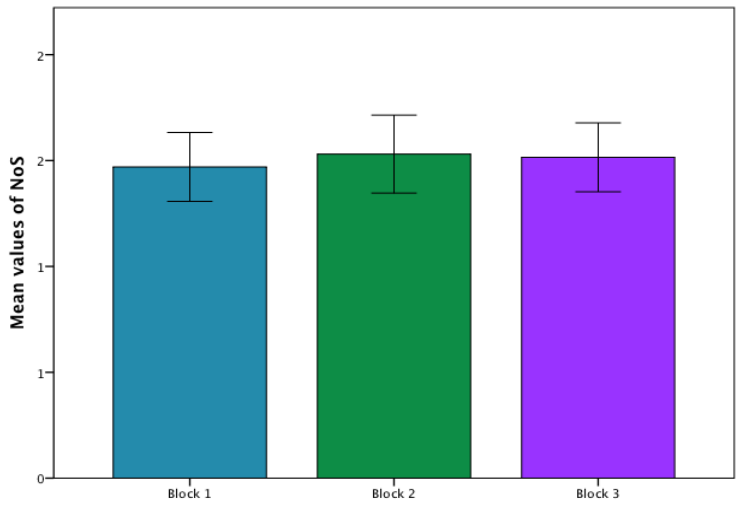
Note: Error bars represent the 95% CI.
Figure 6 Mean values of NoP across three blocks

Number of syllables (NoS)

Skewness and kurtosis values revealed that the assumption of normality was met (see Table 10 below for details). The Levene's test result suggested that the homogeneity of variance was not violated, p = .32 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the NoS, F (2, 195) = .30, p = .74, η_p^2 = .003. Descriptive statistics of the NoS across blocks are presented in Table 7. Results of the comparison between blocks in terms of the NoS are illustrated in Figure 7.

Table 7 Descriptive statistics of NoS in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	1	3	1.45	.64
Block 2	66	1	4	1.55	.75
Block 3	66	1	4	1.52	.66



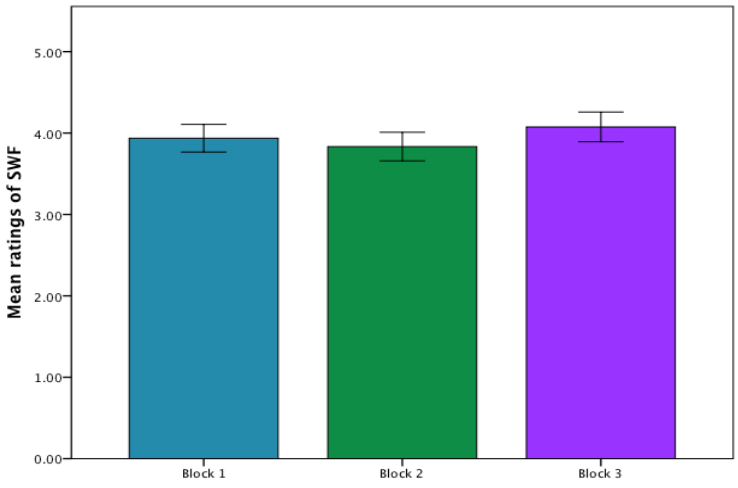
Note: Error bars represent the 95% CI.
Figure 7 Mean values of NoS across three blocks

Subjective word frequency (SWF)

The distribution of SWF was found to be normal within each block (see Table 10 below for details). The Levene's test result suggested that the assumption of homogeneity of variance was not violated, p = .44 (see Table 11 below for details). The ANOVA result suggested that there was no statistically significant difference between blocks in terms of the SWF, F (2, 195)= 1.88, p = .16, η_p^2 = .02. Descriptive statistics of the SWF across blocks are presented in Table 8. Results of the comparison between blocks are illustrated in Figure 8.

Table 8 Descriptive statistics of SWF in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	2.55	5.69	3.94	.69
Block 2	66	2.53	5.65	3.83	.72
Block 3	66	2.78	5.75	4.08	.75



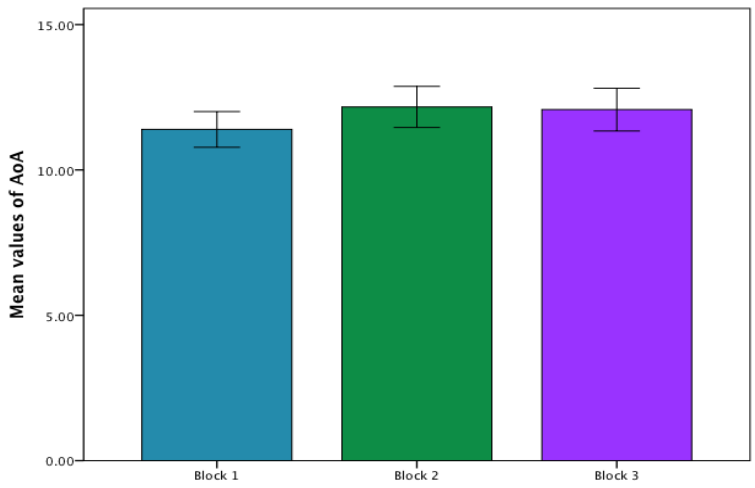
Note: Error bars represent the 95% CI.
Figure 8 Mean ratings of SWF across three blocks

Age of acquisition (AoA)

The normality check showed that the AoA was normally distributed (see Table 10 below for details). The Levene's test result showed that the homogeneity of variance was not violated, p = .25 (see Table 11 for details). The ANOVA result suggested that there was no statistically significant difference between the blocks being compared in terms of the AoA, F (2, 195)= 1.51, p = .22, η_p^2 = .02. Descriptive statistics of the AoA across blocks are presented in Table 9. Results of the comparison between blocks are illustrated in Figure 9.

Table 9 Descriptive statistics of AoA in each block

	n	Minimum	Maximum	Mean	SD.
Block 1	66	9	17	11.39	2.50
Block 2	66	9	17	12.17	2.87
Block 3	66	8	17	12.08	2.98



Note: Error bars represent the 95% CI.
Figure 9 Mean ratings of AoA across three blocks

Table 10 Skewness and Kurtosis values

Variable		Skewness	Kurtosis
WC	Block 1	-.74	-.93
	Block 2	-.68	-1.05
	Block 3	-.72	-.89
WT	Block 1	-.34	-.98
	Block 2	-.14	-1.10
	Block 3	-.11	-.64
Log-SND	Block 1	1.10	.80
	Block 2	.35	-.85
	Block 3	.76	1.59
Log-NoRS	Block 1	.48	1.15
	Block 2	.02	-.03
	Block 3	.10	-.57
NoL	Block 1	1.02	.97
	Block 2	.57	-.55
	Block 3	.76	.29
NoP	Block 1	.73	-.20
	Block 2	.58	-.47
	Block 3	.83	.47
NoS	Block 1	1.10	.14
	Block 2	1.21	.73
	Block 3	1.26	1.84
SWF	Block 1	.32	-.06
	Block 2	.10	-.81
	Block 3	.19	-.73
AoA	Block 1	.79	-.48
	Block 2	.32	-1.28
	Block 3	-.004	-1.11

Table 11 Test of Homogeneity of Variances

Variable	Levene statistic	df1	df2	Sig.
WC	.40	2	195	.67
WT	.46	2	195	.63
Log-SND	.76	2	195	.49
Log-NoRS	.74	2	195	.48
NoL	.12	2	195	.89
NoP	.56	2	195	.57
NoS	1.14	2	195	.32
SWF	.83	2	195	.44
AoA	1.39	2	195	.25

Appendix 5.3 Assumptions of SEM: Do Meaning and Form predict Usage among adult L2 learners?

Preliminary analysis. In order to examine the tenability of assumptions required for the statistical analyses, descriptive statistics for all the variables were examined, including their means, standard deviations, and frequency histograms.

Sample size and missing data. As regards the sample size, according to Bentler and Chou's recommendation (1987), under normal distribution, a ratio of 5 cases or observations per free parameter would be necessary, with a higher ratio (10:1) recommended. The hypothesized model had 27 parameters to be estimated; thus, the rule '5 cases per parameter' indicated that a minimum sample size of 135 cases was needed to assess the hypothesized model; while the rule '10 cases per parameter' suggested that the minimum sample size was 270. Although the larger sample size with 10:1 ratio would increase the power of analysis, it was considered sufficient to conduct SEM with the obtained sample size of 198. There were no missing data in the data set.

Normality and linearity. One of the important distribution requirements for conducting SEM is the assumption of normality, including univariate normality and multivariate normality. This requirement is crucial because, as DeCarlo (1997) pointed out, skewness tends to affect tests of means, and kurtosis tends to impact upon tests of variances and covariances. Univariate normality of the observed variables was assessed through examining skewness and kurtosis values. None of the raw and transformed values was found to be beyond the recommended limits mentioned earlier (+/-2.00). Moreover, the critical ratio (C.R. = statistic/standard error) should be within the recommended limits +/-3.00 (Tabachnick & Fidell, 2012). It was found that the critical ratios of the following three indicators, NoL, NoP and NoS were beyond the range, so log-transformations were applied to normalize them.

The univariate normality is considered as a necessary but not sufficient condition for securing multivariate normal distributions in variables (B. M. Byrne, 2009).

Multivariate normality is usually assessed by Mardia's (1970, 1974) normalized estimate of multivariate kurtosis, and the Mardia's statistic that exceeds 5.00 is believed to be indicative of non-normally distributed data (Bentler, 1995). If the assumption of multivariate normality is violated, bootstrapping will be used. The Bollen-Stine corrected *p* value rather than the usual maximum likelihood-based *p* value will be used to assess the overall model fit as well as the significance level of all regression coefficients. As to the current analysis, the z-statistic of 1.51 indicated that the assumption of multivariate normality was not violated.

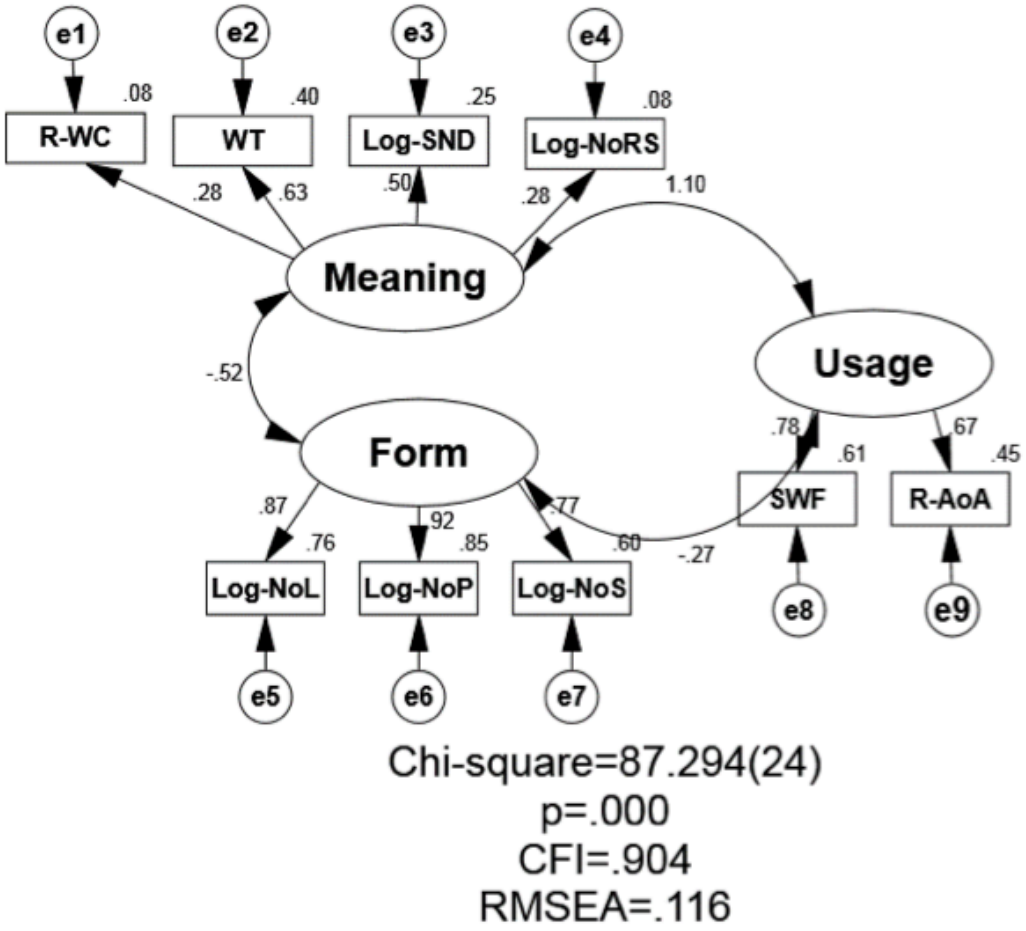
Linearity was assessed using pairwise scatter plots, and it was found that all observed variables appeared to be linearly related.

Outliers. There are two types of outliers: univariate and multivariate outliers. A univariate outlier is any data point with an extreme or unusual value on a single variable, while multivariate outliers are cases with an unusual combination of scores on two or more variables (Tabachnick & Fidell, 2012). One way of testing univariate outliers is to convert raw scores to standardized scores and then to examine the maximum and minimum scores against conventional standards based on a sample size. According to Stevens (2009), conventionally, the z values for a sample size larger than 100 should be less than 4.00 and more than -4.00, and the percentage of observations beyond 3 standard deviations should be less than 1. Following these guidelines, there was no case falling outside the tolerance limits. To assess multivariate outliers, the Mahalanobis distance, a relative measure of data point's distance from the group centroid under the hypothesis of normality, was calculated (B. M. Byrne, 2009). One multivariate outlier (line 61) was identified and removed from further analysis ($p < .001$)

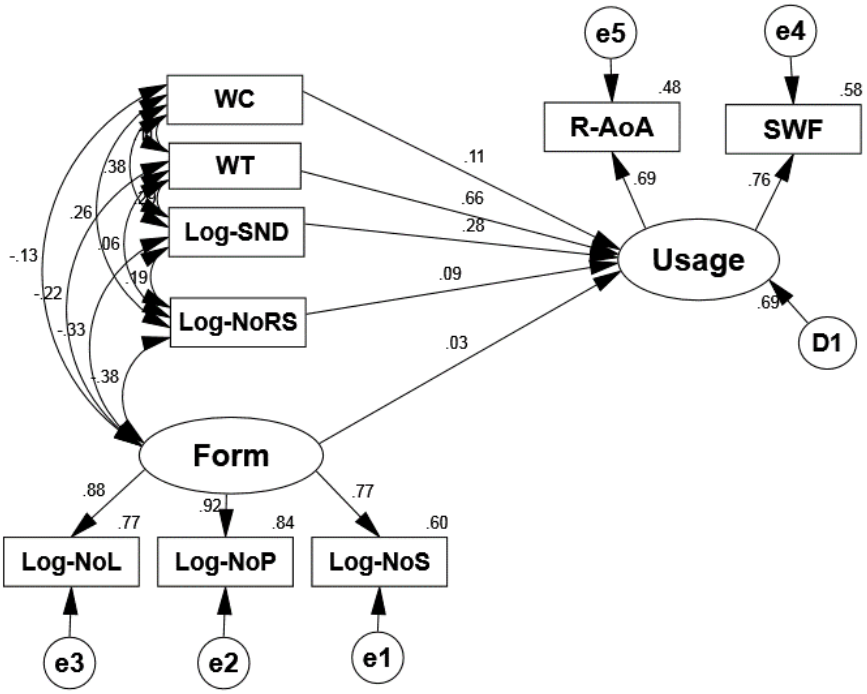
Multicollinearity and singularity. Multicollinearity is a high correlation between two or more independents (.90 and above), and singularity is a perfect or nearly perfect correlation between two or more independents. According to Tabachnick and Fidell (2012), highly correlated independents can lead to a failure of achieving

statistical significance of regression coefficients, and a failure of model converge on a solution. To examine for multicollinearity and singularity, a scan of the correlation matrix of the indicants showed that none of the correlation coefficients exceeded 0.90, indicating the absence of multicollinearity and linearity.

Appendix 5.4 Measurement model with standardized coefficients: Meaning, Form, and Usage of L2 words among adult L2 learners (IBM SPSS AMOS output)



Appendix 5.5 Re-specified SEM model with standardized coefficients: Do Meaning and Form predict Usage of L2 words among adult L2 learners (IBM SPSS AMOS output)



Chi-square=35.432(16)
p=.003
CFI=.970
RMSEA=.079

Appendix 5.6 Results of associations between nominal learner-related variables and the response latency of PE and that of FT among adult L2 learners (n = 94)

A. PE and learner-related nominal variables

Variables	Lambda value	Approx. Sig.
Gender	.01	.32
Major	.01	.32
Ethnicity	.01	.32
Place of birth	.07	.05
Rural/Urban	.01	.32
Ways of English learning	.04	.04

B. FT and learner-related nominal variables

Variables	Lambda value	Approx. Sig.
Gender	.01	.32
Major	.01	.32
Ethnicity	.01	.32
Place of birth	.07	.05
Rural/Urban	.01	.32
Ways of English learning	.04	.04

Appendix 5.7 Assumptions of MR: Predictive effect of learner-related variables on PE and FT response latencies among adult L2 learners (n = 94)

Sample size and missing data. For the present study, with a ratio of 20 observations to each IV, the necessary sample size is 60 for predicting the PE and 80 for predicting the FT. The number of cases in the present study (n = 94) met the requirements of sample size. There were no missing data.

Normality, linearity, and homoscedasticity of residuals. Analysis of the scatterplot showed that all residuals closely formed a rectangular pattern, indicating that the above assumptions were met (see Figure 1 for examples).

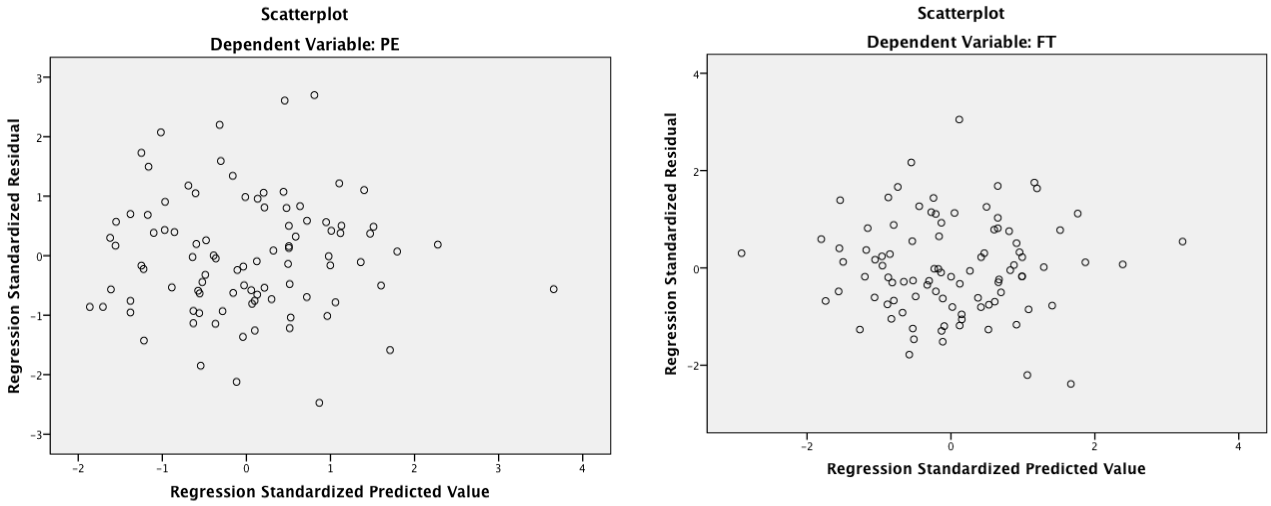


Figure 1 Residual scatterplot (Output from IBM SPSS REGRESSION.)

Outliers. Univariate outliers were checked using standardized scores, which showed that all values were within the limits (+/-3.00 SD). Multivariate outliers were checked by examining the Mahalanobis distance of each case to the centroid of all cases. Mahalanobis distance is distributed as a chi-square variable, with degrees of freedom equal to the number of IVs. The results showed that none of the cases exceeded the value of 16.27 (critical χ^2 at $\alpha = 0.001$ for 3 dfs) in the model predicting PE response latencies, and the value of 18.47 (critical χ^2 at $\alpha = 0.001$ for 4 dfs) in the model predicting FT response latencies. Therefore, no multivariate

outliers were observed in the data set.

Multicollinearity and singularity. None of the tolerances approached zero, and all the variance inflation factor (VIF) values were not significantly far from 1 (see Table 1 below), so there was no inappropriate level of multicollinearity and singularity.

Table 1 Collinearity statistics (Tolerance and VIF)

Model predicting PE response latencies

Variables	Tolerance	VIF
Vocabulary depth	.958	1.044
Age of starting L2 learning	.964	1.038
Age	.942	1.061

Model predicting FT response latencies

Variables	Tolerance	VIF
Vocabulary depth	.951	1.051
Age of starting L2 learning	.952	1.051
Age	.942	1.061
Amount of exposure	.978	1.022

Appendix 5.8 Assumptions of SEM model: Predictive effect of word feature-related variables on PE and FT response latencies among adult L2 learners

Predictive effect of word feature-related variables on PE response latencies

Identifiability of the model. As a preliminary check of the identifiability of the model, the number of data points and parameters to be estimated was counted. With 10 variables, there were 10 (10+1)/2 = 55 data points. The hypothesized model contained 36 parameters to be estimated; therefore, the model was over-identified and tested with 19 dfs.

Sample size and missing data. According to the minimum sample size requirement, it was considered sufficient to conduct SEM with the obtained size of 197 words although the desirable size was not reached. There were no missing data in the data set.

Normality and linearity. The distributions of all lexical variables as well as log-transformed variables were normally distributed. As for PE response latencies, its distribution was normal (skewness value = .38, and kurtosis value = -.19). Multivariate normality was checked by Mardia's statistic (value = 1.40), which was well below the threshold of 5.00; therefore, multivariate normality was met in the current analysis.

As Tabachnick and Fidell (2012) pointed out, it is not feasible to examine all pairwise scatterplots to assess linearity; so pairs of scatterplots were randomly examined using IBM SPSS GRAPHS. All observed pairs appeared to be linearly related, if at all.

Outliers. There were no univariate outliers for word feature-related variables, and there were no univariate outliers for PE response latencies (within +/-3.00 SD). Multivariate outliers were checked using the Mahalanobis distance, and there were no multivariate outliers observed.

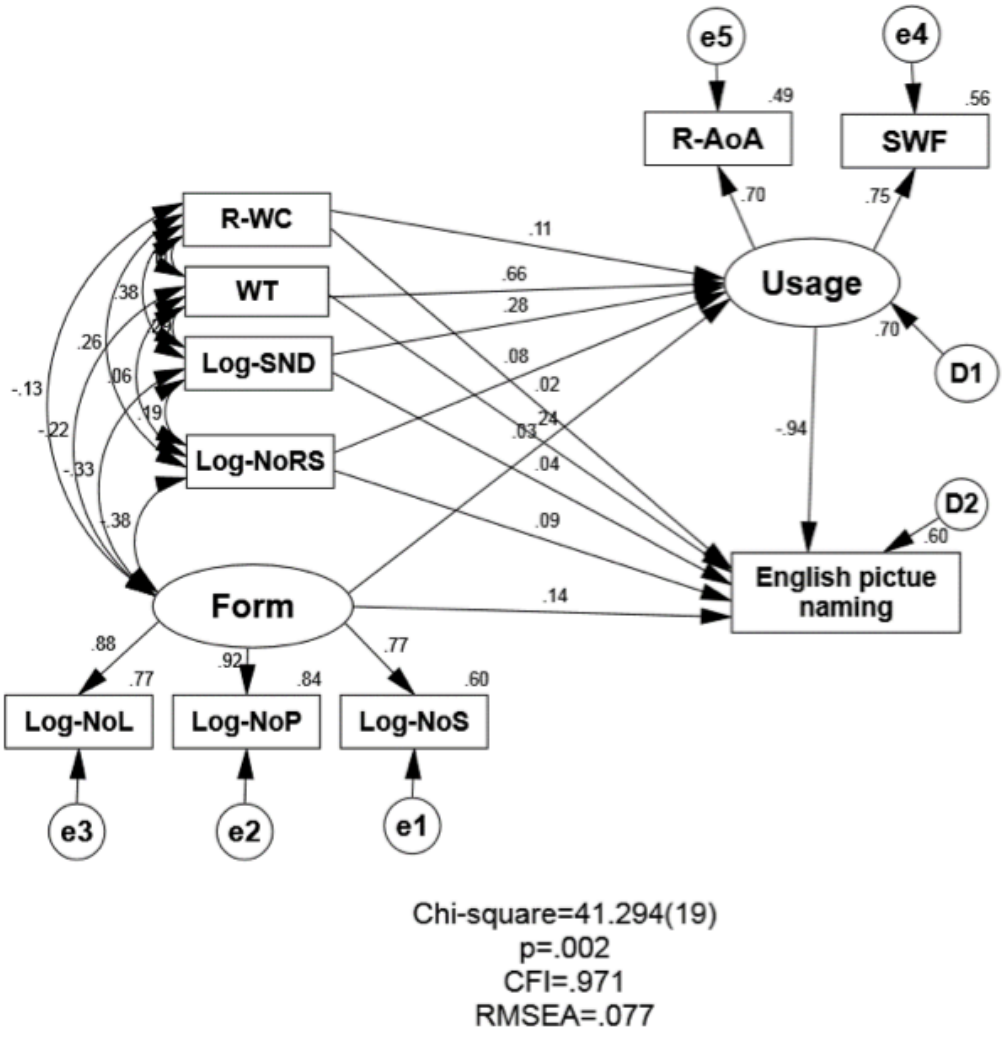
Multicollinearity and singularity. To examine for multicollinearity and

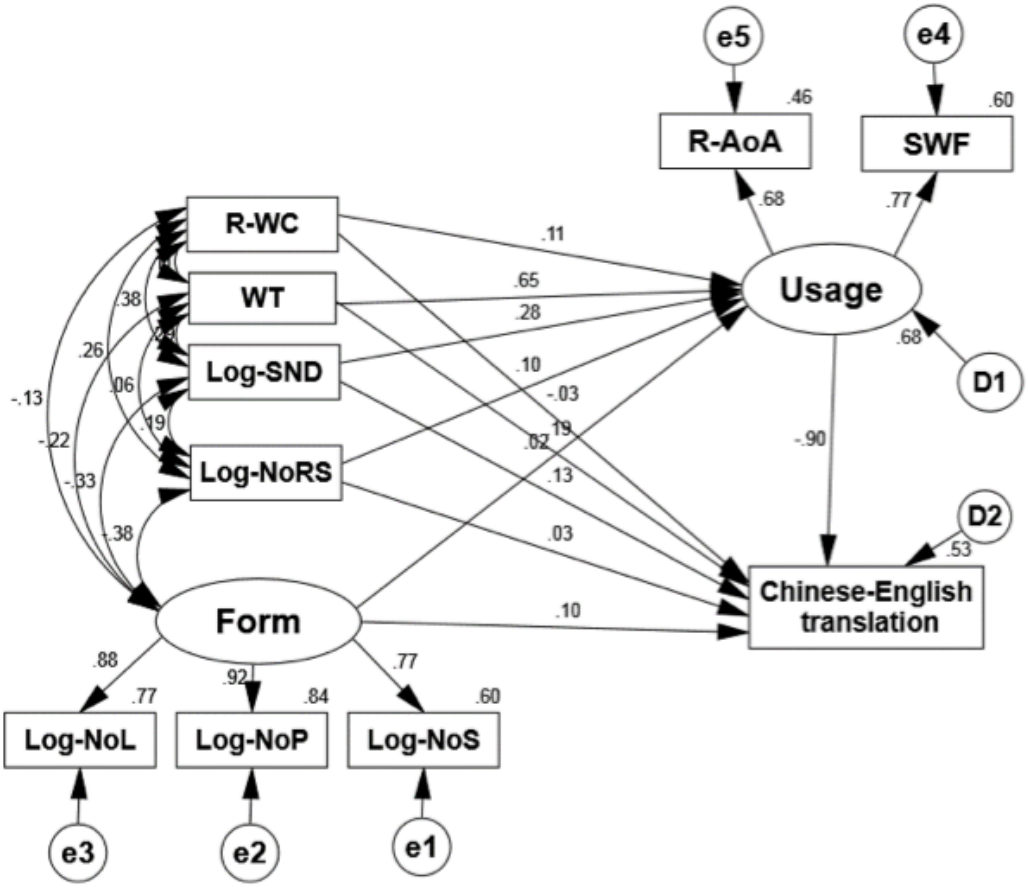
singularity, a check on the correlation matrix of indicants showed that none of the correlation coefficient exceeded 0.90, indicating the absence of multicollinearity and linearity.

Predictive effect of word feature-related variables on FT response latencies

All the assumptions were checked in a similar way in which the PE model was evaluated. The results suggested that none of the assumptions was seriously violated in terms of sample size, missing data, normality, linearity, outliers, multicollinearity and singularity.

Appendix 5.9 SEM model with standardized coefficients: Predict effects of word feature-related variables on PE and FT response latencies among adult L2 learners (IBM SPSS AMOS output)





Chi-square=37.189(19)
p=.008
CFI=.976
RMSEA=.070

Appendix 5.10 Results of associations between nominal learner-related variables and PF-FT response latency differences among adult L2 learners (n = 94)

Variables	Lambda value	Approx. Sig.
Gender	.01	.32
Major	.01	.32
Ethnicity	.01	.32
Place of birth	.07	.05
Rural/Urban	.01	.32
Ways of English learning	.04	.04

Appendix 5.11 Assumptions of MR: Predictive effect of learner-related and word feature-related variables on PE-FT RL differences among adult L2 learners

LEARNERS-RELATED VARIABLES

Sample size and missing data. As for the current analysis, with a recommended ratio of 20 observations to each predictor, the necessary sample size was estimated to be 20 for assessing the effect of academic achievements on the PE-FT response latency difference. The number of cases in the present study (n = 94) was considered sufficient. There were no missing data.

Normality, linearity, and homoscedasticity of residuals. Univariate normality was checked using skewness and kurtosis values. The univariate normality of the predictor variable had been checked and confirmed in Section 5.3.1. As to the outcome variable, its skewness value was .33, and kurtosis value was -.33, which suggested the normality of its distribution. Analysis of the scatterplot showed that all residuals closely formed a rectangular pattern, indicating that the assumption of linearity and homoscedasticity of residuals were not violated.

Outliers. Univariate outliers were checked using standardized scores, which showed that all values were within +/-2.00 SD. Multivariate outliers were checked by examining the Mahalanobis distance of each case to the centroid of all cases. The results showed that none of the cases exceeded the value of 10.83 (critical χ^2 at $\alpha = .001$ for 1 dfs). Therefore, there were no multivariate outliers identified in the data set.

Multicollinearity and singularity. The assumption of multicollinearity and singularity were checked in the same way as described in Section 5.3.1. None of the tolerances approached zero, and all the VIF values were not significantly far from 1, suggesting that there was no inappropriate level of multicollinearity and singularity in the data set.

WORD FEATURE-RELATED VARIABLES

Sample size and missing data. As for the current analysis, with a recommended ratio of 20 observations to each predictor, the necessary sample size was estimated to

be 100 for predicting the PE-FT response latency difference. The number of cases in the present study (n = 198) was considered sufficient. There were no missing data.

Normality, linearity, and homoscedasticity of residuals. Univariate normality was checked using skewness and kurtosis values. The univariate normality of predictor variables had been checked and confirmed in Section 5.3.2. As to the outcome variable, its skewness value was .33, and kurtosis value was -.33, which suggested the normality of its distribution. Analysis of the scatterplot showed that all residuals closely formed a rectangular pattern, indicating that the assumption of linearity and homoscedasticity of residuals were not violated.

Outliers. Univariate outliers had been checked using standardized scores. Values of the outcome variable were within +/-3.00 SD limits. Multivariate outliers were checked by examining the Mahalanobis distance of each case to the centroid of all cases. The results showed that none of the cases exceeded the value of 22.46 (critical $\chi^2\ at\ \alpha = .001$ for 5 dfs). Therefore, there were no multivariate outliers in the data set.

Multicollinearity and singularity. The assumption of multicollinearity and singularity were checked in the same way as described in Section 5.3.2. None of the tolerances approached zero, and all the VIF values were not significantly far from 1, suggesting that there was no inappropriate level of multicollinearity and singularity in the data set.