Testing Applicability of Eye-tracking and fMRI to Translation and Interpreting Studies: An Investigation into Directionality

A thesis submitted to the University of London
in partial fulfillment for the degree of Doctor of Philosophy
and Diploma of Imperial College London

by

Chieh-Ying Chang

Department of Humanities

Imperial College London

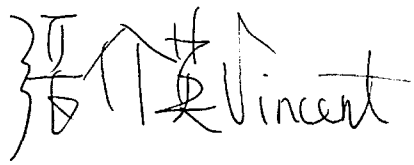
University of London

2009

Declaration

The work presented in this thesis was carried out in the Department of Humanities on the South Kensington Campus and the Imaging Sciences Department on the Hammersmith Hospital Campus of Imperial College London from the 1st of October 2005. This thesis is the result of my own work and any quotation and image from, or description of the work of others is acknowledged herein by reference to the sources, whether published or unpublished.

This thesis is not the same as any that I have submitted for any degree, diploma or other qualification at any other University. No part of this thesis has been or is being concurrently submitted for any such degree, diploma or other qualification.



Signature of Ph.D. Candidate: _____

To My Beloved Family in Taiwan

Acknowledgements

There are so many people who I would like to thank for the completion of my Ph.D. I came from a humanities background and ventured out into the world of hard-core science to measure cognitive loading accompanying translation and interpreting directions. Along the way, many people kindly and generously provided me with help and guidance.

Firstly, I would like to extend my heart-felt appreciation to my supervisors, Mr. Mark Shuttleworth and Dr. Gabriela Saldanha, for their supervision, encouragement and support throughout the course of my Ph.D. pursuit. They gave me the freedom to engage in an exciting line of research, using eye-tracking and functional magnetic resonance imaging (fMRI), in collaboration with Imaging Sciences Department (ISD). On the same note, I would also like to thank Professor Basant K. Puri and Dr. Alex Dresner at ISD for their intellectual, experimental and technical support regarding the fMRI study presented in my Ph.D. Special gratitude is also extended to Mr. Stephen Oliver, who generously loaned me an eye-tracker for my eye-tracking studies free of charge. Further, were it not for Dr. Gabriela Saldanha, who took the trouble to travel with me to Mr. Oliver's office to negotiate over the use of the eye-tracker, the eye-tracking experiments in the study would not have been possible. Moreover, many thanks go to Mr. Mark Shuttleworth and Professor Basant K. Puri who helped me to obtain ethical approval for my studies. I am also very much indebted to Dr Fabiana Gordon at the Statistical Advisory Service (SAS) office at Imperial College London for her expertly and friendly statistical advice.

Secondly, there were many of my good friends who mentally supported me along the Ph.D. journey, and they are Edward Liu, Marcus Yeh, Scott Hsieh, Ian Franks, Pei-Chun Kao, Ting-Hsuan Huang, Baron Henrik von Platen, Karen Hsiao-Ling Chung, Liya Wu, Tzu-Hui Chuang, Chi-kai Hsu, Jason Lee Mitchell, Patrick Reimer, Robin Trischler, Aaron Brown, Marc Fernandez, Marcus McCown and Winz Lin. It is they that helped me to overcome

many mental hurdles throughout this journey, and I am indebted to them. Further, I would like to extend my appreciation to all those people who participated in my experiments to make possible my doctoral study.

Thirdly, I would like to thank my fellow Ph.D. colleagues, and they are Elina Lagoudaki, Idalina Gomes, Ana Mialdea, and Panayiota Vatikioti. Our strong and close bond kept me going till the very end of my Ph.D., and I am honored that I can call them my life-long friends.

Fourthly, two of the most important people in my life, Iain McNaughton and Gregory Koulisis, have my deepest gratitude. Both of them have been mentors, friends, and even family members to me. Their most generous, tolerant, and endearing personality and character have made them guardian angels to me in different stages in my doctoral pursuit. Naturally, my Ph.D. would not be at all possible without their support and love.

Fifthly, I would like to thank Professors Jennifer Thomson and Steven Pinker at the Harvard University, whose world-class knowledge inspired me and further solidified my Ph.D. thesis when I collaborated with them for neuroimaging experiments as a visiting fellow at Harvard University.

Last but not least, I would like to thank my beloved family. It is their love for and faith in me that saw me through the entire doctoral journey. I am most proud of my family, and I hope I have made them proud.

Abstract

The Revised Hierarchical Model (Kroll and Stewart, 1994) in the fields of bilingualism and second language acquisition has argued that the task of translating/interpreting 'single words' into the second language is more cognitively demanding than into the first language. Past empirical studies focused mostly on the behavioral aspect of professional translators and interpreters or highly proficient bilinguals. However, there is a scarcity of neurological and physiological evidence to support the above suggestion, and the population of novice translators and interpreters has long been neglected as potential research subjects. This study aims to *explore* if the predictions suggested by the Revised Hierarchical Model are valid at a textual level and can be extended to a textual level by collecting physiological and neurological measures accompanying the cognitive loading caused by different translation/interpretation directions on the part of novice translators and interpreters. In order to collect such measures as data on the cognitive loading, methods of eye-tracking and fMRI are employed; by doing so, the study also aims to *test* the applicability of the two methods specifically to the field of Translation and Interpreting Studies.

This is an area of importance as the findings can help us understand novice translators' physiological reactions and novice interpreters' brain activation patterns. With this in mind, it may in future be possible to design more effective educational or training regimens to help novices improve their translating or interpreting skills. The potential findings may also be useful, if combined with software development in future research, to develop software to stimulate brain activities to help acquire a skill, e.g. translation or interpreting. This aims to lay the foundation for possible future development.

Regarding the structure of this doctorial thesis, **Chapter One** presents the background to the study. **Chapter Two** surveys related literature related to the research aims

and questions in the study. **Chapter Three** describes the methodology adopted for the two eye-tracking experiments reported in the research program. **Chapter Four** details the first eye-tracking experiment with the language combination of Mandarin (A) and English (B). **Chapter Five** reports the second eye-tracking experiment with the language combination of Spanish (A) and English (B). **Chapter Six** describes an exploratory fMRI study with the language combination of Mandarin (A) and English (B). **Chapter Seven** presents a discussion regarding the entire doctorial study, limitations and recommended future work.

Table of Contents

Thesis Title:
Testing Applicability of Eye-tracking and fMRI to Translation and Interpreting Studies: An Investigation into Directionality 1
Declaration 2
Acknowledgements 4
Abstract 6
List of Abbreviations 14
List of Figures 16
List of Tables 19
Chapter One Introduction 21
1.1 Background to the Study 21
1.2 Aims and Research Questions 24
1.3 Research Significance and Motivation of the Study 24
Chapter Two Positioning Research Perspective 26
2.1 Translation and Interpreting Studies (T&I Studies) as a Discipline 26
2.2 Process-oriented Approaches to Translation and Interpreting Studies 26
2.3 Translation Directionality 29
2.4 Recent Further Behavioral Evidence for the Revised Hierarchical Model 32
2.5 Controversies over the Revised Hierarchical Model 35
2.6 The Effect of Age on Language Acquisition 37
2.7 Neurological Investigations of Translation Asymmetry 40
2.8 Eye-tracking Studies on Translation and Interpreting 46
2.9 Conclusion 53
2.9.1 Two Eye-tracking experiments (Language Combinations: Mandarin as A – English as B vs. Spanish as A – English as B) 54
2.9.2 One Exploratory Functional MRI experiment (Language Combination: Mandarin as A – English as B) 54
Chapter Three Methods: Eye-tracking 56
3.1 Introduction 56
3.2 Basic Eye Biology 56
3.3 Principles of Eye-tracking Technology 57
3.3.1 Pupil and Cornea Reflection Discrimination 57
3.4 Optimal Eye-tracking Laboratory Setup 58
3.4.1 Laboratory Room 58
3.4.2 Light Meter 59

3.4.3 Eye-tracking System 59
3.4.3 Sampling Frequency 61
3.4.4 Software Installation 61
3.4.5 Camera Setup 62
3.4.6 Eye Calibration 63
3.4.7 Testing Calibration 64
3.5 Experimental Designs 66
3.5.1 Designs 66
3.5.1.1 Mandarin (A) – English (B) Eye-tracking Experiment 66
3.5.1.2 Spanish (A) – English (B) Eye-tracking Experiment 67
3.5.2 Material Design and Selection 68
3.5.2.1 Mandarin (A) – English (B) Eye-tracking experiment 69
3.5.2.2 Spanish (A) – English (B) Eye-tracking experiment 73
3.5.3 Definitions of Dependent Variables 76
3.5.3.1 Pupil Diameter 76
3.5.3.2 Fixation Count 76
3.5.3.3 Task Time(sec) 77
3.5.3.4 Fixation Frequency (decimal numbers) 77
3.5.3.5 Blink Frequency (decimal numbers) 78
3.5.3.6 Fixation Duration/Time(sec) 79
3.6 Data Collection 79
3.6.1 Participants 79
3.6.1.1 Mandarin (A) – English (B) Eye-tracking Experiment 79
3.6.1.2 Spanish (A) – English (B) Eye-tracking Experiment 80
3.7 Stimuli Presentation 80
3.8 Pilot Testing 82
3.9 Task Sequence 82
3.9.1 Mandarin (A) – English (B) Eye-tracking Experiment 82
3.9.2 Spanish (A) – English (B) Eye-tracking Experiment 82
3.10 Handedness 83
3.10.1 Mandarin (A) – English (B) Eye-tracking Experiment 83
3.10.2 Spanish (A) – English (B) Eye-tracking Experiment 84
3.11 Ocular Dominance 84
3.11.1 Mandarin (A) – English (B) Eye-tracking Experiment 85
3.11.2 Spanish (A) – English (B) Eye-tracking Experiment 85
3.12 Questionnaire Information 85
3.12.1 Mandarin (A) – English (B) Eye-tracking Experiment 86
3.12.2 Spanish (A) – English (B) Eye-tracking Experiment 89
3.13 Fixation Plotting 91
3.14 Experimental Procedures 92

3.14.1 Description Given to the Participants 92
3.14.2 Participant Preparation 93
3.14.3 Practice Session 94
3.15 Conclusion 95
Chapter Four Mandarin (A) – English (B) Eye-tracking Experiment 96
4.1 Introduction 96
4.2 Behavioral Results 97
4.3 Key Concepts behind *t* Test in Statistics 99
4.4 Difference in Global Cognitive Loading between Mandarin Chinese → English vs. English → Mandarin Chinese Translations 105
4.4.1 Pupil Size(mm) 105
4.4.2 Overall Fixation Count across Texts on the Screen 105
4.4.3 Task Time (measured in milliseconds) 106
4.4.4 Fixation Frequency across Texts on the Screen (decimal numbers) 106
4.4.5 Blink Frequency (decimal numbers) 106
4.5 Secondary Results: Attention Allocation to Target Text (TT) and Source Text (ST) 107
4.6 Repeated Measure 2-way ANOVA on Pupillometry 108
4.6.1 Post-hoc Bonferroni-corrected Paired Samples t Test 111
4.6.2 Fixation Data for Typing Mandarin Chinese and Typing English 113
4.7 Repeated Measure 2-way ANOVA on Fixation Count 117
4.8 Repeated Measure 2-way ANOVA on Fixation Frequency 119
4.9 Repeated Measure 2-way ANOVA on Task Time (sec) 120
4.10 Further Exploratory Statistical Analyses 122
4.10.1 Correlation Analysis 122
4.10.2 Principal Components Analysis (PCA) 128
4.10.3 Longitudinal Analysis 133
4.11 Conclusion 139
Chapter Five Eye-tracking Experiment Two: Spanish (A) – English (B) 140
5.1 Introduction 140
5.2 Behavioral Results 141
5.3 Difference in Global Cognitive Loading between Spanish → English and English → Spanish Translation 142
5.4 Secondary Results: Attention Allocation to Target Text (TT) and Source Text (ST) 143
5.5 Repeated measures 2-way ANOVA on Pupillometry 144
5.5.1 Bonferroni-corrected Post-hoc Paired Samples *t* Tests 146
5.6 Repeated Measures 2-way ANOVA on Fixation Count 149
5.7 Repeated measures 2-way ANOVA on Fixation Frequency 151

5.8 Repeated measures 2-way ANOVA on Task Time 153
5.9 Longitudinal Analysis 156
5.10 Conclusion 161
Chapter Six Exploratory fMRI Experiment: Mandarin-English 163
6.1 Introduction 163
6.2 Physical Principles of MRI vs. Functional MRI (fMRI) 163
6.2.1 MRI 164
6.2.2 Functional Magnetic Resonance Imaging (fMRI) 168
6.2.3 Brain Volume Formation 170
6.3 Brain Anatomy and Cognitive Functions 171
6.4 fMRI Experimental Design 175
6.4.1 Participants 176
6.4.2 Psychological State 177
6.4.3 Handedness 178
6.4.4 Informed Consent 178
6.5 fMRI Sparse Paradigm 180
6.6 Experimental Paradigm 180
6.6.1 Pure Insertion 181
6.6.2 Material 183
6.7 Experiment Setup on E-Prime 186
6.7.1 Sparse Scanning 188
6.8 Trial Arrangement 190
6.9 MRI Scanner and Exam Card 190
6.10 Piloting Using Phantom Scanning 191
6.11 Human Piloting 192
6.12 Pre-scanning Procedure 193
6.13 Data Storage 196
6.14 Data Analysis 196
6.14.1 Behavior Results 196
6.14.2 Neurological Results 198
6.15 Steps Involved in Analyzing the Neurological Data 199
6.15.1 Slice Timing 199
6.15.2 Realignment 199
6.15.3 Coregistration (Coregistering Functional Images with Structural Images) 200
6.15.4 Normalization 201
6.15.5 Smoothing 204
6.15.6 First-Level Analysis 204
6.15.7 Second-Level Analysis 204
6.16 Results 205

6.16.1 Interpreting from English into Mandarin Chinese vs. Interpreting from Mandarin Chinese into English 205
6.16.2 Brain Activation Patterns during Interpreting from English into Mandarin Chinese 208
6.16.3 Brain Activation Patterns during Interpreting from Mandarin Chinese into English 219
6.17 Findings 231
6.18 Discussion 233
6.19 Conclusion 239
Chapter Seven Discussion, Limitations and Recommended Future Work 240
7.1 Summary of the Research Program 240
7.2 Relevance of the Research Findings to Theoretical Considerations 241
7.3 Summary of Research Conclusions 241
7.4 Recommended Future Work 243
7.4.1 Extending the Languages Investigated 243
7.4.2 Extending the Eye-tracking Work 244
7.4.2.1 Potential Hypotheses for Future Testing 245
7.4.3 Extending the fMRI Work 246
7.5 Conclusion 248
References 249
Appendix A: English Text for Translation 279
Appendix B: English Text for Typing 280
Appendix C: Mandarin Text for Translation 281
Appendix D: Mandarin Text for Typing 282
Appendix E: Spanish Text for Translation and Reading 283
Appendix F: Spanish Text for Typing 284
Appendix G: Consent Form for the Mandarin – English Eye-tracking Experiment 285
Appendix H: Consent Form for the Spanish – English Eye-tracking Experiment 287
Appendix I: Random Permutation Table 289
Appendix J: Language and Translation Background Questionnaire (Mandarin – English Eye-tracking Experiment) 290
Appendix K: Language and Translation Background Questionnaire (Spanish – English Eye-tracking Experiment) 294
Appendix L: Description of the Procedure of the Study 297
Appendix M: Practice Material for English → Mandarin Translation 298
Appendix N: Practice Material for Typing English 299
Appendix O: Practice Material for Mandarin → English Translation 300
Appendix P: Practice Material for Typing Mandarin 301
Appendix Q: Practice Material for Spanish → English Translation 302

Appendix R: Practice Material for Typing and Reading Spanish **303**
Appendix S: Hospital Anxiety and Depression Scale (HADS) **304**
Appendix T: Consent Form for the fMRI Experiment **305**

List of Abbreviations

A: First Language or A Language or Language A
A→B: Translate or Interpret from the First Language into the Second Language
AOI: Area of Interest
ART: Attention Responsive Technology
ASL: Applied Science Laboratory
B: Second Language or B Language or Language B
B→A: Translate or Interpret from the Second Language into the First Language
bmp: bitmap (file format)
BOLD: Blood Oxygenation Level Dependent
CE: Translate from Mandarin Chinese into English
CI: Confidence Interval
COGAIN: Communication by Gaze Interaction
cm: centimeters
CR: Corneal Reflection
DICOM: Digital Imaging and Communications in Medicine
EC: Translate from English into Mandarin Chinese
EEG : Electroencephalogram
ERP : Event-related Potentials
ES: Translate from English into Spanish
df *or df*: Degrees of Freedom
DTI: Diffusion Tensor Imaging
fMRI: Functional Magnetic Resonance Imaging
hz: Hertz
ICREC: Imperial College Research Ethics Committee
L1: First Language or A Language or Language A
L2: Second Language or B Language or Language B
M: Mean
MRI: Magnetic Resonance Imaging
mm: millimeters
ms: milliseconds
mW: milliwatts
N: Number
p: *p* value; Level of Statistical Significance
PCA: Principal Component Analysis
PET: Positron Emission Tomography
r: Effect Size
REC: Reading Mandarin Chinese
REE: Reading English

SD: Standard Deviation
SE: Translate from Spanish into English
SE: Standard Error of Mean
sec: seconds
Sig.: Significance
SIEC: Sight Interpreting from English into Mandarin Chinese
SICE: Sight Interpreting from Mandarin Chinese into English
SPM: Statistical Parametric Mapping
ST: Source Text
Std. Error: Standard Error
T: Tesla
T&I Studies: Translation and Interpreting Studies
TMS: Transcranial Magnetic Stimulation
TT: Target Text

List of Figures

Figure 1. Revised Hierarchical Model 30
Figure 2. Basic Eye Biology 56
Figure 3. ASL 504 Model Monocular Eye-tracker and Laboratory Setup 60
Figure 4. Camera Setup and Participant Placement 63
Figure 5. Split Window for Stimuli Presentation and Text Production 81
Figure 6. FixPlot Interface 92
Figure 7. A Bell-shaped Normal Distribution 100
Figure 8. Two Normal Distributions with Different Means 102
Figure 9. Two Normal Distributions with the Same Mean Value 103
Figure 10. Interaction of Input Language and Task Type (Pupillometry, mm) 110
Figure 11. Pupil Sizes (mm) during Typing English and Typing Mandarin Chinese 112
Figure 12. Fixation Counts (integers) during Typing English and Typing Mandarin Chinese 113
Figure 13. Interaction between Input Language and Task Type Based on Fixation Count 118
Figure 14. Interaction between Input Language and Task Type Based on Fixation Frequency (decimal numbers) 119
Figure 15. Interaction between Input Language and Task Type Based on Task Time (sec) 121
Figure 16. Scatter Plot for Pair Wise Correlations of the five Variables during Mandarin → English Translation 124
Figure 17. Scatter Plot for Pair Wise Correlations of the five Variables during English → Mandarin Translation 126
Figure 18. Pupillometric Data (mm) during Mandarin → English Translation 135
Figure 19. Pupillometric Data (mm) during English → Mandarin Translation 136
Figure 20. Interaction of Input Language and Task Type (Pupil Size, mm) 146
Figure 21. Interaction of Input Language and Task Type (Fixation Count, integers) 149
Figure 22. Interaction of Input Language and Task Type (Fixation Frequency, decimal numbers) 152
Figure 23. Interaction of Input Language and Task Type (Task Time, sec) 154
Figure 24. Pupillometric Data (mm) during Spanish → English Translation 157
Figure 25. Pupillometric Data (mm) during English → Spanish Translation 158
Figure 26. Alignment of Protons to Magnetization 166
Figure 27. States of Protons When Radio Frequency is Applied vs. Off 167

Figure 28. T_1-weighted Image 168
Figure 29. T_2 Image 168
Figure 30. Linearity of a Stimulus and BOLD Signal 169
Figure 31. PLANAR Image Sequential Acquisition 170
Figure 32. Brain Anatomy 172
Figure 33. Limbic System 173
Figure 34. General Cognitive Functions of the Four Lobes of a Brain 175
Figure 35. Philips 3T Intera Scanner Used at the Robert Steiner MR Unit at Imaging Sciences Department, Hammersmith Hospital Campus, Imperial College London, UK. 176
Figure 36. One Volume Acquisition 189
Figure 37. A Phantom Used for Phantom Scanning 192
Figure 38. Visual Stimuli Presentation in the Scanner 195
Figure 39. MatLab Environment 198
Figure 40. Coregistration 201
Figure 41. Variations of Brain Sizes of the Five Subjects in the fMRI Study. 202
Figure 42. Stereotaxic Space on the Left and a Talairach Space on the Right 203
Figure 43. Brain Activation Patterns during Interpreting from English into Mandarin Chinese 209
Figure 44. Coronal, Sagittal, and Horizontal Brain Images 210
Figure 45. Activation of Right Parahippocampal Gyrus. 211
Figure 46. Activation of Left Parahippocampal Gyrus. 212
Figure 47. Activation of the Left Angular Gyrus. 213
Figure 48. Activation of the left inferior parietal lobule 214
Figure 49. Activation of the Right Putamen 215
Figure 50. Activation of the Left Inferior Frontal Gyrus 216
Figure 51. Activation of the Uncus 217
Figure 52. Activation of the left caudate head 218
Figure 53. Brain Activation Patterns during Interpreting from Mandarin Chinese into English 219
Figure 54. Activation of the Right Insula. 220
Figure 55. Activation of the Right Putamen 221
Figure 56. Activation of the Cingulate Gyrus 222
Figure 57. Activation of the Left Dorsolateral Prefrontal Cortex 223
Figure 58. Activation of the Right Lingual Gyrus 224
Figure 59. Activation of the Left Superior Temporal Gyrus 225
Figure 60. Activation of the Left Superior Temporal Gyrus 226
Figure 61. Activation of the Left Parahippocampal Gyrus 227
Figure 62. Activation of the Right Superior Temporal Gyrus 228
Figure 63. Activation of the Precentral Gyrus 229

Figure 64. Activation of the Cerebellum **230**
Figure 65. Activation of the Right Fusiform Gyrus **231**

List of Tables

Table 1. Comprehensibility Test. 71
Table 2. Readability Test 72
Table 3. Comprehensibility Test. 74
Table 4. Readability Test. 75
Table 5. Descriptive Statistics on Handedness of the Mandarin – English Participants 83
Table 6. Descriptive Statistics on Handedness of the Spanish – English Participants 84
Table 7. Descriptive Statistics of Percentages of B→A (English into Mandarin Chinese) Translation Training in the Three Institutes. 88
Table 8. Descriptive Statistics of Percentages of A→B (Mandarin Chinese into English) Translation Training in the Three Institutes. 88
Table 9. Descriptive Statistics of Percentages of Professional Translation from Chinese Mandarin into English vs. from English into Mandarin Chinese 89
Table 10. Descriptive Statistics of Percentages of A→B (Spanish into English) and B→A (English into Spanish) Translation Training. 90
Table 11. Descriptive Statistics of Percentages of Professional Translation from Spanish into English vs. from English into Spanish 91
Table 12. Descriptive Statistics on Scorings of the Outputs (Mandarin - English) 98
Table 13. Reasons Why Typing Mandarin Chinese or Typing English Was Considered More Difficult 115
Table 14. Pair Wise Correlation Coefficients of the Five Variables during Mandarin → English Translation 125
Table 15. Pair Wise Correlation Coefficients of the Five Variables during English → Mandarin Translation 127
Table 16. Component Matrix with Correlation Coefficients (Mandarin → English) 130
Table 17. Total Variance Explained by the Components (Mandarin → English) . 131
Table 18. Component Matrix with Correlation Coefficients (English → Mandarin) 132
Table 19. Total Variance Explained by the Components (English → Mandarin) . 132
Table 20. Mean Pupillometric Values (mm) of the Three Intervals during Mandarin → English Translation 137
Table 21. Pair-wise Comparisons between the Three Intervals during Mandarin →

English Translation 137
Table 22. Mean Pupillometric Values (mm) of the Three Intervals during English → Mandarin Translation 138
Table 23. Pair-wise Comparisons between the Three Intervals during English → Mandarin Translation 138
Table 24. Descriptive Statistics on Scorings (Spanish - English) 142
Table 25. Reasons Why It Is More Difficult to Type Spanish or Type English 148
Table 26. Mean Pupillometric Values of the Three Intervals during Spanish → English Translation 159
Table 27. Pair-wise Comparisons between the Three Intervals during Spanish → English Translation 159
Table 28. Mean Pupillometric Values of the Three Intervals during English → Spanish Translation 160
Table 29. Pair-wise Comparisons between the Three Intervals during English → Spanish Translation 160
Table 30. Anxiety and Depression Scoring for Each Participant (0 - 7 = normal, 8 - 10 = mild, 11 - 14 = moderate, 15 - 21 severe) 177
Table 31. Handedness Scores 178
Table 32. Cognitive Subtraction for Semantic Processing 182
Table 33. Cognitive Subtraction for Simultaneous Interpreting 182
Table 34. Comprehensibility Test. 184
Table 35. Readability Test 185
Table 36. Descriptive Statistics on Scorings 197
Table 37. Coordinates of Gray Matter Areas Activated during Interpreting from English into Mandarin Chinese. 206
Table 38. Coordinates of Gray Matter Areas Activated during Interpreting from Mandarin Chinese into English. 207
Table 39. Comparison of the Present Results and Those from Prior Studies 238

Chapter One Introduction

1.1 Background to the Study

Economic globalization and the existence of international political initiatives, such as the United Nations and the European Union, have greatly increased the demands made upon translators and interpreters. In the *Routledge Encyclopedia of Translation Studies*, Lonsdale (1998: 63-67) provides a succinct historical account of the effects of these recent social and commercial pressures. One important consequence is that translators and interpreters frequently have to work in both directions, that is, not only *into* their first language (B→A) but also *from* that language (A→B). Pöchhacker (2004) argues that, although the western tradition of conference interpreting prefers translation to be performed in the former direction, the practical reality in organizations such as the UN and the EU requires interpreters to interpret bilaterally, i.e. both A→B and B→A.

Williams and Chesterman (2002: 52) point out that psycholinguistics researchers have simplified the process of translation as 'Input → Black box → Output', with black box referring to what is happening in a translator's mind during the very act of translation. Past studies regarding cognitive loading of translation and its association with directionality have been based mainly on behavioral evidence (for a review, see Harley, 2001). Visual attention and physiological studies investigating 'translation directionality' are scarce.

Interpreting requires the ability of 'concurrent' comprehension of a text in an input language (presented either in an auditory or visual format) and verbal production of the same message in an output language. Conference interpreters need to process a number of cognitive components: (1) listen to or read large chunks of information in a source language; (2) extract and temporarily store the information; and (3) reproduce it in a target language (Gile, 1995). Past studies have shown that the interpreters' output speech production in the target language significantly overlaps with the speaker's speech in the source language and

this has suggested 'concurrent processing' of the cognitive components. Chernov (1979: 277) reports that English-Russian interpreters' speech output overlaps with the speaker's to a maximum of 89% during 70.5% of the total simultaneous interpretation performance. Other past studies have demonstrated that English-French interpreters' overlapping time range between 65% and 75% (Barik, 1969; Gerver, 1975; Goldman-Eisler, 1974). Lee (1999) reports that English-Korean interpreters' overlapping time is between 40% and 60%. However, the immense cognitive effort required in interpreting and its association with directionality remains equally under-researched.

It is often assumed that directionality makes a significant difference to the competence of most translators and interpreters (see Chang, 2005; Pavlović, 2007). A theory named the Revised Hierarchical Model (Kroll & Stewart, 1995; see **Section 2.3**) supports this suggestion. However, most of the research literature in translation and interpreting studies documenting this issue is based on self-reported and behavioral data or is related to how translators or interpreters adopt different strategies during translation or interpreting (see Chang, 2005; Pavlović, 2007). Although some preliminary work has been done, as yet there is no empirically verified theory or model to explain the effect of directionality on cognitive loading induced by translation or interpreting at a 'textual' level.

Technological advances, e.g. eye-tracking and functional magnetic resonance imaging (fMRI) can help researchers obtain information regarding pupillary, ocular and neurological activities accompanying translation and interpreting. The two techniques have been widely applied and firmly established as useful methods of research in a variety of disciplines (see **Chapter Two**). However, their general or broader applicability, if any, specifically to translation and interpreting studies has not yet been demonstrated. This thesis represents a first step in covering this gap.

This study concerns novice bilingual translators and novice interpreters who acquire their B language later in their lives (i.e. toward the end of puberty or post-puberty, or at an early age with minimally sporadic exposure to a second language until puberty; the significance of this point is explained in **Section 2.6**) and who become proficient enough to

be trained as novice translators or interpreters in adulthood. It seeks to empirically investigate the difference, if any, in cognitive loading induced by reversing the translation direction in this novice population. Two experimental methods are used: (1) eye-tracking and (2) fMRI.

This thesis begins with a literature survey in the area of bilingualism and second language acquisition, especially focusing on the important early as well as the latest studies offering a theoretical background as well as experimental and empirical evidence. The focus of the literature survey is mainly on previous experimental studies because the study presented here is also experimental in nature. The available literature providing physiological and neurological evidence of the effect of directionality on translation and interpreting is also reviewed. Given that this thesis primarily focuses on methodological issues, physical principles of eye-tracking and fMRI are introduced as a crucial part of the literature review. Further, the available literature demonstrating the potential applicability of eye-tracking and fMRI specifically to the field of Translation and Interpreting Studies is also reviewed in **Chapter Two**. The eye-tracking methodology of the research program is then outlined in **Chapter Three**, alongside a detailed description of the eye-tracking laboratory setup, because this thesis primarily deals with testing research methods, and because the laboratory was set up by the researcher of the study from ground up at the Department of Humanities, Imperial College London. Since this is a novel method in Translation and Interpreting Studies, it is expected that this type of information will be useful to ensure replicability for future studies along the same lines. The thesis includes two eye-tracking studies which are discussed respectively in **Chapter Four** and **Chapter Five,** while **Chapter Six** presents an exploratory fMRI study, alongside an introduction on brain gross anatomy as well as physical principles of the fMRI method. In **Chapter Seven**, a discussion of the findings, relevance of the research findings to the theoretical considerations and limitations of the entire study is provided, alongside suggestions for further investigation.

1.2 Aims and Research Questions

The research questions to be answered can be summarized as follows:

- Can the method of eye-tracking be adequately applied to reveal cognitive loading associated with A→B and B→A translations of novice translators[1]?
- Apart from measuring cognitive loading, can the eye-tracking data obtained yield additional results and findings that are useful in formulating potential research questions or hypotheses to be tested by future studies (see **Section 2.8**)?
- Can the method of functional MRI be adequately applied (with an aim of avoiding orofacial movement as an artefact during data collection, see **Chapter Six**) to reveal brain activation patterns associated with A→B and B→A interpreting at textual level by novice interpreters?
- Can the predictions suggested by the Revised Hierarchical Model be valid at a textual level, based on the results of the three experiments?

1.3 Research Significance and Motivation of the Study

Providing satisfactory answers to these research questions has a number of potential benefits. Firstly, it will provide powerful additional techniques for further investigation of this field of study. Secondly, potential findings regarding directionality may help to inform pedagogy.

In the translation and interpreting market place, bilateral translation and interpreting is commonplace (Malkiel, 2004) and translation into the second language is gaining ground (Wilss, 1996: 8). It has also been revealed that most professional translators do work into their second language in their profession (Weatherby, 1998: 21). If it can be

[1] One might argue that the notion of A→B translation is more cognitively more demanding than B→A is intuitively plausible. However, the primary spirit of 'positivism' is to test an assumption by repeated experimentation, even if the assumption is normally considered *'common sense'*. Positivists believe either *confirming* or *disconfirming* such an assumption via repeated experimentation is a contribution to knowledge. Therefore, testing an assumption that appears to be intuitively cognizable still adds value to the realm of knowledge (see also **Footnote 6** in **Section 2.5**) (for a detailed discussion, see Tolman, 1992; Schick, 2000).

established that translating or interpreting into the second language is generally more cognitively demanding than working into the first language, then it is arguable that translator/interpreter-training should take this into account in its curricula, so that future students will be better equipped to face future market challenges.

A further possible application is in the area of software development. Potential neurological findings regarding 'translation asymmetry' (see **Section 2.3**), if interfaced with software development, can be used to develop software-based behavioral tasks with the aim of establishing cortical connections that would help develop translation and interpreting skills (for how behavioral tasks can help to build cortical connections that would help develop a skill, see Buonomano and Merzenich, 1998; Olesen *et al.*, 2004).

Chapter Two Positioning Research Perspective

2.1 Translation and Interpreting Studies (T&I Studies) as a Discipline

This study considers translation studies, combined with interpreting research, as one discipline called Translation and Interpreting Studies or T&I Studies. Snell-Hornby (1991) argues that translation studies has long neglected interpreting research and that translation studies should be expanded as an *interdiscipline* for it to be comprehensive. Opposed to this view, in a translation conference held in Dublin in 1995, some scholars argue that interpreting research should be separated from translation studies as a stand-alone discipline (Baker, 1998: 279). Baker (1998) contends that translation and interpreting scholars must not split the discipline into opposing sub-factions and urges collaborative unity in working toward maturity of research related to both translation and interpreting. Pöchhacker (2004: 1) also urges 'unity in diversity' for the field of translation and interpreting studies. This argument has gradually been reflected in reality, as a large number of academic institutions are now offering Master's degrees or doctoral programs in T&I Studies as a discipline, e.g. University of Bath and University of Newcastle upon Tyne. The study presented here subscribes to the case for an integrated discipline of 'T&I Studies'.

2.2 Process-oriented Approaches to Translation and Interpreting Studies

The study involves quantitatively measuring cognitive loading induced by directionality during the cognitive *process* of translation and interpreting, using such methods as eye-tracking (during translation) and fMRI (during interpreting). The study presented here therefore falls within the scope of 'process-oriented approach' to T&I Studies. However, very little has been done on this front using state-of-the-art technologies. Most of the process-oriented research conducted in the past was based on behavioral data

and/or helped to reveal translators' or interpreters' use of strategies during translation and interpreting. Even though such previous research is not entirely related to the current study on directionality supported by physiological and neurological data, a brief overview regarding the 'process-oriented' T&I studies is presented (for a more detailed review, see Hansen, 2002).

Process-oriented research into translation has long been underway. Lörscher (1989) gives a critical review of how the field of translation studies modeled the process of translation. Séguinot (1989) gave a lucid description of the translation process, and used pauses and hesitations as measurements to investigate the process. Campbell (1991) administered translation tests to explore the process of translation and suggested a model that was based on translation competences. T&I Studies borrowed paradigms or methods from cognitive linguistics (e.g. Schilperoord and Sanders, 1997) to use pauses and hesitations as indicators of cognitive effort during the translation process (O'Brien, 2006b). Think aloud protocols (TAPs) to tap into the cognitive process of professionals vs. non-professionals or student translators have been employed by authors like Jääskeläinen (1989, 2002) and Jääskeläinen and Tirkkonen-Condit (1991); further, overviews regarding how different methods can be employed to explore the translation process have been presented by Alves (2003) and Jääskeläinen (2005). Danks *et al.* (1997) gave a overview of how researchers had been using a cognitive approach to investigate translation and interpreting. Translog (www.translog.dk), a key stroke logging application developed at the Copenhagen Business School, has been widely used to study the translation process too (e.g. Atari, 2005; for a review, see Jakobsen, 2006). Retrospective interviews and TAPs have been employed in connection with Translog to study translation strategies (e.g. Tirkkonen-Condit and Jääskeläinen, 2000). Kring (2001) has investigated the cognitive aspect of the machine translation post-editing process. Methodological complexities accompanying studies regarding the translation process and the effect of expertise on the translation process have been thoroughly introduced by Englund-Dimitrova (2005). Choice Network Analysis or CNA has been presented as another method to model the cognitive process of translation (Campbell, 2000; O'Brien, 2005). PACTE (2000) is a research group at Universitat

Autònoma de Barcelona in Spain that has been investigating how translation competence is acquired. Minimax theory has been adopted to explain translation processes where translators seek optimal solutions (e.g. Lörscher, 1992; Levỳ, 2000). Recommended methodological issues, in particular suggestions regarding experimental designs, to explore the translation process are presented by Alves and Vale (forthcoming).

In terms of interpreting, process-oriented research has also gained momentum. Gran and Fabbro (1988) used 'dichotic listening' [2] and 'finger-tapping manual-verbal interference'[3] experimental paradigms to determine cerebral lateralization for syntactic and semantic components in simultaneous interpreting. Moser-Mercer (1997) provides an information-processing model of simultaneous interpreting and the model indicated the importance of lexical, semantic, syntactic processing and of long-term memory. Setton (1999) collected corpora of interpreters' outputs and developed a cognitive-pragmatic paradigm to explain an interpreter's cognitive processes. Interdisciplinary approaches to study the cognitive process of interpreting were reported by Englund-Dimitrova and Hyltenstam (2000). The association of working memory and conference interpreting has been studied (e.g. Christoffels *et al.*, 2003; Christoffels and De Groot, 2004; Padilla *et al.*, 2005; De Groot and Christoffels, 2006; Christoffels *et al.*, 2006). Chang (2005) utilizes retrospective interviews to explore strategy use on the part of interpreters.

The studies listed above, situated within the long-standing psychological or linguistic research paradigm, collected merely behavioral data and elicited self-reported data. However, there has been a paucity of physiological and neuroscientific evidence accompanying the phenomenon of translation or interpreting. Tommola (1999) argues that future research, with the help of technological advances, should be able to transcend paradigmatic boundaries, and argues for a 'neurophysiological' or 'neurolinguistic'

[2] Dichotic listening is an experimental paradigm where a participant wears a headset. During an experiment, the participant listens to two different sound stimuli simultaneously or sequentially fed into the participant's right and left ears respectively. This paradigm was used because it resembled *simultaneously listening to both an input and an output* during interpreting.

[3] Finger-tapping manual-verbal interference is an experimental paradigm where a participant has to either tap a button or utter a word whilst listening or watching a stimulus. This causes interference, and has been employed to investigate conference interpreting because it is a simulation of 'multitasking' of interpreting.

approach to investigate the process of translation or interpreting. To date, however, little research has been done using this 'neurological' and 'physiological' approach in the field of T&I Studies with the aid of newly developed technologies. This thesis therefore aims to bridge this gap.

2.3 Translation Directionality

It has been argued by Kroll and Stewart (1994) that more cognitive effort has to be made when a person performs a single-word translation task from his/her first language into his/her second language (A→B), whereas relatively less is required in the opposite direction (B→A). They reported a key piece of experimental research in support of this suggestion. Bilingual undergraduate students were instructed to perform two kinds of tasks: (1) verbally naming a particular item presented in a list of pictorial representations of concrete objects, and (2) translating a particular word presented in a word list. Sixteen bilingual subjects with the language pair of Dutch (A) and English (B) participated in the first task, and twenty four bilingual subjects with the same language combination participated in the second. Cognitive loading generated to provide the required word in response to the picture or word cue was measured by reaction times. The list in which the pictures to be named or the words to be translated were located could comprise objects either in a similar conceptual category (e.g. scarf, shoes, stockings) or in random categories (e.g. horse, skates, cherry), and the subjects could be required to name the pictorial object either in their first or second language as well as to carry out single word A→B or B→A translation.

The experimenters' hypothesis was that, when a general category was stimulated by a list of related objects or words, many lexical representations would be simultaneously activated. As a result, when required to name a particular picture or translate a single word within the list, other words would compete in the subject's mind, creating 'category interference' and *slowing the mental process.*

Interestingly, their results confirmed that category interference did indeed occur (i.e.

longer reaction times were observed) when the bilingual subjects worked A→B, but no such interference was found when they worked B→A. Kroll and Stewart argued that this 'translation asymmetry' demonstrated the existence of *conceptual* as well as *lexical* associations in the minds of bilingual individuals. They devised a Revised Hierarchical Model (Figure 1) to *a posteriori* explain this theoretical interpretation of their findings. Further, the Revised Hierarchical Model (Kroll and Stewart, 1994) explicitly suggests a strong correlation between reaction time and cognitive loading (or mental effort); in other words, cognitive loading is measurable by reaction time – the longer the reaction time is, the higher the cognitive loading. This point has been demonstrated by a number of studies (e.g. Heredia, 1996; Brysbaert and Dijkstra, 2006).

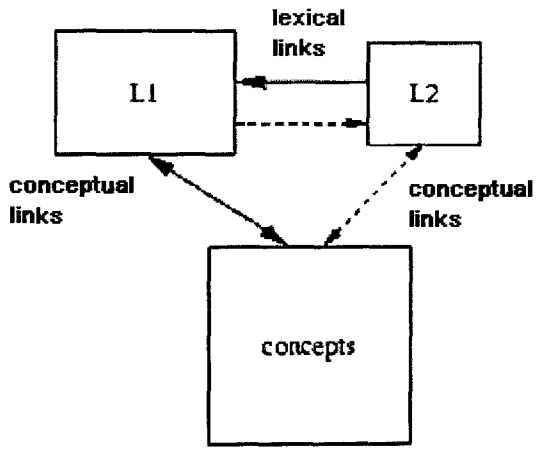


Figure 1. Revised Hierarchical Model

Source: Kroll & Stewart (1994: 158).

Arrows represent the directions of associations or links. Bold arrows represent comparatively stronger links and dotted-line arrows represent weaker links.

In Figure 1, the box representing A language[4] (L1) is larger than the box representing B language (L2). This indicates that an individual's vocabulary in his/her first language is

[4] In this thesis, L1, A language and language A all refer to an individual's first language, whereas L2, B language and language B all mean a person's second language.

always larger than that in his/her second language. When a learner acquires a new word in his/her second language, a strong lexical link is created to a corresponding word in his/her first language, so it can be remembered effectively. For instance, if an English (A) – French (B) bilingual is presented with the word 'chat' in French, s/he will experience a strong and immediate association with the word 'cat' in English. On the other hand, words in an individual's first language primarily arouse conceptual associations. Presented with the word 'cat', the bilingual individual does not immediately think 'chat', but rather experiences imaginative associations linked with other English words and/or non-verbal mental images. The direct lexical link to the second language is thus comparatively weak.

Using this model, Kroll and Stewart (1994) reasoned that, when their experimental subjects were required to work B→A, vocabulary selection exploited the direct lexical links that had been learned, whereas when working A→B, conceptual associations were inevitably aroused which slowed the cognitive process. They further suggested that, although growing improvement in second-language proficiency could moderate category interference, it could never completely eliminate it. Second language learners always rely on their dominant language to process second language meaning to some extent, although direct conceptual processing exclusively within the B language of course becomes more possible as the learners' B language proficiency improves over time.

The Revised Hierarchical Model is intuitively plausible and some further experimental evidence has been reported that supports it. However, the theory has been challenged as an over-simplification. In the following two sections, further evidence both for and against the model is surveyed.

2.4 Recent Further Behavioral Evidence for the Revised Hierarchical Model

Latest research has offered substantial behavioral evidence in support of the Kroll and Stewart hypothesis (Sholl *et al.*, 1995; de Bot, 2000; Kroll *et al.*, 2002; Kroll and Tokowicz, 2001; Sunderman and Kroll, 2006; Smith, 2007; Schwartz *et al.*, 2007; Prior *et al.*, 2007; Hoshino and Kroll, 2007). For example, de Bot (2000) compared reaction times for lexical processing in both directions for participants at three levels of bilingual proficiency. Translation asymmetry was clearly observed. In other words, producing equivalent words in language B took a longer time than in language A, although the effect became less marked as the level of proficiency rose. In the same study, the results also showed that producing the syntax for language B was less automatic and often required conscious monitoring and hence more cognitive effort.

Kroll *et al.* (2002) conducted an experiment where two groups of English native speakers with French as their second language — one group more proficient in French than the other — were instructed to carry out word-naming and translation tasks. In terms of word-naming, the subjects were asked to read aloud a word visually presented in front of them as quickly and accurately as possible in either language. Where a translation task was involved, they were instructed to name the translation of a word stimulus as quickly and accurately as possible in another language. Their results of reaction times, consistent with what the Revised Hierarchical Model predicted, showed that both groups were less quick to name a translation from English into French (A→B) than one from French into English (B→A). However, the phenomenon of 'translation asymmetry' was less prominent for the more proficient group (see also Kroll *et al.*, 1998).

Smith (2007) recruited twenty nine participants who were English native speakers with Spanish as their second language. The participants' working memory spans were tested and they were grouped into 'high span' and 'low span' groups. The subjects were asked to verbalize translations as quickly as possible, based on visually presented word stimuli of varying levels of concreteness or with varying numbers of possible target

language translations. The results revealed that, regardless of the memory span, the concreteness of the word stimuli or the number of target language translations, reaction time for A→B translation was significantly longer than that for B→A translation.

Schwartz *et al.* (2007) conducted an experiment where eighteen English (A) – Spanish (B) bilinguals were instructed to translate Spanish or English words that were either cognate or non-cognate as quickly as possible. Regardless of whether the words were cognate or not, the reaction time for A→B translation was significantly greater than that for B→A translation.

Prior *et al.* (2007) recruited both Spanish (A) – English (B) and English (A) – Spanish (B) bilinguals who were grouped based on their second-language proficiency levels to carry out lexical decision tasks, using nouns, verbs and word-class ambiguous items that were controlled for such factors as imageability, concreteness, familiarity, word frequency and cognateness. The factors will be further detailed in **Section 2.5**. In the study, a lexical decision task is to write down the translation they could think of for each stimulus word on the word list. Translation ambiguity was a factor in question. They reported that when it comes to dealing with A→B and B→A translations, participants with higher second-language proficiency performed better at negotiating translation ambiguity. Further, the results revealed that, irrespective of language combinations and varied second-language proficiency levels, working A→B was slower than in the reverse direction. The results supported the assumptions of the Revised Hierarchical Model, because (1) A→B translation was demonstrated to be more cognitively demanding than B→A, and (2) if a bilingual's second-language proficiency was higher, s/he was quicker to produce translations for both A→B and B→A.

Hoshino and Kroll (2007) conducted a study where Spanish (A) – English (B) and Japanese (A) – English (B) subjects were requested to name a picture in English. The pictures for the picture-naming tasks were chosen in such a way that some of the usual names for the objects in the pictures were cognates in the subject's A and B languages, and the others were not. For example, a picture of guitar would be 'guitar' in English, 'guitarra' in Spanish and

'gi-ta-a' in Japanese, which makes the picture cognate with the two languages within a language combination. Their results revealed that the bilinguals were quicker to name a picture in English when it was cognate with their language combination. This demonstrated that cognateness facilitates conceptual mediation during cross-language production. At the very least, this suggests that conceptual mediation was prevalent and required for cross-language activation, especially A→B, and might be susceptible to the effect of other factor(s).

However, up to now, such investigations into the effect of directionality on translation have suffered a number of weaknesses. Although claims are made regarding the effect of improved proficiency, little work has been done with novice translators or novice interpreters. In addition, the experimental tasks have involved only associations between 'single words', as opposed to authentic translation tasks with texts. In fact, the Revised Hierarchical Model seems to assume that meaning resides in 'single words'; however, whether or not the predictions regarding 'translation asymmetry' suggested by the model are valid at a *textual* level has yet to be explored. A further criticism is that the distinction between lexical and conceptual links within the brain has been inferred from purely behavioral evidence (reaction times) and lacks direct physiological or neurological confirmation. It is likely that the verbal and imaginative associations (conceptual processing) evoked by language, especially at a textual level, are in fact much more complex than this model suggests. Last but not least, the thesis presented here argues that, due to the nature of the RHM, 'predictions' can definitely be made, but made only on single-word translations of, for instance, 'different language combinations' and 'varying levels of word lengths and frequencies, but NOT on 'texts'. This point has been confirmed by Kroll (2009). The model's 'predicting capacity' is therefore limited in this sense. This is why experiments were designed (see **Chapters Four, Five and Six**) to *explore* whether or not the predictions associated with the asymmetry suggested by the Revised Hierarchical Model are valid at a *textual* level without a definitive hypothesis (see also 'two-tailed *t* test' in **Section 4.3**).

2.5 Controversies over the Revised Hierarchical Model

Some researchers have, therefore, challenged Kroll and Stewart's Revised Hierarchical Model, arguing that the model is based on experimental results using 'concrete', 'simple' words, e.g. nouns, without taking into account other factors. It is worth mentioning that de Groot (de Groot, 1992b; de Groot, 1992a) argues that such factors as imageability, context availability, definition accuracy, familiarity, word frequency, cognateness and word length had an effect on A→B translation or forward translation, well before researchers, e.g. Kroll and Stewert (1994), started to explore the different processes of translation in both directions.

De Groot *et al.* (1994) conducted an experiment to investigate whether the model only predicted correctly for concrete lexical items. Dutch native speakers with non-fluent English as a second language were recruited to carry out only B→A translation. Words of different levels of imageability, context availability and definition accuracy[5] were selected. Their results demonstrated that all the three factors had a significant effect on the reaction time for B→A translation. The researchers concluded that, even for non-fluent bilinguals, B→A translation involved semantic processing as well as mere lexical association.

In a second experiment (de Groot, 1994), Dutch native speakers were recruited whose English proficiency was higher and instructed them to conduct both A→B and B→A translation. Words of varying levels of concreteness were selected. Interestingly, their results showed that the reaction time for A→B translation was significantly shorter than that for B→A translation (1011 ms vs. 1076 ms) when translating words of a relatively high level of concreteness. However, when translating words of a relatively low level of concreteness, the reaction for B→A translation was significantly shorter than that for A→B translation (1351

[5] Before the experiment, the words used as stimuli in this experiment underwent ratings on a scale of 1 to 7 by Dutch-English bilinguals to determine their levels of imageability, context availability and definition accuracy. If a presented word could arouse a mental image very quickly and easily, the rating of imageability would be 7; if not, the rating would be 1. On the same rating scale, context availability meant 'how easy it is to come up with a particular context or circumstance' (de Groot, 1994: 607) where a word might appear. Definition accuracy meant how accurately the bilinguals thought they could define the words.

ms vs. 1406 ms). These results seemed to suggest that the cognitive process used to translate a word differs depending on its concreteness[6]. Concrete words rely more heavily on the direct lexical link. At the very least, this implies that the Revised Hierarchical Model is an over-simplification, and *may not apply at the sentence level where complex meaning is communicated and where textual stimuli other than lexical stimuli are involved.*

Altarriba and Mathis (1997) further investigated the effect of different levels of second language proficiency. They compared English native speakers who had received just one single Spanish lesson to English (A) – Spanish (B) bilinguals. Both groups were asked to carry out a translation recognition task. Word stimuli in the second language were presented for 300 ms, followed by a target word in the first language presented for 1250 ms. The words used as stimuli and targets were either orthographically or semantically related, to see if this would cause any interference. The subjects were instructed to respond by pressing a response key as quickly as possible if the target word was the correct translation of the stimulus. Their results showed that both orthographical and semantic relatedness caused significant interference to both groups. This suggested that, even for novice learners, word-processing in the second language went beyond the word surface form and involved semantic engagement.

Jiang (1999) carried out five experiments to explore the cross-language priming effect during A→B vs. B→A translation recognition tasks. Priming, in an experiment, refers to presenting a stimulus or stimuli that are morphologically, phonologically, semantically, orthographically or syntactically related to the ensuing stimulus in question. Words used for priming in an experiment are called 'prime words'. He recruited bilingual participants for

[6] This result contradicts Kroll and Stewart's suggestion of 'translation asymmetry' (1994). Firstly, this shows that there are other factors apart from the different cognitive loading required by lexical processing and conceptual processing that need to be taken into account. Secondly, this is also helpful in showing the necessity of testing via repeated experimentation an assumption that is generally considered 'common sense' (see **Section 1.2**), as other potential variables or factors may have an effect on the assumption that is seemingly intuitively cognizable. For instance, the study by de Groot (1994) reported that the level of word concreteness was likely to have an effect on reaction time during translation, and her result did not support 'translation asymmetry' suggested by the Revised Hierarchical Model (Kroll and Stewart, 1994).

all the experiments with the language combination of Chinese (A) and English (B). In the study, he presented prime words (both in Chinese and English) and their target translations, non-translations and nonwords. A nonword is a word designed by a researcher for it to seem/sound like a word and yet it is NOT a word, such as 'rekull' or 'findel'. The subjects were instructed to press a YES button if a target word was a real word after presentation of a prime; reaction times were recorded for analysis. He reported a significant priming effect during A→B translation recognition. However, the priming effect for B→A was inconsistent and insignificant. This somehow contradicts the Revised Hierarchical Model, because if B→A translation exploits a direct lexical link, presenting a prime word in a B language should have facilitated recognition of its translation in an A language in the study; however, the results showed that the priming effect was weak for B→A in the study.

A more recent study conducted by Tokowicz and Kroll (2007) involved an experiment on Spanish-English and English-Spanish bilinguals, using both abstract and concrete words whose context availability and concreteness were rated and controlled, along with word frequency and word length. Words were presented on a computer screen and the subjects were instructed to conduct verbal translation as quickly and accurately as possible. They reported, based on reaction time data, that the results were consistent with the predictions of the Revised Hierarchical Model.

Despite the controversies over the Revised Hierarchical Model, it has still been selected as the theoretical framework within which the experimental study was conducted, because this is the only model that specifically points out and explains the phenomenon of 'translation asymmetry' and it is based on scientific and experimental evidence.

2.6 The Effect of Age on Language Acquisition

A further issue that the Revised Hierarchical Model fails to address is the effect of when the second language is acquired. Lennenberg (1967) claimed that, to be fully functional, a language had to be acquired within a 'critical period', before the age of twelve, with intensive exposure. This is called the Critical Period Hypothesis (for a review, see

Birdsong, 1999). Cook (1996) and Gass and Selinker (2001) supported this hypothesis, arguing that younger learners are better than older ones, due to a higher level of brain plasticity. Once brain lateralization is completed during puberty, i.e. around the end of the critical period, it is difficult to acquire a language to the extent that this language can be as functional as a native language.

This hypothesis has been tested in the fields of bilingualism and second language acquisition, and studies have shown that there is indeed a critical period effect in second language acquisition in bilinguals. Lennenberg (1967) reported that second language learners who acquired a second language post-puberty could never obtain a native-like accent, whereas those who acquired a second language pre-puberty could. McLaughlin (1978) also demonstrated that younger learners were better at phonology than older individuals. Johnson and Newport (1991) revealed that the more mature the learners of a second language were, the less proficient they were at the second language and the broader the deterioration of universal properties[7] of the second language.

Bilinguals can also be categorized into three major types (Weinreich, 1968). *Simultaneous* bilinguals are those who acquire both languages from infancy onwards with exposure to both languages with approximately the same level of intensity. *Successive* or *sequential* bilinguals are those who acquire a second language during childhood with extensive second-language exposure; however, such factors as the intensity of exposure and time of acquisition (primary school period, pre-puberty, or post-puberty) have an effect on the level of proficiency. *Late* bilinguals are those who acquire a second language post-puberty.

Mechelli *et al.* (2004) recruited twenty five 'early' bilinguals who acquired a second European language before the age of five and who had practiced it regularly since, as well as thirty-three 'late' bilinguals who started acquiring a second European language between the ages of ten and fifteen and had practiced it regularly for a minimum of five years. They

[7] 'Universal properties' means the overall performance of a language, e.g. grammar, vocabulary, intonation, and accent.

reported that age had a significant effect on the grey matter density in the left inferior parietal cortex (a planar description of neuroanatomy will be introduced in **Section 6.3**). This means that the 'early' bilinguals have a larger volume of cortical connections compared with 'late' bilinguals. Their further analysis also revealed that the overall second language proficiency of the different groups of subjects was highly correlated with grey matter density in exactly the same left inferior parietal region.

A famous case study (cited in Mehler *et al.*, 2000: 51-75) that is worth mentioning in support of the Critical Period Hypothesis is the 'feral child' by the name of Genie in the United States. Up until the age of thirteen, Genie had been abused by her father at home. She was tied to a potty chair during the daytime and bound to a sleeping bag at night, deprived of any form of human communication. When she was discovered by social workers, she only had a vocabulary of approximately forty words and could only manage to say such short negative sentences as 'Stop it' and 'No more'. Even though psychologists and linguists tried to teach her to acquire her first language, her speech was still very short, limited and fragmented. She frequently had to resort to body movement, e.g. stomping feet and slamming doors, in order to express her anger because she did not have the ability to express herself like a native speaker. Throughout Genie's life, she has never managed to obtain native-like English proficiency, as she was given the opportunity to learn English only post-puberty. This case study may not be direct proof of the Critical Period Hypothesis; however, at the very least, it does form some degree of support for the idea.

The Critical Period Hypothesis thus seems to rest on substantial evidence. At the very least, this suggests that the Revised Hierarchical Model of Kroll and Stewart (1994) may need to take into account the age at which translators or interpreters acquire their second language (see also **Section 3.12** regarding the questionnaire design). It seems likely that bilingualism acquired in childhood exploits different cognitive mechanisms within the brain from second language proficiency acquired later in life. Further, it is worth reiterating that the study presented here only concerns the population of 'non-simultaneous' bilingual novice translators or interpreters with specific language combinations, and therefore does

not aim to generalize its results to other populations.

2.7 Neurological Investigations of Translation Asymmetry

A major weakness in the experimental evidence upon which the debate about the Revised Hierarchical Model has been based is that it is exclusively behavioral. However, recent technological developments in the field of medical diagnosis have produced tools that can explore the functioning of the living brain. These have led to an increased understanding of the cerebral activity involved in translation and interpreting. It is now known that the number of activated areas of the cortex is correlated with the cognitive load required to perform a cognitive task (Hasegawa *et al.*, 2002; Fridriksson and Morrow, 2005; Whelan, 2007). This section presents a number of neurological investigations of translation asymmetry, and **Chapter Six** will introduce the significance and cognitive functions of cortical results of these studies.

The first study in the area of translation and interpreting was conducted by Petsche *et al.* (1993). Using an electroencephalogram (EEG), they reported that there was more right-hemispheric coherence when German-English interpreters translated into their B language. EEG records waveforms, i.e. coherence, of various joint electrical activities of dispersed populations of neurons in different grey matter area in the brain during a cognitive task. In terms of language processing, right hemispheric activation is associated with dealing with 'contextual, holistic information', whereas left hemisphere has been demonstrated to be related to 'word retrieval' and 'analytic processing' (Gazzaniga *et al.*, 2002; 423-425). In this case, left-hemispheric coherence was comparable during both A→B and B→A interpreting, whereas the right hemisphere of the brain was recruited to a lesser extent during B→A interpreting than during A→B interpreting, as manifested by the coherence. The results that more coherence existed during A→B interpreting than during B→A interpreting suggests that A→B interpreting required more cognitive effort than B→A interpreting. Later, Petsche and Etlinger (1998) replicated this original study, increasing the number of subjects to ten, and confirmed their previous findings. However, in order to avoid speech artefacts in the

EEG that might have confused the results, the interpreters were required to carry out simultaneous interpreting (SI) silently and mentally, without any vocal articulation. This was an intrinsic experimental flaw, because there was no behavioral data to verify whether or not the subjects were genuinely carrying out the interpreting task as instructed. This means that the researchers might have been unaware of the possibility that the interpreters were in fact thinking of something else, instead of interpreting. Moreover, the subjects recruited were all professional interpreters; no investigation of less proficient interpreters was performed.

Kurz (1994) also used EEG to explore cortical activities during SI. The results showed that the temporal regions were most activated during interpreting and confirmed that 'more coherence increases - particularly in the right hemisphere - were found while interpreting into the foreign than into the native language' (Petsche *et al.*, 1993: 385). However, this study suffered from the same weaknesses as those of Petsche *et al.*, as the task selected was also *silent interpreting*.

Klein *et al.* (1995) used positron emission tomography (PET) to determine the neural correlates of single-word translation conducted by bilinguals. Sixteen bilinguals who were English native speakers with French as their second language carried out a translation task from English into French, alongside repeating auditorily presented single words as the baseline task. They observed activation mainly in the left dorsolateral prefrontal cortex during the translation task. However, results of the study are not generalizable to translation or interpreting, because the subjects recruited were merely bilinguals, not translators or interpreters. In addition, only one translation direction was studied, namely from language A to language B, and the study related only to the translation of isolated single words, not authentic texts.

Price *et al.* (1999) used PET to explore the neural system underlying translation and language switching, using six German native speakers with English as their second language as participants. The subjects were instructed to articulate a translation in response to visually displayed stimuli which were single words either in German or in English of

varying degrees of frequency. The baseline task in this experiment was enunciating the visually presented words in the same language. However, if the subjects did not know the correct translation in the target language, they could say 'No' when translating from German into English or 'Nein' when translating from English into German. The researchers observed increased activation in the anterior cingulate and bilateral subcortical structures, such as putamen and head of caudate, irrespective of word frequency and translation direction. This shows that such a 'verbal translation' task requires conscious 'movement control' associated with subcortical structures (Ward, 2006; 28). One point that is worth mentioning is that this study did not find any significant activation in the dorsolateral prefrontal cortex for the translation task, as described by Klein *et al.*(1995), even though the two studies employed the same experimental task, i.e. single word translation. Further, this PET study suffered from the same weaknesses as those of the earlier study by Klein *et al.* (1995). Additionally, the number of subjects recruited in the study was relatively small.

Rinne *et al.* (2000) used PET to explore brain activation patterns during simultaneous interpreting by Finnish (A) – English (B) 'professional' interpreters. They reported that even though the quality of interpretation outputs in each direction did not differ significantly, interpreting into Finnish elicited more left frontal activation while interpreting into English involved more extensive left-sided fronto-temporal activation. Frontal activation, during language tasks, is often associated with language comprehension, production and decision-making. The human brain's left hemisphere is often associated with language functions. Temporal activation is often related to long-term memory retrieval, e.g. a subject's retrieving a word in a target language upon hearing a word in a source language (see also **Section 6.3**). However, once again only professional interpreters were recruited, so it is unclear whether similar results would have been obtained from novice translators/interpreters. In fact, the authors specifically recommended that further imaging research using 'less proficient students of translation studies as subjects' (ibid.: 88) would be useful.

The first functional magnetic resonance imaging (fMRI) study to measure brain activation at a sentential level was conducted by Lehtonen *et al.* (2005). Eleven bilingual

Finnish (A) – Norwegian (B) speakers participated in this study. The subjects were instructed to silently translate a visually presented Finnish sentence into Norwegian; afterwards, a Norwegian probe sentence was visually presented, and the subjects had to decide whether or not this probe sentence was the correct translation by pressing the 'yes' or the 'no' button on a response pad. The baseline task was similar in nature. The subjects were instructed to silently read the Finnish sentence and keep it in mind, decide whether the Finnish probe sentence visually presented afterwards was identical to the test sentence, and press the 'yes' or 'no' button accordingly. The results showed that the activated areas were all in the left hemisphere of the brain, specifically the ventrolateral prefrontal cortex, globus pallidus, inferior frontal gyrus and basal ganglia. However, once again the design of this experiment was open to criticism. Because the subjects were required to 'silently' translate the sentences and then compare this retained translation with the probe sentence, the results may be more related to recognition, memory storage and decision-making than to translation *per se*. Certainly, the task was not one which translators or interpreters would do in real life. Besides, the subjects recruited were not trained translators or interpreters. Further, fMRI research is very sensitive to orofacial movement as a potential artefact, and this is why the participants in this study were instructed to translate silently. This also means that an applicable method of using texts for trained translators or interpreters to translate whilst avoiding orofacial movement as an artefact during an fMRI experiment has yet to be designed and tested.

The second piece of research that applied fMRI to T&I Studies was conducted by Krick *et al.* (2006). Participants with the language combination of German (A) and English or French (B) were instructed to silently read a text that was scrolled down before their eyes. The text was a short story and after approximately every third sentence, the language would change at sentence boundaries from the subjects' first language into his/her second language. Although this study reported preliminary findings of the neural correlates of language switching, i.e. the dorsolateral prefrontal cortex, the insula, the medially situated anterior cingulated cortex, the superior temporal sulcus and the areas in the inferior parietal lobe, it is not related to the issue of directionality in T&I Studies. It is more concerned with visual

perception during language switching.

Ahrens *et al.* (in press) conducted a third fMRI study where six student interpreters with the language combination of German-Spanish participated. They were instructed to simultaneously interpret synchronized audiovisual speech stimuli from Spanish into German, with free speech production in German as the baseline task. By free speech production, the participants were instructed to produce free, verbal German speeches irrelevant to the audiovisual stimuli being watched. Brain volumes were acquired in a continuous paradigm, without any behavioral data obtained to ascertain if the subjects did carry out the instructed tasks. The study reported the left medial temporal gyrus as the most active area during SI. The study yielded interesting results; however, the study could have been potentially strengthened because (1) no behavioral data were obtained to validate the neurological data; (2) adopting a 'continuous' paradigm in an fMRI study was extremely susceptible to artefects incurred by orofacial movement during volume acquisition with the potential of contaminating the data obtained; (3) choosing free speech production as a baseline task did not justify the assumption behind 'cognitive subtraction' or 'pure insertion' (see **Section 6.6.1**); and (4) only one direction of interpreting was involved in the study. Therefore, an improved experimental design with a paradigm less susceptible to orofacial movement is desirable (see **Section 6.7.1**).

Phillips *et al.* (2006) applied EEG to look at the N400 component during A→B and B→A translation priming. The N400 (or negative 400 ms) component is a negative electrical activity component that peaks at approximately 400 ms after presentation of a stimulus and it has a concentrated distribution in the central scalp area collected by electrodes particularly in an auditory task (Luck, 2005). Its amplitude is inversely associated with conceptual processing; therefore, where minimal conceptual processing is involved during an auditory task, a strong presence of the N400 component shows (Luck, 2005). In the study, fifteen subjects (language combination: English as A – French as B) were instructed to passively listen to a list of five words in different trials. One of the trials was to listen to the same word four times (priming) before listening to the last word in the form of its

translation. They reported that minimal presence of N400 during B→A translation and strong presence of N400 during A→B translation. This meant that B→A translation required more conceptual processing than A→B translation, which contradicted the assumption of the Revised Hierarchical Model. However, the experimental paradigm used in the study, i.e. translation priming, did not reflect what bilinguals or translators/interpreters would do in the real world. Due to the experimental design, the findings may at best explain conceptual processing required to *recognize* translation words, as opposed to *producing* translations. Further, no translators or interpreters were recruited as participants in the study.

The Eye-to-IT project yielded an interesting EEG study (Grabner *et al.*, 2007), reporting the presence of two electrical frequency bands (theta and alpha) (for a detailed introduction, see Luck, 2005: 168-169) sensitive to lexical and semantic processing required in language translation. However, the study did not involve translation directionality.

To summarize, the review of prior studies on translation or interpreting using neuroimaging techniques has shown that a neuroscientific approach to T&I Studies is at its inception. The studies have demonstrated (1) the brain areas of activity involved in translation and interpreting, and (2) the use of neuroimaging to study translation/sight interpreting of 'single words' suggests a *potential* applicability of neuroimaging to investigate translation of whole 'texts'. However, nothing has been done using fMRI to investigate cognitive loading of different interpreting directions specifically on the part of novice interpreters with the language combination of Mandarin (A) – English (B). In addition, given the methodological complexities and technical limitations of fMRI (see **Chapter Six**), an optimal experimental paradigm (see **Sections 6.6.1** and **6.7.1**) that incorporates *textual* (as opposed to singe word) stimuli has yet to be designed and tested. This thesis therefore aims to explore the methodological implications of investigating novice interpreters' cognitive loading during A→B and B→A interpreting using an improved experimental design that includes textual stimuli and allows for oral-facial movement.

2.8 Eye-tracking Studies on Translation and Interpreting

To date, there seems to have been only a limited number of published studies applying eye-tracking to T&I research: Tommola and Niemi (1986), Hyönä *et al.* (1995), O'Brien (2006a), Sharmin *et al.* (2008), Dragsted and Hansen (2008), Sjørup (2008), Jakobsen and Jensen (2008), Caffrey (2008), Carl *et al.* (2008), Jensen (2008), Balling (2008), and Pavlović and Jensen (2009). There is ongoing eye-tracking research within the EU-sponsored Eye-to-IT project (Online), which is to some extent related to the cognitive processes involved in translation; however, only one eye-tracking study (Janyan *et al.*, forthcoming) has been submitted pending publication produced by this project so far, and this study is more related to reading than translating.

Tommola and Niemi (1986) were the first to present a preliminary investigation of cognitive loading of on-line simultaneous interpreting, using data on pupil diameter. A larger pupil diameter indicates a heavier amount of cognitive loading (Rayner, 1998). One female subject was recruited to carry out simultaneous interpreting of five Finnish texts into English. Due to the syntactic differences between Finnish and English, restructuring of the output English sentences was required. In the study, they reported peaks of pupil diameter where restructuring was necessary in the English output.

This seminal study yielded interesting and yet preliminary results, and tentatively showed that eye-tracking was potentially useful to investigate cognitive loading during simultaneous interpreting. However, the study was inconclusive, given its single participant recruited to investigate only one translation direction. Further, only one eye-tracking data-type, i.e. pupil diameter, was used in the study, which leaves ample room for testing if eye-tracking has a broader applicability to the field of T&I Studies.

Hyönä *et al.* (1995) were the second to apply the method of eye-tracking to translation and interpreting studies. They tested whether or not pupil diameter change could be used to investigate the cognitive processing load for simultaneous interpreting (SI) and other language tasks. In their first experiment, nine Finnish-English translation and

interpretation students were recruited to carry out three tasks: SI, shadowing (listening to and repeating a text without translating it) and passive listening. Their results demonstrated that pupillometric measures strongly correlated with the difficulty of the task. The pupil dilated significantly more during SI than during shadowing, and more during shadowing than during passive listening. In their second experiment, eighteen Finnish-English translators were recruited to conduct three tasks: (1) single-word passive listening, but with the added instruction to say YES when they recognized the aurally presented word; (2) single-word shadowing; and (3) single-word translation. Both Finnish and English vocabulary was presented and the words were classified as difficult or easy. A multifactorial analysis showed that there was a significant correlation between pupil size and the type of task, the input language and the difficulty of the words. Pupil size was largest for single-word interpretation, followed by shadowing and passive listening; it was larger for processing the English words than processing the Finnish words; and the more difficult the words were, the more the pupil dilated.

This vital study demonstrated that pupillometric data from eye-tracking was a useful indicator of the cognitive load involved in various language tasks (for a review see, Janisse, 1977; Beatty, 1982). Further, another interesting finding yielded by the first experiment was that there was a 'familiarity effect' during text-based SI from English into Finnish, and this means cognitive loading measured by pupil size decreased over time. Their interpretation was that 'the subject becomes acquainted with the conceptual content and other properties of the texts, as well as the task itself' (Hyönä *et al.*, 1995: 602). This suggests that pupillometric data was useful in detecting changes of cognitive loading within a single task 'over time' (see also Kahneman, 1973).

However, given the task of interpreting involved in the study, whether or not such a pupil-based 'familiarity effect' also potentially exists during *translation* has yet to be verified (see **Sections 4.10.3** and **5.9**). Moreover, the nature of the tasks in the second experiment of this study lacked 'ecological validity', since translators never carry out single word translation in real life. 'Ecological validity' refers to how well an experimental task reflects

the task conducted in real life (see also **Section 3.5.2**). To be reliably relevant to the actual practice of translation, subjects need to be tested using an 'authentic text'. A further limitation of the work was that pupil size was the only kind of ocular data measured. The results would have been far more robust if they had been supported by other kinds of eye-tracking data, e.g. the number of fixations, fixation frequency, and blink frequency.

A third study that had tested the usefulness of eye-tracking to translation studies to date was conducted by O'Brien (2006a). In her study, four professional translators were recruited to translate a text using a translation memory tool. Their pupil dilation data were recorded and used as a measure of the cognitive loads required to process different types of matches: No Match, Fuzzy Match, MT Match and Exact Match. Her results showed that identifying No Match required the most cognitive effort and Exact Match the least. Gaze replay, another type of eye-tracking data, was also used in combination with retrospective interviews that prompted the subjects to recall what they did during the translation task. In this way, the translation strategies used when the subjects were working with the Translation Memory tools were explored, and the study tried to identify the information provided by the software that translators tended to use, or at least to look at more.

Although it has a limited number of subjects, this study did add to the evidence provided by previous studies that suggest eye-tracking is a viable technique for measuring cognitive loads in translation tasks. However, once again only a very limited range of eye-tracking data was obtained, namely pupil size and gaze replay. Given the wealth of collectable data available from eye-tracking technology, whether or not eye-tracking has a broader applicability to T&I Studies remains to be tested.

The latest eye-tracking study (Sharmin *et al.*, 2008) aimed to explore the effect of time pressure and text complexity on fixation patterns of translators. They reported a significant effect of text complexity and no significant effect of time pressure on the translators' fixations on their computer screen. They also indicated that more fixations appeared in the target text (TT) than the source text (ST) during translation; this suggests more attention paid to the target text. In the study, only fixation data was obtained for analysis.

Dragsted and Hansen (2008) employed keystroke-logging and eye-tracking as research methods, and, based on data of eight participants, they suggested a concept of *eye-key span*: the temporal duration between 'a given ST fixation and the production of its TT equivalent' (ibid.: 27). Further, their results suggested that a *translation segment* involves reading and comprehending a segment of the ST, as well as writing and producing its equivalent in the TT. This indicates that the translation process consists of various cognitive components, such as reading, comprehending, and typing, and the result may be helpful in informing experimental designs involving translation (see **Chapter Three**). However, the findings of the study were extremely tentative, due to the limited number of participants. Also, in terms of eye-tracking data, the study only used fixations for analysis.

In a pilot study, Sjørup (2008) explored how to optimally design an eye-tracking experiment, using three professional translators, to determine potentially different patterns of eye movement behavior when processing metaphors during translation. Fixation was the only eye-tracking data collected for analysis. The study suggested that processing metaphors, in comparison with literal expressions, required more fixation time, and that processing metaphors was cognitively more demanding. However, Sjørup (ibid.: 74) self-reported two major weaknesses of the study: (1) a very small number of participants involved in the study; and (2) failure to control and manipulate the textual stimuli so as to eliminate other confounding variables in the study. Therefore, the findings of the study were extremely preliminary and non-conclusive.

Jakcobsen and Jensen (2008) recruited a group of six professional translators and a group of six translation students to carry out four tasks: (1) reading for comprehension only; (2) reading in preparation for translating; (3) reading while verbally translating into their native language, i.e. Danish (also referred to as B→A 'sight translating or interpreting'); and (4) reading while producing a translation by typing. Data of task time, fixation count, fixation duration and fixation transitions between the ST and TT were used in the study. Results of the experiment suggested that, compared with the first two tasks, the third and the fourth tasks generally required significantly longer task time, more fixations and longer

fixation durations, regardless of participant grouping. The results also suggested that translation students needed significantly longer task time, more fixations, longer fixation durations, and more transitions between the ST and TT for all translation-related tasks. Despite interesting results yielded from the study, potential values of other eye-tracking data, such as fixation frequency and blink frequency, and the issue of 'translation directionality' have yet to be addressed.

Using fixation-based and pupillometric data, Caffrey (2008) recruited twenty participants to explore how they processed TV anime with four different subtitling conditions: (1) one-line subtitling; (2) two-line subtitling; (3) one-line subtitling with pop-up glosses that were used to convey cultural information intrinsic to Japanese anime; and (4) two-line subtitling with pop-up glosses. Results revealed that both the number of lines used for subtitling and the use of pop-up glosses had a significant effect on cognitive processing. The study suggested that processing two-line subtitling generated more cognitive loading, compared with one-line subtitling, and that two-line subtitling with glosses created the most cognitive loading. Also, it was suggested that the presence of glosses contributed to a re-distribution of attention between the subtitling and the glosses, and that whenever glosses were presented, more subtitles were omitted. Nevertheless, the study only used two types of eye-tracking data, i.e. fixation and pupil size.

To investigate the phenomenon of translation, Carl *et al.* (2008) proposed a new cognitive research method of 'user-activity data' (UAD) that incorporates keystroke logging and eye movement data. Three participants were requested to translate a number of texts from English into Danish, and their preliminary results based on keystrokes and fixations suggested a usefulness of the research method. It is worth noting that the study pointed out that 'reading while translating' vs. 'reading continuously for comprehension' are two different cognitive tasks, and that 'reading purpose' and 'reading task' are factors that have a strong effect on eye movement (ibid.: 24). This issue will be addressed in the eye-tracking experimental designs in **Chapter Three**.

Another study that only used fixation data was conducted by Jensen (2008) to assess

how accurately an eye-tracking equipment and its associated software could identify what words were fixated upon by eight participants recruited to carry out two tasks: reading and translation. Results suggested that there was no significant difference in accuracy in identifying fixated words between the two tasks.

Balling (2008) proposed a new statistical analysis, i.e. Principal Component Analysis (PCA), for eye-tracking data, using a non-eye-tracking study involving only reaction times to demonstrate the rationale of employing the analysis. Balling (ibid.) presented mixed stimuli of words and nonwords of varying word lengths and frequencies, and requested twenty participants to accurately determine to the best of their ability whether or not a stimulus was a word by pressing a button as quickly as possible. After yielding correlations between the reaction times and the different stimuli, the study suggested a usefulness of PCA, and this kind of analysis was introduced and conducted in the study presented in this thesis (see **Section 4.10.2**).

It is worth noting that Arnáiz Uzquiza (2008) considered how eye-tracking could help explore audience perception of subtitling and the potential of the use of eye-tracking as a method of research to improve the practice of subtitling. However, it only involved a theoretical discussion of how eye-tracking could be applied without presenting any empirical results.

The latest eye-tracking study was carried out by Pavlović and Jensen (2009), and it was related to measuring cognitive loading of translation directionality with the language combination of Danish (A) and English (B). Such eye-tracking data as fixation time, average fixation duration, task time and pupil size were collected from a group of eight translation students and another group of eight professional translators. Text stimuli were used and controlled for genre, syllable count, and sentence length. On the same computer screen, split windows were displayed with one on the top for ST presentation and the other on the bottom for TT production. Their results suggested that (1) processing TT generated more cognitive loading, and this was consistent with the findings of Sharmin *et al.* (2008); (2) regardless of participant grouping, A→B translation was significantly more cognitively

demanding than B→A translation, based on the data of task time and pupil size; and (3) regardless of translation directions, there was significantly more cognitive loading on the part of translation students, compared to professional translators. Two hypotheses the study set out to test were non-conclusive, and they were (1) processing ST during A→B translation would be more cognitively demanding than during B→A translation; and (2) producing TT would be more cognitively demanding during A→B translation than during B→A translation.

However, the study by Pavlović and Jensen (2009) could have been potentially strengthened. First, the text stimuli used and controlled for only 'genre', 'readability', 'grade level'[8], 'syllable count' and 'sentence length' were assumed to be comparable. However, for text stimuli to be comparable in a study closely related to translation directionality, more tests have to be conducted in order to avoid potential confounding variables at a *textual* level (see **Section 3.5.2**). Secondly, the study did NOT at all report behavioral data (how accurate the translation outputs by all the participants were), and therefore the validity of the eye-tracking (physiological) data remained questionable. Thirdly, whether or not the presentation of split windows in the study was suitable is debatable (see **Section 3.7** for an improved condition). Fourthly, the study did not include other types of eye-tracking data for analysis, e.g. fixation frequency and fixation count. Last but not least, the results were based on the Danish (A) –English (B) language combination, and hence are not generalizable to other language pairs.

To sum up, the previous research has to a certain extent suggested eye-tracking as a viable method to study translation or interpreting. However, the number of the eye-tracking studies specifically investigating the phenomenon of translation or interpreting is very limited. Therefore, whether or not eye-tracking has a broader applicability to the field of T&I Studies warrants further testing. By 'broader', the research presented in this thesis means the potential usefulness of other data specific to eye-tracking, e.g. fixation frequency, apart from pupil size and fixation already employed as measurements in

[8] See **Section 3.5.2**.

prior research. By 'broader', it also means the potential, if any, of other eye-tracking results and findings in forming a basis for more concrete hypotheses to be tested by future research. In addition, all of the prior eye-tracking research only collected pupillometric measures, or fixation data, or a combination of both for analysis. Much more work remains to be carried out to test how useful other types of collectable eye-tracking data could be.

2.9 Conclusion

Although there have been a few pioneering studies on the way cognitive loading varies with different translation tasks, much still needs to be learned. Further, even though the Revised Hierarchical Model provides a theory of some usefulness for 'single-word' translation, whether or not its predictions are valid at a *textual* level has yet to be explored. In particular, 'translation asymmetry' has only been investigated in the case of highly competent professional interpreters or non-translator/interpreter bilinguals. The limited evidence suggests it is likely to be a much more significant phenomenon among novice translators and interpreters.

The techniques that have so far been applied to this area of research have also been fairly limited. The general usefulness of eye-tracking has been clearly established, but many of the types of data that this technology can provide have yet to be further explored within T&I Studies, and it has not yet been specifically applied to the question of translation directionality. As for direct neurological evidence, the potential of EEG and PET has been demonstrated, but fMRI, arguably an even more powerful tool for the study of brain activity, has yet to be applied to this field of study at all.

The **Introduction** pointed out questions that require answering: (1) whether 'eye-tracking' and 'fMRI' can be adequately applied to the study of the effect of directionality on cognitive effort; (2) whether eye-tracking has a broader applicability to T&I Studies, based

on other types[9] of eye-tracking data accompanying 'translation directionality', and their potential to form the basis for new hypotheses to be tested by future research; and (3) whether the predictions suggested by the Revised Hierarchical Model regarding the asymmetry are valid at a *textual* level. Therefore, this research project aims to address these gaps with two eye-tracking experiments and an exploratory fMRI experiment as follows:

2.9.1 Two Eye-tracking experiments (Language Combinations: Mandarin as A –English as B vs. Spanish as A – English as B)

- Pupillometry will be used to quantitatively measure the cognitive loading induced in novice translators required to work both A→B and B →A on texts rather than just isolated words with a view to establishing the difference in cognitive loading, if any, between different translation directions.
- The usefulness of other kinds of eye-tracking data (including the number of fixations, fixation frequency, task time and blink frequency) will be *explored*, to find out whether they can potentially serve as a basis for new hypotheses to be tested by future research, and to test if eye-tracking has a broader applicability to T&I Studies.
- The results and findings from the two eye-tracking experiments will help to 'explore' *a posteriori* whether or not the predictions suggested by the Revised Hierarchical Model are valid at a *textual* level.

2.9.2 One Exploratory Functional MRI experiment (Language Combination: Mandarin as A – English as B)

- A preliminary fMRI experiment has been designed to *explore* if fMRI could be adequately applied to elicit brain activation patterns as well as behavioral data

[9] By 'other types of data', the present study refers to fixation frequency and blink frequency.

during A → B and B → A interpreting by novice interpreters, in order to *test* its applicability to T&I studies.

- The results and findings from the fMRI experiment will help to 'explore' *a posteriori* whether or not the predictions suggested by the Revised Hierarchical Model are valid at a *textual* level.

Chapter Three Methods: Eye-tracking

3.1 Introduction

This chapter aims to explain details of the two eye-tracking experiments. Basic biology of the eye is introduced. Details on how the eye-tracking laboratory was set up are described, because the laboratory was set up from ground up by the researcher at the Department of Humanities, Imperial College London. Design, data collection, and procedures of the two eye-tracking experiments are described.

3.2 Basic Eye Biology

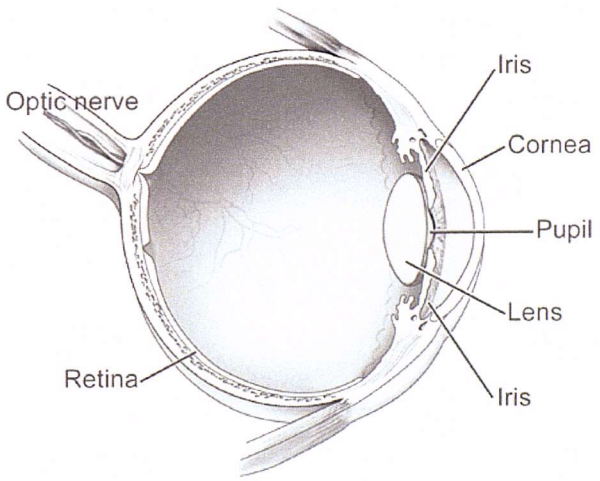


Figure 2. Basic Eye Biology

Source: (NEI, Online)

Figure 2 shows the basic anatomy of the eye: most of the surface of the back of the eye is made up of retina that contains cells responsible for detecting any stimulation of light

input into the eye and sending information to the brain via optic nerves. The pupil is an opening in the middle of the iris (the colored part of the eye) and it is the pupil that allows light into the eye. The amount of light it allows into the eye is controlled by the dilator and the sphincter muscles of the iris; therefore, the size of the pupil is related to the iris, which is like blindfolds that open and close. Lastly, the sclera is the white part of the eye.

Before light enters the pupil, the light will pass the cornea, which is like a thin, smooth film-like tissue that covers the eye. The cornea is basically transparent. However, when the light passes through the cornea, there is always a significant and detectable amount of reflection of the light passing through.

The front of the eye consists of the pupil, the iris and sclera (the white part of the eye). Light passes the cornea and enters the pupil until it reaches the retina. The eye serves as a retroreflector, which means that the eye reflects an amount of light back to where the light comes from via the same route. This means that the pupil normally appears black as the light entering the pupil rarely comes from the same point of observation. However, if the light or radiation comes from roughly the same point as observation, the light reflected can be observed and the pupil will become bright (for a more detailed introduction, see Tortora and Anagnostakos, 1990). An eye-tracker radiates non-invasive infrared light to brighten the pupil and detects and records the variations of pupil size over time during a cognitive task.

3.3 Principles of Eye-tracking Technology

Two fundamental principles, i.e. 'pupil discrimination' and 'cornea reflection discrimination' are introduced in this section, so as to illustrate how an eye-tracker detects eye movement data.

3.3.1 Pupil and Cornea Reflection Discrimination

The infrared light emanated from the eye-tracker's illuminator brightens the pupil and generates cornea reflection (CR), called Purkinje reflection, using edge detection logic

(Deb, 2003). During pupil and cornea reflection discrimination, a white circle appears on the edge of the pupil and the white crosshairs show the center of the pupil. Simultaneously, a black circle outlines the Purkinje reflection and the black crosshairs show the cornea reflection. Due to the geometrical structure of the eye, the distance between the center of the pupil and Purkinje reflection is constant. Therefore, the center of the pupil and CR are used as a reference points for an eye-tracker to detect eye movement data.

3.4 Optimal Eye-tracking Laboratory Setup

It is crucial to present how the eye-tracking laboratory was set up, because the entire laboratory was established from scratch by the investigator, with only a one-off tutoring visitation by the provider of the eye-tracker used in the study regarding wire assembly and eye calibration.

3.4.1 Laboratory Room

A room was secured at the Department of Humanities of Imperial College London exclusively for the eye-tracking experiments in the study. The room had one translucent ceiling-to-floor glass wall as well as an entrance door with a transparent rectangular window pane embedded in the middle of the door. In order to avoid any artificial lighting from outside the room, the wall and the window pane were fully covered by sheets of brown-colored, double-layered paper firmly taped together with duck tapes. A security sign was put up on the door to avoid any accidental entry into the laboratory during any ongoing experiment, and the sign read, '(1) This Room Has Been Reserved for Research Purposes! (2) Please Do Not Enter This Room without Permission! (3) Surveillance Camera at Work!' A chair was prepared and placed right in front of the entrance door during an ongoing experiment for the next subject to sit on while waiting, on the off chance that the ongoing experiment overran the allotted time. Figure 2 shows the laboratory room setup for the study.

Ideally, the duration of a whole experiment was less than one hour. Sometimes, multiple subjects were arranged for participation back to back, and a chair was provided outside the laboratory to (1) help the participants understand that they had identified and reached the correct venue for experimental participation; (2) offer them a chair to sit on while waiting; and (3) prevent them from opening the door while an experiment was still on-going. The chair was placed to also stop non-subjects (Imperial College students or staff) from accidentally opening the door in the middle of an experiment.

3.4.2 Light Meter

An individual's pupil size variations are extremely sensitive to any possible changes in lighting level (Morimoto *et al.*, 2000), however slight. Therefore, in order to ensure that the lighting level would stay constant both across participants and across tasks throughout all experimental sessions over the course of data collection, only two light bulbs were allowed on during all experimental sessions in the laboratory, and a light meter, i.e. Digital Light Meter 100,000 Lux LCD Lab Photo Camera, was used to measure the lighting level in the laboratory room. The lighting level of the room was consistently 305 lux. In this way, a potential confounding variable regarding lighting variations was avoided.

3.4.3 Eye-tracking System

An ASL (Applied Sciences Laboratory) 504 Model Monocular Eye-tracker was set up, via negotiations conducted jointly by the investigator and Dr. Gabriela Saldanha with Stephen Oliver Associates[10], UK, who loaned the system (both hardware and software) for the entire duration of data collection without any monetary request. In exchange, it was offered that the name of the company and contact information would be mentioned in all

[10] Contact information: Mr. Stephen Oliver, S. Oliver Associates, Northumberland House, Popes Lane, London, W5 4NG, UK. Phone: + 44 (0) 20 8840 9910, Fax: + 44 (0) 20 8840 9915, Mobile: + 44 (0) 7798 798 496, Email: oliverstephenb@aol.com, Websites: www.s-oliver-associates.com & www.eyetracker.co.uk.

future presentations or publications in connection with the doctoral study. Figure 3 shows the ASL 504 Model Monocular Eye-tracker set up and used in the study. The entire system includes two computers (one as an investigator interface and the other as a subject interface), a 504 infrared Pan/Tilt camera, video converter, control unit, and two mini monitors (eye movement monitor and subject screen monitor).

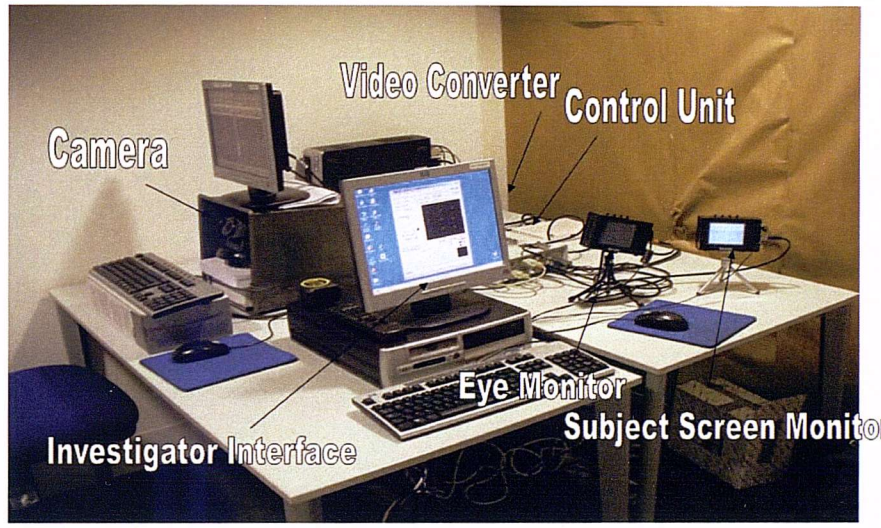


Figure 3. ASL 504 Model Monocular Eye-tracker and Laboratory Setup

The Pan/Tilt camera has an illuminator embedded right in the middle that emanates non-invasive infrared light. The infrared light level used for the ASL 504 Model is 0.04 mW (milliwatts) per square cm and the safe level for long term exposure is 10mW per square cm. A simple comparison is that if an individual is walking on the street on a sunny day, the eye is being exposed to 100mW per square cm (Oliver, 2007). Therefore, the Pan/Tilt camera and illuminator is non-invasive and has no known risks (Duchowski, 2003). This issue was presented to the Imperial College Research Ethics Committee for ethical approval, and was explained to all participants in the study.

The 'video converter' intercepts the on-going 'scene' (what is being displayed) on a subject's computer screen and sends the information to the control unit, which then feeds

the scene to one of the mini screens for investigator use. This requires use of a mini screen called 'subject screen monitor', which displays what is happening on a subject's computer screen during an experiment for the investigator to monitor and ensure the subject is carrying out a task as instructed. The other mini screen is called 'eye monitor', which displays the movement and calibration crosshairs of the subject's tracked eye during an experiment.

The control unit is equivalent to a central processing unit and performs all processes within the entire eye-tracking system, including scene transmission, pupil size calculation, gaze calculation, synchronization of all components involved, and camera movement control.

To sum up, all mechanical procedures and information regarding the laboratory setup and system installation were followed, as instructed by ASL's manual (ASL, 2006). Future researchers should refer to the manual to ensure replicability.

3.4.3 Sampling Frequency

The ASL 504 Model Monocular Eye-tracker has a sampling frequency of fifty hertz or Hz. Hertz is a unit of temporal frequency and fifty Hz means that a sample is collected every twenty milliseconds or ms. In the case of pupil size data sampled at a rate of 50 Hz in the study, for example, every twenty ms a sample regarding the diameter of the pupil of the tracked eye was collected and recorded. Therefore, for every second, there are exactly fifty data points regarding the pupil diameter collected and recorded. In addition, the data sampling frequency (Hz) is also used to refer to how frequently eye coordinates on the computer screen are sampled, and this information helps to determine where fixations take place.

3.4.4 Software Installation

A program 'Eye-Trac 6000' was installed on the investigator's terminal to control the

eye-tracker. Eye-Trac 6000 is downwardly compatible with ASL 504 Model Monocular Eye-tracker, so the software can control the eye-tracker in the study. Eye-Trac 6000 initiates the control unit's synchronizing of all components in the eye-tracking study, helps calibrate the eye and records data.

3.4.5 Camera Setup

According to the ASL manual (ASL, 2006) for the 504 Model, a number of considerations have to be taken into account when setting up the Pan/Tilt camera. When a participant is sitting in front of a computer, to optimize the visual scene captured by the participant's eyes, the distance between the camera and the eyes is optimally twenty four inches (Figure 4). In this way, the subjects will not be sitting too close to capture *all* that is presented on the screen or too far to *clearly* see the screen scene. The optimal distance will also allow the camera to track the eye properly. Therefore, a tape measure was employed to ensure that the distance between the camera and a participant's eye position was always kept at twenty four inches before s/he carried out any task.

Further, for the camera to track the widest horizontal visual field, it should be placed in centerline with the scene presented. For the camera to track the widest vertical visual field, it should be placed as close to the bottom area of the scene on the computer screen as possible. To collect the most robust data, the optics or lens of the camera should be on the same plane as the scene presented. When a subject is sitting in front of the computer and when s/he is gazing at the scene on the screen, his or her gaze should be on the same plane as the center point of the screen (see also **Section 3.4.6**). This is why a chair that is adjustable in height to meet different participants' sitting positions was installed in the laboratory.

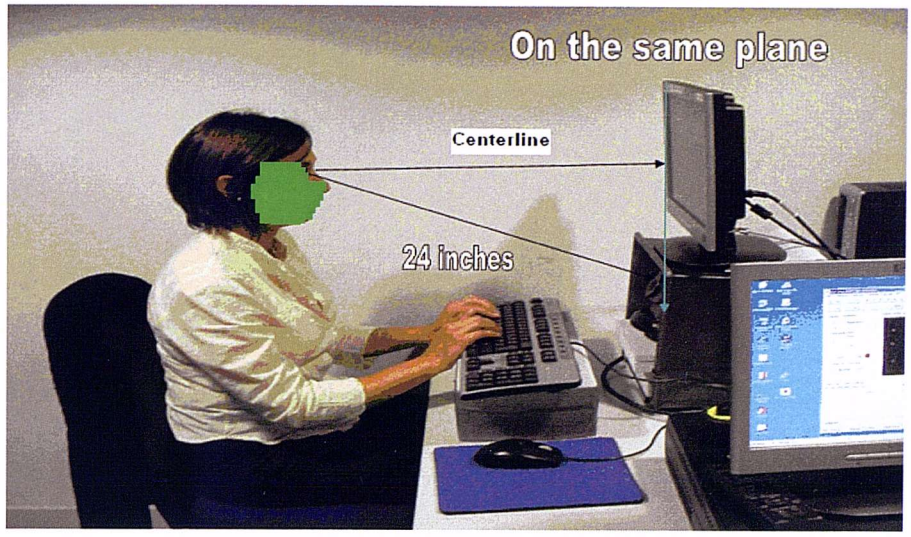


Figure 4. Camera Setup and Participant Placement

Due to confidentiality, the face of the subject was smudged.

3.4.6 Eye Calibration

For eye calibration, a bitmap file (background color: gray) was created and carried nine reference points. The background color was set as gray because of the following reasons. First, if the background color had been white (the regular color of a computer screen), the lighting level would have been too strong and this would have affected a participant's pupil size (e.g. constant constriction). Second, the background color of the calibration map was gray because it had to be consistent with the background color of the split work windows (see **Section 3.7**) during the experiment to avoid lighting inconsistency to which the pupil is extremely sensitive.

The calibration map was displayed on the user monitor. The nine reference points were set in a way that point '5' was in the center of the computer screen and the rest of the points were close to the periphery of the scene. All the nine points were manually synchronized with the nine points preset on the Eye-Trac 6000 interface on the investigator's computer.

Eye calibration involved the following steps: Once seated in front of the computer, a

participant was instructed by the investigator to stare at point 5 to ensure that pupil and cornea reflection discrimination was completed. Afterwards, the investigator instructed the participant to stare at the nine numbers one at a time on the screen 'sequentially' and to fix his/her gaze on a particular number until they received the investigator's instruction to move the gaze to the next number. This allowed the investigator to calibrate the subject's eye coordinates on the scene. Once the calibration was completed, the investigator would then 'randomly' select numbers on the user monitor and ask the participant to stare at them so as to double-check whether or not the subject's eye movement corresponded with the movement of the crosshairs on the 'subject scene monitor'. This was a 'precision check' because the investigator needed to ensure that the crosshairs on the 'subject scene monitor' landed on the number that the investigator randomly requested the participant to stare at. This concluded the process of eye calibration.

3.4.7 Testing Calibration

Seven people volunteered to sit in front of the user computer to help the investigator to practice calibrating their eyes. It was observed that it was sometimes difficult to carry out pupil and cornea reflection discrimination if a participant was wearing glasses. This is because the frames, when slightly sliding down the nose, could create a false CR, generating erroneous eye coordinates or even rendering the eye undetectable. To prevent this, an elastic band for glasses was provided for participant use, should there be any sliding of their frames. It was also noticed that if the diameter of contact lenses worn by a participant was too small, the rim of the contact lens would often be mistaken by the eye-tracker as that of the pupil, which would cause problem to precise eye calibration. In this case, the participant had to take off his/her contact lenses and put on glasses. Therefore, when booking an appointment with participants, the investigator asked all the participants wearing contact lenses to bring along their glasses to the laboratory. A contact lenses container and a bottle of saline solution were provided for use in the laboratory. In addition, it was also discovered that, in elderly participants, their eyes were not wide open enough for accurate eye

calibration as their eyelids could sometimes cover a large portion of the iris. Therefore, all participants, regardless of their age group, were instructed to try and open their eyes as widely as possible without over-flexing their facial muscles. The participants were allowed to blink their eyes comfortably. Another issue that was encountered by the researcher during testing eye calibration was that, when a participant was wearing heavy make-up, the make-up could potentially generate light reflection that could be construed by the eye-tracker as the cornea reflection. In light of this, the researcher asked all the participants not to wear make-up during the experiment. The last problem was that many of the participants would smile or laugh for various reasons, e.g. being slightly nervous about participating in an experiment, making a mistake during a practice session, or being unsure of what was expected of them. However, when a participant smiled or laughed, her/his eyes would squint, and this would make it very difficult for the camera to track her/his eye. To cope with this issue, the researcher firstly tried to help the participants feel at ease by conducting a casual chat irrelevant to the study, followed by explaining what the experiment was about and asking the participants to refrain from smiling or laughing throughout the course of the entire experiment to ensure eye calibration and data collection with a better quality.

On the Eye-Trac 6000 interface controlled by the researcher, once eye calibration is completed, the function of eye tracking can generally be put on AUTO for automatic eye-tracking by the camera. However, due to the nature of the task (translate vs. type; see **Section 3.5.1**) that invariably involved head movement or looking down at the keyboard to search for the correct keys to press for text production, it was observed that AUTO eye tracking function was sometimes problematic. This is because the movement of the camera to track a participant's eye was not as swift as the person's head movement. To tackle this issue, a remote control accompanying the camera was used, so that if the camera failed to track the eye accurately on AUTO, the researcher could fall back on the remote control to help the camera to track the eye accurately. Sometimes, the camera could auto-track the eye throughout an experimental session. Occasionally, AUTO and manual eye tracking were used in a mixed fashion, depending on whether or not the camera was tracking the eye properly. Occasions arose where the researcher had to resort to the remote control to track

the eye throughout all experimental sessions. As using a remote control required the researcher to move his hand relatively closer to a participant, this possible scenario was explained in detail to all the participants and a number of mock situations were carried out, so that should this happen in the middle of an experimental session, the participants would not feel perturbed or distracted.

3.5 Experimental Designs

This section introduces criteria involved in text selection for the two eye-tracking experiments approved by the Imperial College Research Ethics Committee (ICREC, Online), the concept of ecological validity and definition of each variable measured in the study.

3.5.1 Designs

3.5.1.1 Mandarin (A) – English (B) Eye-tracking Experiment

To test if there is any significant difference between cognitive loading for A→B and B→A translation using eye-tracking, an experimental study was designed. The Mandarin (A) – English (B) experiment included two 'factors', or independent variables. In an experiment, 'levels' refers to different conditions of a factor. These factors were both within-subject factors, meaning they were designed to collect data from the same subject on various tasks within a specific period of time. This type of design is referred to as a 'repeated measures' experimental design, as the data samples from the same person collected within the same period of time are repeatedly measured and recorded. An advantage of repeated measures experiments is that, as many samples can be obtained from the same person, they do not require many participants (Craig, 2007: 2), so long as the participants recruited are representative of the population in question (Goodwin, 1995). In the first eye-tracking experiment that involved the language combination of Mandarin (A) – English (B), one factor is Input Language (two levels: Mandarin Chinese vs. English) and the other factor is

Task Type (two levels: Translating vs. Typing). Therefore, the experiment was a 2 x 2 (read as two by two) repeated measures within-subject factorial design. To illustrate, if a researcher wishes to investigate if the strength of lighting has an effect on pupil size of a healthy person, s/he may design an experiment with one factor (lighting) with different levels (a room with fifteen light bulbs on vs. ten bulbs on vs. five bulbs on, and this generates three levels of the factor) (for a detailed discussion on experimental designs, see Schweigert, 1994; Shaughnessy *et al.*, 2000). Regarding criteria of participant selection for both eye-tracking experiments in the study, given practical and geographical constraints, the population the study aimed to investigate was specific and limited. Non-simultaneous bilinguals (see **Section 2.6**) who acquired English as a second language and completed a one-year master's program in translation as novice translators in the UK were approached and recruited.

In the first eye-tracking experiment, a typing task was introduced as a control task because Mandarin Chinese is a square-shaped, ideogram-based (or logographic) and non-alphabetic language and this might contribute to extra cognitive loading (Li and Green, 2007; Jeong *et al.*, 2007). By introducing the typing task, the study was able to explore if Input Language and Task Type would have an effect on 'dependent variables' or variables which might vary as a function of the independent variables.

3.5.1.2 Spanish (A) – English (B) Eye-tracking Experiment

The second eye-tracking experiment involved using Spanish native speakers with English as their second language. The introduction of the second eye-tracking experiment had three main merits. First, some of the concerns discovered by the experimenter during the first eye-tracking experiment could be addressed (see **Chapter Five**). Second, the results of the second experiment with the language combination of Spanish (A) - English (B) could be used to further verify whether or not the first Mandarin (A) - English (B) eye-tracking results were generalizable to an alphabet-based language combination.

In light of the experience obtained from the first eye-tracking experiment, 'reading' was introduced as an extra task, alongside typing and translating. 'Reading' was

incorporated as an additional level within the factor of Task Type. In this case, six cognitive tasks were involved, i.e. (1) Read Spanish; (2) Read English; (3) Type Spanish; (4) Type English; (5) Translate from English into Spanish; and (6) Translate from Spanish into English. Since there were three levels of Task Type and two levels of Input Language (Spanish vs. English), the experiment used a 3 x 2 within subject repeated measures factorial design.

3.5.2 Material Design and Selection

Text selection was a critical aspect of the study design. Several elements were considered, in terms of criteria for text selection. First, the study aimed to achieve as much 'ecological validity' as possible. 'Ecological validity' refers to how well an experimental task reflects the *real-world* situation under study and how well the experimental results can be generalized to the *real-life* scenario under investigation. Shlesinger (2000) argues against decomposability of translation and interpreting into components, subcomponents and sub-subcompenents for interpreting research in an experimental setting. Setton (2002) contends against 'decontextualized' experimental assumptions about 'task modularity' of conference interpreting and criticized Fabbro *et al.*'s (1991) dichotic listening research as a paradigm performed on activities unrelated to the nature of interpreting. This argument boils down to 'ecological validity', a bone of contention addressed by a number of T&I researchers (Setton, 1999; Liu, 2001; Chang, 2005). Therefore, to *maximize* ecological validity of the study, controlled written 'texts', as opposed to 'single words', were used as stimuli to *simulate* real-life situations where translators are requested to conduct textual translations. Nevertheless, no claim of having achieved perfect ecological validity can be made in relation to this study because the fact that the texts were selected and controlled and because the study was conducted in a laboratory setting that could have a potentially adverse effect on the level of ecological validity.

In addition, since the research program presented in this thesis was interested only in investigating 'cognitive loading' of 'translation directions', simple and non-technical texts

were selected. Therefore, it was necessary to exclude possible text-related confounding variables, so the simple, non-technical texts were tested to the extent that they were comparable in grade level, word count, comprehensibility, readability and 'translatability'. It is worth pointing out that the use of the 'translatability test' was an improvement on the study by Pavlović and Jensen (2009), because the test further ensured text comparability.

3.5.2.1 Mandarin (A) – English (B) Eye-tracking experiment

Two English texts from one children's story book (Milne, 1980) were selected, one for translating (Appendix A) and the other for typing (Appendix B); two Mandarin texts were selected from a Mandarin textbook used in Taiwan (Lai, 2006), one for translating (Appendix C) and the other for typing (Appendix D).

Grade Level: 'Grade level' refers to educational stages in a system of grades. The two selected English texts scored 5.5 and 4.5 respectively on the Flesch-Kincaid Grade Level Test (available on Microsoft Word 2002). This means that the two English texts were usable and understandable by pupils in grade four or five in an elementary school in the US educational system. As Flesch-Kincaid Grade Level Test could not score Mandarin texts and no test was available to score Mandarin texts, a Mandarin textbook used in the second semester of grade 4 in Taiwan[11] was obtained (grade level = 4.5). Two short texts were extracted from the textbook to ensure that they were comparable with the English texts in terms of grade level. A low grade level minimized text-related 'difficulty' factor in the study.

Word Count: All four texts consisted of fifty 'words' in total. The concept of an orthographic 'word' as understood in the English language is different from the idea of 'word' in an ideogram-based language as Mandarin, where 'words' are often composed of two 'characters', forming a 'compound' unit that is equal to a word unit in English. For example, '電' as a Chinese character means 'electricity' and '腦' means

[11] The educational system in Taiwan is modeled on the US educational system. Therefore, it is argued that the grade levels of the two countries are comparable.

'brain'. Juxtaposing the two characters would yield a compound word of '電腦' (meaning 'computer'). Therefore, to control word counts of all the texts, two Mandarin language and translation teachers with Mandarin as their first language and English as their second agreed on the total word counts of all the four texts after modifications. First, they used Microsoft Word 2000 to compute the word counts of the two English texts. To ensure that they composed fifty words exactly, a number of English words were deleted or changed without affecting grammar and meanings of the texts. Second, the two teachers individually calculated the word counts of the two selected Mandarin texts. After careful deliberation and modifications without affecting grammar and meanings of the texts, they reached a consensus that the two Mandarin texts consisted of fifty 'words'. By doing so, the potential effect of 'word count' was *minimized.*

Comprehensibility: Ten bilinguals [12] with Mandarin as their first language and English as their second were instructed to read and rate all the four texts in terms of comprehensibility level, using a scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. They were asked to rate the texts, based on how easy it was for them to understand the texts. Table 1 shows that all four texts were approximately rated at 1.6 or 1.7. This indicates that the four texts were fairly easy to comprehend and minimized the potential factor of text-related variable of 'comprehensibility'.

[12] The numbers of the people who 'volunteered' for the different tests varied from test to test, due to practical constraints, such as people's willingness to volunteer and the point in time when a given test was administered.

	N	Minimum	Maximum	Mean	Std. Deviation
English Text for Translation	10	1	2	1.6	.516
English Text for Typing	10	1	2	1.7	.483
Mandarin Text for Translation	10	1	2	1.7	.483
Mandarin Text for Typing	10	1	2	1.6	.516
Valid Number of Participants	10				

Table 1. Comprehensibility Test.

Readability: The same ten bilinguals with Mandarin as their first language and English as their second were asked to read and rate all the four texts in terms of readability level, using the same scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. They were instructed to rate the texts, based on how easy it was for them to read the texts. It is worth mentioning that 'comprehensibility' differs from 'readability'. For example, a newspaper advertisement that is succinct yet full of jargon may be easy to comprehend but difficult to read. Contrarily, a high school literary piece may be easy to read and yet difficult to comprehend. Table 2 shows that all the four texts were approximately rated between 1.5 and 1.7. This shows that the four texts were fairly easy to read and minimized the potential factor of text-related variable of 'readability'.

	N	Minimum	Maximum	Mean	Std. Deviation
English Text for Translation	10	1	2	1.6	.516
English Text for Typing	10	1	2	1.6	.516
Mandarin Text for Translation	10	1	2	1.5	.527
Mandarin Text for Typing	10	1	2	1.7	.483
Valid Number of Participants	10				

Table 2. Readability Test

Comprehensibility and readability tests in the study were considered generic tests that could be administered to non-translator sequential bilinguals; therefore, to reiterate, given the practical constraints, as many recruitable bilinguals as possible were requested to conduct the tests. However, 'translatability' test could only be carried out by qualified translators that were difficult to come by. Cohen (1960: 37) argues that at least 'two' judges are needed to 'determine the degree, significance, and sampling stability of their agreement'. Therefore, two evaluators were approached to complete the translatability test, again given the time and practical constraints.

Translatability: In the study, 'translatability' was defined as how difficult or easy it was for a qualified translator to translate a text, based on their translation expertise. It has often been assumed that the fact that a text is easily readable and comprehensible and used in an elementary level means that it is easy to translate.

However, this is not necessarily the case. Some children's literary works are cases in point. In order to tease out another confounding variable, 'translatability', another test was administered. Two professional translation teachers with Mandarin as their first language and English as their second were asked to read and rate all the four texts in terms of translatability level, based on the scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. Both of the professional translation teachers gave 1 to all the four texts, and Cohen Kappa coefficient that represents inter-rater agreement was 1. This indicated almost perfect agreement (Landis and Koch, 1977: 167-171) and suggested that the texts were very easy to translate.

3.5.2.2 *Spanish (A) – English (B) Eye-tracking experiment*

The same criteria for text selection were applied to the Spanish (A) – English (B) eye-tracking experiment. Results of the tests are reported below.

The two English texts employed in the first Mandarin (A) – English (B) eye-tracking experiment were also used in the second experiment. Regarding Spanish texts, two Spanish texts were selected with the help of a Spanish language teacher (also a Spanish native speaker) (Mialdea, 2007), one for translating and reading (Appendix E) and the other for typing (Appendix F).

Word Count: All the four texts consisted of fifty orthographic words in total. As both Spanish and English are alphabet-based languages, the function of word count in Microsoft Word 2002 was used to ascertain the word counts of the four texts.

Grade Level: The two selected English texts scored 5.5 and 4.5 respectively on the Flesch-Kincaid Grade Level Test. This means that the two English texts were usable and understandable by pupils in grade four or five in an elementary school in the US educational system. As Flesch-Kincaid Test could not score Spanish texts and there was no available test to score Spanish texts, a Spanish book used in the second semester of grade 4 in Spain was obtained (grade level = 4.5) (Mialdea, 2007). Two

short texts were extracted from the book to ensure that they were comparable with the English texts in terms of grade level. By doing so, a low grade level minimized text-related 'difficulty' factor.

Comprehensibility: Given the time and practical constraints, as many as thirteen bilinguals with Spanish as their first language and English as their second volunteered to read and rate all the four texts in terms of comprehensibility level, based on the scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. Table 3 shows that all the four texts were approximately rated between 1.69 and 1.62. This shows that the four texts were fairly easy to comprehend and minimized the potential factor of text-related variable of 'comprehensibility'.

	N	Minimum	Maximum	Mean	Std. Deviation
English Text for Translation	13	1	2	1.62	.506
English Text for Typing	13	1	2	1.69	.480
Spanish Text for Translation	13	1	2	1.69	.480
Spanish Text for Typing	13	1	2	1.62	.506
Valid Number of Participants	13				

Table 3. Comprehensibility Test

Readability: The same thirteen bilinguals with Spanish as their first language and English as their second were asked to read and rate all the four texts in terms of readability level, based on the scoring system of 1 = very easy, 2 = easy, 3 = medium, 4

= difficult, 5 = very difficult. Table 4 shows that all the four texts were approximately rated between 1.69 and 1.62. This showed that the four texts were fairly easy to read and minimized the potential factor of text-related variable of 'readability'.

	N	Minimum	Maximum	Mean	Std. Deviation
English Text for Translation	13	1	2	1.62	.506
English Text for Typing	13	1	2	1.69	.480
Spanish Text for Translation	13	1	2	1.69	.480
Spanish Text for Typing	13	1	2	1.62	.506
Valid Number of Participants	13				

Table 4. Readability Test

Translatability: Two professional translation teachers with Spanish as their first language and English as their second were asked to read and rate all the four texts in terms of translatability level, based on the scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. Both of them gave 1 to all the four texts. The Cohen Kappa coefficient was 1, and this indicated almost perfect inter-rater agreement (Landis and Koch, 1977; 167-171). This suggests that the texts were very easy to translate.

3.5.3 Definitions of Dependent Variables

The study aimed to measure such dependent variables as (1) pupil diameter, (2) fixation count, (3) task time, (4) fixation frequency, (5) blink frequency, and (6) fixation duration/time. Definition of each dependent variable is introduced in this section. It is worth noting that multiple dependent variables were collected to *converge* different kinds of data, and this would help *strengthen* the potential results and findings presented in this thesis (for a detailed discussion, see Stawarski and Phillips, 2008).

3.5.3.1 Pupil Diameter

'Pupil diameter' is defined as 'the measure through the center of the adjustable opening in the iris of the eye, terminated at both ends by its circumference' (COGAIN, Online); it has been used as a measure of cognitive loading (Rayner, 1998; Rayner *et al.*, 2006). In the study, with the lighting level of the laboratory room kept constant as well as the background color of the computer screen preset as always gray, the dependent variable of pupil size was measured. The eye-tracker recorded data on pupil diameter variations over time during a task. The unit of the pupil diameter data was in millimeters. The pupillometric data was also used to test whether or not a 'familiarity effect' existed during *written translation*, as previously stated in **Section 2.8**.

3.5.3.2 Fixation Count

In a generic reading experiment, a fixation is normally defined as a gaze at a particular point for 200 to 300 milliseconds (ms) (Rayner, 1998). A number of people define a fixation as a gaze at a particular point between 200 to 600 ms, on the assumption that the more cognitively demanding a task is, the longer a fixation it will entail (Majaranta, Online: 6). 'Translating' is obviously more cognitively demanding than mere 'reading', because it involves reading comprehension, language conversion and text production. Given a

relatively higher complexity level of 'translating', the study thresholded the definition of a fixation at 500 ms, a level that was higher than 200 to 300 ms often associated with a fixation during reading (Rayner *et al.*, 2006). It is worth mentioning that, before finalizing the threshold of 500 ms for a fixation in the study, testing was carried out in advance, using 100 ms, 200 ms, 300 ms, and 400 ms as thresholds. However, the testing yielded extremely cluttered fixation plots, whereas 500 ms generated clear fixation plots across all participants. Based on Majaranta's argument (Online) as well as the test results, it was decided that 500 ms was to be employed as the threshold for a fixation in the present study. Further, it has to be kept in mind that, in any given eye-tracking experiment, so long as the fixation threshold remains constant across all statistical analyses, the definition of the threshold will not adversely affect the results.

As the system in the study was fifty Hz in frequency, a fixation consisted of at least twenty-five samples or instances of gazing that stayed in a fixed point on the computer screen. The fixed point was determined by the crosshairs (indicating coordinates on the subject screen) recorded by the Eye-Trac 6000 system.

3.5.3.3 Task Time(sec)

Time spent on completing each task by each participant was recorded in the study. Such data is referred to as 'task time', and its unit is in seconds (sec). Task time has been demonstrated to be a useful indicator of cognitive loading of a task (Pavlović and Jensen, 2009).

3.5.3.4 Fixation Frequency (decimal numbers)

'Fixation frequency' refers to the frequency at which fixations appear on a particular area, and has been demonstrated to be a useful measure of cognitive loading (Orden *et al.*, 2001). The algorithm involved in computing 'fixation frequency' is temporal definition of a fixation in a study (500 ms in this study) divided by the number of fixations that take place

in an area in question (COGAIN, 2008).

It has to be acknowledged that there has been a controversy regarding how 'fixation frequency' is compute. For example, Radach and Kennedy (2004: 6) explicitly defined fixation frequency as 'mean absolute number of fixations per word, for the current pass'. However, the algorithm for the computing of fixation frequency already built in the software program named EYENAL used in the present study adopted the COGAIN's definition. Using EYENAL, values regarding fixation frequency were automatically yielded for data analysis, and the researcher had no control over how to compute fixation frequency for the study. Even so, at the very least, COGAIN's computation can still to a certain extent create useful values for statistical analyses. When the numerator within a fraction stays constant, whilst the denominator varies, the values generated are still useful for comparisons. In the study, 500 ms was the constant numerator within the fraction, whereas the denominator changed whenever the total fixation counts for different conditions of different participants varied. The EYENAL in-built computation method created different values for different conditions for different subjects, and the computation remained constant across all conditions and subjects, and this computation therefore at the very least generated a certain extent of value and usefulness for further analyses and did not negatively affect the statistical analyses in the study. However, it is worth noting that future researchers should select a relatively more suitable definition of fixation frequency.

3.5.3.5 Blink Frequency (decimal numbers)

'Blink frequency' refers to how frequently the eye blinks during a task. When an individual blinks, the value(s) of pupil size sample(s) during blinking recorded by Eye-Trac 6000 will be zero(s). Upon completion of the task, the total count of 'zeros' divided by the total number of samples obtained will yield 'blink frequency'.

3.5.3.6 Fixation Duration/Time(sec)

Once the fixation count is established, the total duration of the fixations on a particular area can be computed in a unit of seconds; fixation duration has been shown to be a reliable measure of cognitive loading (Rayner, 1998; Rayner *et al.*, 2006). For example, three fixations appear in a specific area under investigation, and their respective durations are 600 ms, 650 ms, and 750 ms. In this case, total fixation duration or fixation time is two seconds.

3.6 Data Collection

3.6.1 Participants

All the subjects participated with informed consent (Appendices G and H), and received a cash honorarium of ten pounds, alongside an acknowledgement letter issued by the Department of Humanities, Imperial College London. The letter confirmed that they had participated in the study.

3.6.1.1 Mandarin (A) – English (B) Eye-tracking Experiment

Participant recruitment emails were sent to Imperial College London, University of Westminster, University of Bath, University of Newcastle upon Tyne, Middlesex University, City University, University of Leeds, School of Oriental and Asian Studies, University College London, and University of Manchester in the United Kingdom. Sixteen healthy, right-handed, female subjects with a language combination of Mandarin Chinese as the first language and English as the second language who had completed a one-year postgraduate translation and/or interpretation program at one of the three universities (Imperial College London, University of Westminster or Middlesex University) participated. The data from

one subject was excluded from analyses because she did not follow the investigator's instructions. Therefore, data from fifteen participants were analyzed.

3.6.1.2 Spanish (A) – English (B) Eye-tracking Experiment

For the Spanish (A) – English (B) eye-tracking experiment, recruitment emails were sent to all the universities from which the researcher previously sought to recruit participants for the Mandarin (A) – English (B) eye-tracking experiment. Further, TransPerfect Translations [13] was contacted to help recruit potential participants. Eight healthy participants with a language combination of Spanish as the first language and English as the second language who completed their one-year postgraduate translation M.A. program only at Imperial College London, UK, replied and participated with informed consent.

3.7 Stimuli Presentation

Two professional translators' opinions were sought regarding how the windows should be displayed in order to simulate the way they would normally translate, and both reported that they usually had their source text on their left hand side to read, while working on the computer screen on the right hand side for text production. The window environment in the participants' monitor was arranged so that there were two split windows side-by-side, with the text (presented or to be produced) on the top of the window because this would help the participants to keep their eyes open (previously discussed in **Section 3.4.7**) to facilitate precise eye-calibration. Regardless of the task type, the source text was always presented in the split window on the left-hand side, whereas the window on the right was used for target text production. The text alignment was set as 'align to the left'. The font of both source and target texts were set as size 14 (so that words or characters were

[13] Ms. Idalina Gomes, Project Manager, TransPerfect Translations, 120 Fenchurn Street, 4th Floor, London, EC3M 5BA. Tel: + 44 (0) 20 7398 8200, Fax: + 44 (0) 20 7398 8202. Email: London@transperfect.co.uk.

large enough to be easily visible), bold, and black in color. The background colors of the two split windows were switched to gray, a color consistent with that of the calibration map. This prevented light-sensitive pupils from sudden constriction when the participants embarked on the tasks after eye calibration. Figure 5 shows how the split windows and stimuli were displayed during the experiments. Further, where the Mandarin (A) – English (B) participants were concerned, they were all native speakers of Mandarin Chinese, but came from different parts of the Mandarin-speaking world[14]. The format of Mandarin used for source text display and target language production was adjusted accordingly, based on the Language and Translation Background Questionnaire they filled out (see **Section 3.12**), in order to avoid extra cognitive loading for having to produce the other form of Mandarin they were not familiar with.

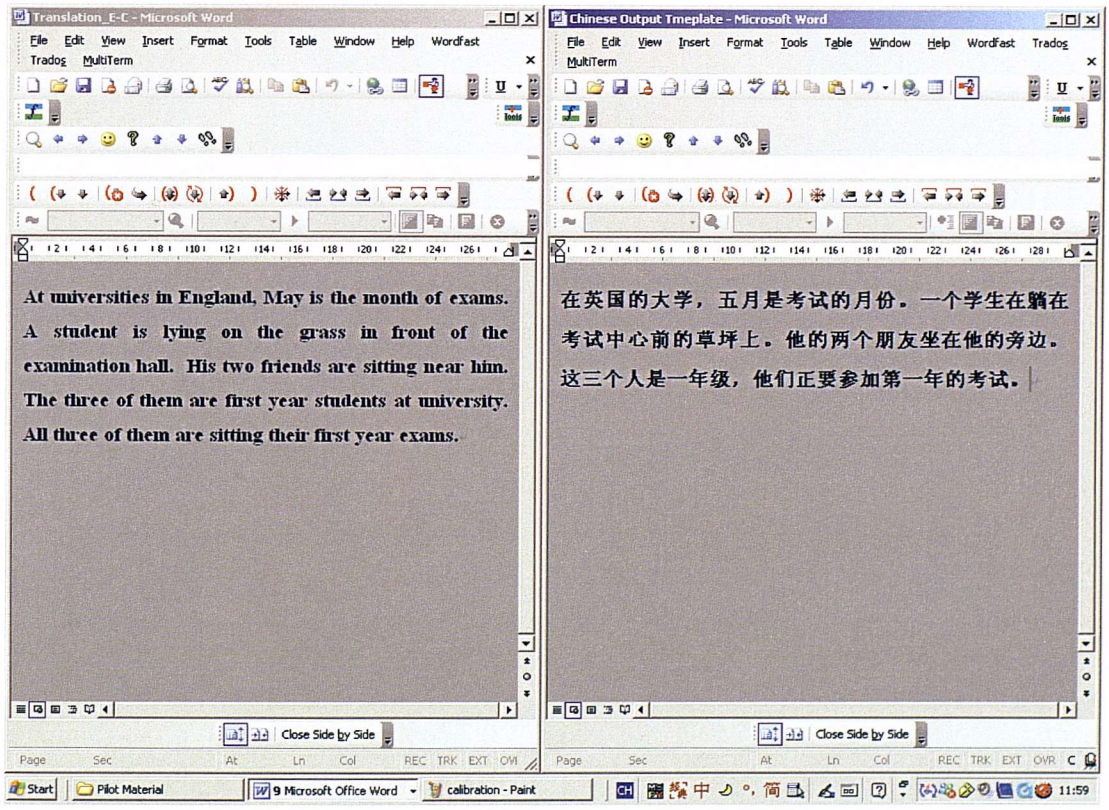


Figure 5. Split Window for Stimuli Presentation and Text Production

[14] Taiwan and Hong Kong use traditional Mandarin Chinese whereas Mainland China and Singapore use simplified Mandarin.

3.8 Pilot Testing

Once the laboratory and the experiment was set up, five non-translator bilinguals with Mandarin as their first language and English as their second volunteered to pilot run the flow of the experiment, so as to ensure that the investigator would be fully competent with dealing with all potential experimental details involved and that contingency plans were put into place, should anything go wrong either with the subject or the equipment.

3.9 Task Sequence

3.9.1 Mandarin (A) – English (B) Eye-tracking Experiment

For each participant, there are in total four tasks in one experimental session: (1) translate from English into Mandarin; (2) type English; (3) translate from Mandarin into English; and (4) type Mandarin. The task sequence was randomized across subjects. A random permutation table was produced (Appendix I), using Excel 2002, and the randomization sequence was based on the table. Further, before a participant was instructed to carry out a particular task, the text selected for that task was presented to the participant to read out aloud once. This was done because a translator is generally recommended to read a text to be translated in advance before s/he starts to translate it. The participants were requested to read aloud, so the researcher could ensure that (1) they read *once*; and (2) they read the *entire* text.

3.9.2 Spanish (A) – English (B) Eye-tracking Experiment

Based on results yielded by the first eye-tracking experiment (see **Chapter Four**), task sequence for the second eye-tracking experiment was adjusted as follows: (1) Read Spanish; (2) Read English; (3) Type Spanish; (4) Type English; (5) Translate from English into

Spanish; and (6) Translate from Spanish into English. Factors such as fatigue (Krichmar *et al.*, 1997), familiarity and practice cause constriction of the pupils. Therefore, if there were a fatigue, familiarity or practice effect, the very last task, i.e. 'Translate from Spanish into English' would be most strongly affected, as the effect would potentially decrease pupil diameter especially for 'Translate from Spanish into English', which would increase the conservativeness of the design.

3.10 Handedness

3.10.1 Mandarin (A) – English (B) Eye-tracking Experiment

On-line Edinburgh Handed Inventory or EHI (CSE, Online) was administered to determine the handedness of all participants. On EHI, handedness can be revealed on a numerical scale between the value 100 (100% right-handed) and -100 (100% left-handed). Table 5 shows that all fifteen participants were right-handed, which was consistent with the self-reported handedness results in Language and Translation Background Questionnaire (see **Section 3.12**).

	N	Minimum	Maximum	Mean	Std. Deviation
Handedness %	15	33.0	100	89.640	18.2650
Valid Number of Subjects	15				

Table 5. Descriptive Statistics on Handedness of the Mandarin – English Participants

3.10.2 Spanish (A) – English (B) Eye-tracking Experiment

Two of the participants were left-handed and the other six were right-handed. Table 6 shows the descriptive statistics of the overall handedness of the eight subjects.

	N	Minimum	Maximum	Mean	Std. Deviation
Handedness %	8	-100	100	52.50	88.115
Valid Number of Subjects	8				

Table 6. Descriptive Statistics on Handedness of the Spanish – English Participants

3.11 Ocular Dominance

'Ocular dominance' refers to an individual's eye preference for visual input. An ASL Model 504 Monocular eye-tracker was used in the study, so only one eye for each participant could be tracked. It is worth clarifying that eye dynamics or vergence that involves both eyes, i.e. the inter-ocular adjustment to focus at different depths, were not relevant for the purposes of this study. In addition, Niehaus *et al.* (2001) reported significant agreement between the degree of pupillary diameter for both eyes and that eyedness or handedness did not have a significant effect on the pupil data of both eyes. Even so, for the sake of selecting one of the eyes to track in the study, a test for ocular dominance or 'eyedness test' named Porta Test (Slowinski, Online) was administered to determine the dominant eye or ocular dominance for each subject for eye calibration and data collection in the study. Based on the Porta Test, first, a participant is instructed to use the index finger of his/her right hand to point at an object visible far afield with the entire right arm straightened and 'both' eyes open until s/he reported to the researcher that s/he is pointing at the object. Second, the individual is instructed NOT to move his/her arm. Third, the researcher instructs the

participant to close his/her right eye (leaving the left eye open and arm still unmoved) and to report to the researcher whether or not the object has removed from the point of the index finger. Fourth, the researcher asks the participant to open both eyes, close his/her left eye (leaving the right eye open and arm still unmoved) and to report whether or not the object has removed from the tip of his/her index finger. If the object has not removed from the tip of the individual's index finger when a participant closes his/her right eye, then his/her ocular dominance is right-eye; the same rule is applicable to determining if someone is left-eyed.

3.11.1 Mandarin (A) – English (B) Eye-tracking Experiment

Twelve participants were right-eyed, and the other three participants had a left-eyed dominance. During the experiment, their dominant eyes were selected accordingly for calibration, tracking and data collection.

3.11.2 Spanish (A) – English (B) Eye-tracking Experiment

Three of the participants showed a left-eyed preference for visual input, and the other five were right-eyed.

3.12 Questionnaire Information

A Language and Translation Background Questionnaire (Appendices J and K) was administered, and its design was based on methodological recommendations by Chang (2005) for investigating the population of translators or interpreters using an experiment that yields *quantitative* data. In such a questionnaire, specific questions regarding the participants' language and translation background that are normally considered factors impacting on translators'/interpreters' or general language workers' performances are included to obtain further *qualitative* details of the participants, and the information describes variability between the individuals involved (Brown, 1988; Bradburn *et al.*, 2004;

Chang, 2005). The aim is that, if post-analysis *quantitative* results appear problematic, the *qualitative* information will be helpful in interpreting, clarifying or explaining the *quantitative* results (e.g. **Sections 4.6.2** and **5.5.1**). The information obtained via the Language and Translation Background Questionnaire used in the present study is reported as follows:

3.12.1 Mandarin (A) – English (B) Eye-tracking Experiment

The fifteen subjects were right-handed females. Ten of them were from the People's Republic of China; three from Taiwan; and two from Hong Kong. In terms of age, nine of the fifteen subjects were aged between eighteen and twenty five years old, whereas six of them had an age range between twenty six and thirty five.

In terms of language use, five of the fifteen subjects self-reported that their language of habitual use was Mandarin Chinese; two indicated English as their language of habitual use; and eight reported that Mandarin Chinese and English were languages of habitual use in equal measure. Twelve of the fifteen participants reported that Mandarin Chinese was more difficult to type on an English-based computer keyboard, and the other three found English more difficult to type. Thirteen out of the fifteen subjects used an input system that was alphabet-based, whereas the other two used an input system that was based on radicals. To illustrate, a radical-based input method for Mandarin Chinese deals with the orthography of a Chinese character. It requires an individual to internally transform the orthography of a Chinese character into radicals (like pieces of a jigsaw puzzle) that form the character, and the radicals are designed onto an English keyboard with no reference to the English alphabet. Therefore, the individual will have to press the right keys carrying the correct radicals without having to key in a number to represent the tone of the character. If the individual manages to key in all the correct radicals for a Chinese character, the exact Chinese character will simply show up on the screen, without having to choose from a list of homophones. A radical-based input method comes in different names. Based on the questionnaire results, the names of the alphabet-based input method were 'Cangjie' and 'Wuxiami'. The participants using one of the radical-based methods had to undergo the

above process.

In essence, the participants who used an alphabet-based input method for Mandarin Chinese had to decompose the sound of a Chinese character into alphabetical representations and its tonal property, press the right keys in an English keyboard, and had frequently to deal with a list of homophones from which to select the correct character for use. In contrast, those who used a radical-based input system only had to decompose a character into radicals, press the right keys that carry the correct radicals, and generally did not have to deal with choosing from a list of homophones.

Four out of the fifteen subjects could not speak any other dialect or any other language. Four of them could speak one other dialect or another language. Six of them could speak two dialects or languages. One of them could speak three dialects or languages. In addition, fourteen of the fifteen participants self-reported that they were more comfortable with translating from English into Mandarin Chinese (B→A) translation, whereas only one of them felt equally comfortable with both translation directions.

Further, eight of the participants have a B.A. degree related to language or translation, whereas the other seven subjects' first degrees were unrelated to those fields. In terms of scoring on the International English Language Testing System or IELTS (9.0 is the highest score of the IELTS test), one of the fifteen subjects scored 6.0; five scored 6.5; three scored 7.0; two scored 7.5; and four scored 8.0. Ten of the participants were from the Imperial College London; three from the University of Westminster; and the other two from the Middlesex University.

At the three institutes from which the participants were recruited, they offered courses on both A→B and B→A translations; however, the percentages of translation training of different translation directions varied between the institutes, based on the questionnaire answers. Table 7 shows the descriptive statistics of the percentages of B→A translation involved in the postgraduate training at the three institutes, whereas Table 8 reveals those of A→B in these schools.

	N	Mini-mum	Maxi-mum	Mean	Std. Deviation
English→Mandarin Work (Imperial) %	10	50	100	88.5	14.9164
English→Mandarin Work (Wesminster) %	3	50	50	50	.00000
English→Mandarin Work (Middlesex) %	2	80	100	90	14.1421

Table 7. Descriptive Statistics of Percentages of B→A (English into Mandarin Chinese) Translation Training in the Three Institutes.

	N	Mini-mum	Maxi-mum	Mean	Std. Deviation
Mandarin→English Work (Imperial) %	10	.00	50.0	11.5	14.9164
Mandarin→English Work (Westminster) %	3	50.0	50.0	50.0	.00000
Mandarin→English Work (Middlesex) %	2	.00	20.0	10.0	14.1421

Table 8. Descriptive Statistics of Percentages of A→B (Mandarin Chinese into English) Translation Training in the Three Institutes.

In the Language and Translation Background Questionnaire, 'working as professional translator' was defined as 'after translator training, having worked as a translator as a paid profession'. Ten out of the fifteen participants reported to have worked

as a professional translator, whereas the other five mentioned that they had not.

Ten out of the fifteen participants reported that they had worked or had been working as a professional (paid) translator. Table 9 indicates that the proportion of their professional work consists of more translation from English into Mandarin Chinese (B→A translation; Mean = 62%, SD = 28.83%) than translating from Mandarin Chinese into English (A→B translation; Mean = 38%, SD = 29.83%). Lastly, fourteen of the fifteen participants reported that they had had other work experience in the past, whereas one of them had not.

	N	Minimum	Maximum	Mean	Std. Deviation
Professional English → Mandarin Translation Work %	10	0.00	100.00	62	29.83
Professional English → Mandarin Translation Work %	10	0.00	100.00	38	29.83
Valid Number of Subjects	10				

Table 9. Descriptive Statistics of Percentages of Professional Translation from Chinese Mandarin into English vs. from English into Mandarin Chinese

3.12.2 Spanish (A) – English (B) Eye-tracking Experiment

Two of the eight participants were male and the other six were female. Two of the participants were left-handed, and the other six were right-handed. Four of the eight subjects are from Spain; two from Argentina; one from Chile; and one from Mexico. Two of them had an age range between eighteen and twenty five; two of them were aged between twenty six and thirty five; and the ages of the other four fell between thirty six and forty five.

As for language use, five out of the eight participants reported that English was the

language of habitual use; only one participant revealed that his/her language of habitual use was Spanish, and two people indicated that both English and Spanish were their languages of habitual use. Seven out of the eight subjects reported that Spanish was more difficult for them to type, whereas only one subject reported otherwise.

One of the eight subjects spoke no other dialect or language. Five of them reported to be able to speak one other dialect or language. One of them indicted to be able to speak two other languages. One of them revealed to be able to speak three other dialects or languages. In addition, five out of the eight subjects reported that it was easier for them to translate from English into Spanish (B→A), whereas the other three indicated that they found translating the other way around was less difficult.

Moreover, four out of the eight subjects had a B.A. degree related to language or translation studies, whereas the other four people's B.A. degrees were unrelated. Four of the eight participants scored 7.0 on IELTS; two scored 7.5; one scored 8.0; and the other one scored 9.0. All of the eight subjects attended the Imperial College London for their postgraduate translation training course. Table 10 shows the percentages of Spanish → English and English → Spanish translation training during the post-graduate program.

	N	Mini-mum	Maxi-mum	Mean	Std. Deviation
English→ Spanish Work %	8	2	100	67.13	36.806
Spanish→ English Work %	8	0	98	26.63	32.641
Valid Number of Subjects	8				

Table 10. Descriptive Statistics of Percentages of A→B (Spanish into English) and B→A (English into Spanish) Translation Training.

In addition, five out of the eight subjects indicated that they had worked as a professional translator, whereas the other three reported that they had not. Table 11 indicates that the proportion of their professional work consists of more translation from English into Spanish (B→A translation; Mean = 68%, SD = 38.34%) than translating from Spanish into English (A→B translation; Mean = 32%, SD = 38.34%). Lastly, all eight participants reported that they had had other work experience in the past.

	N	Minimum	Maximum	Mean	Std. Deviation
Professional English → Mandarin Translation Work %	5	10.00	100.00	68	38.34
Professional English → Mandarin Translation Work %	5	0.00	90.00	32	38.34
Valid Number of Subjects	5				

Table 11. Descriptive Statistics of Percentages of Professional Translation from Spanish into English vs. from English into Spanish

3.13 Fixation Plotting

A program called FixaPlot was used to plot all fixation points across the screen. Figure 6 shows a typical view of FixPlot. For exploratory data analysis (see **Chapters Four** and **Five**), two side-by-side split windows that carried the source text on the left and target text on the right were defined as two separate Areas of Interest (AOI) to see whether or not significantly more fixations appeared in the source text or the target text, if there was a difference at all.

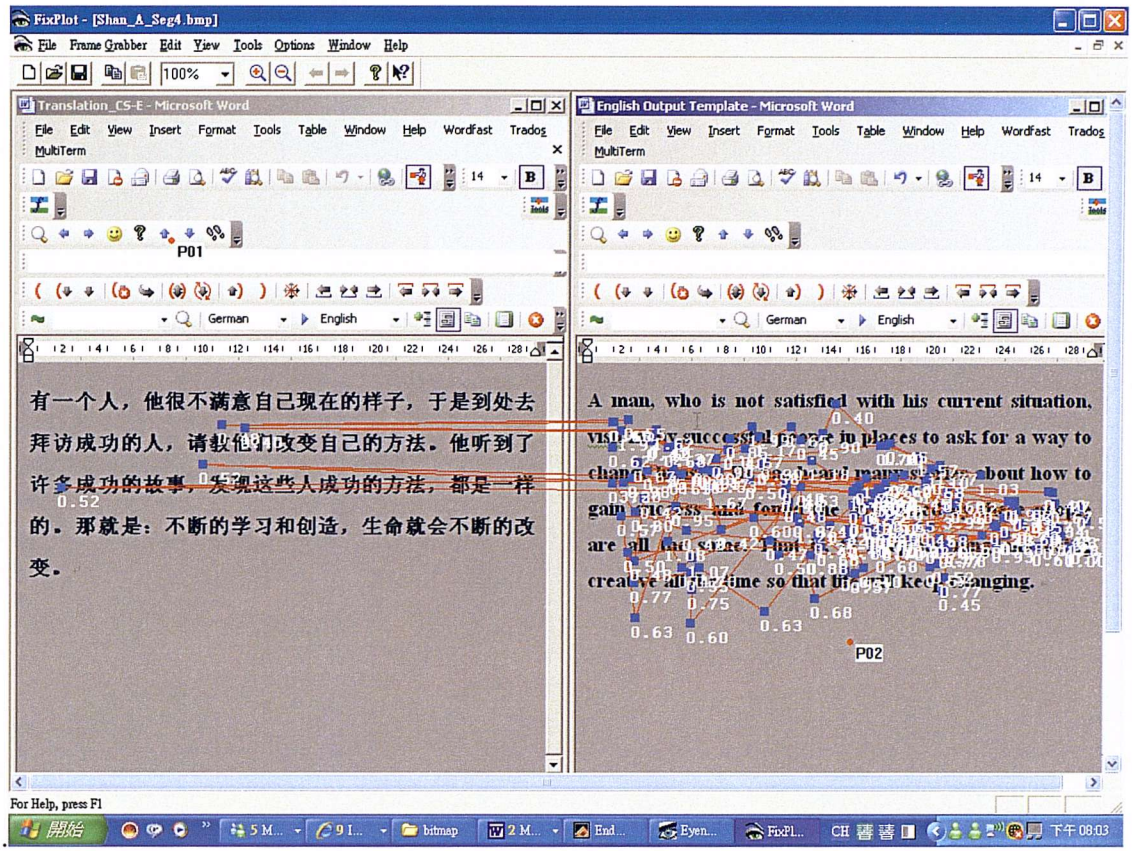


Figure 6. FixPlot Interface

Blue squares in the figure represent a fixation, and the red lines are the paths along which the gazing moved.

3.14 Experimental Procedures

3.14.1 Description Given to the Participants

An information page 'Description of the Procedure of the Study' (Appendix L) was presented to the participants to ensure that they would carry out the tasks as instructed. The investigator read out the entire content of the information page, and afterwards explained and emphasized the importance of following all instructions, especially those that specifically required them to minimize their overt body and head movement and that they needed to finish all tasks to the best of their ability in one go without proofreading or post-

editing. They were also requested to remain silent and, where possible, fully concentrate on all tasks until they finished, upon which moment, they had to say the word DONE immediately to the researcher. This indicated completion of a task on the part of the participant and the investigator would stop data collection and recording on the investigator interface. It has to be acknowledged that, due to technical constraints of the eye-tracking experiments, all of the instructions had to be followed by the participants, and that this might have affected the ecological validity of the study to a certain extent.

3.14.2 Participant Preparation

Upon arrival at the eye-tracking laboratory, the subjects were introduced to the entire laboratory setup. The risk of the study surrounding the camera's infrared lighting was explained to them in order to ensure that they knew that the study was non-invasive. They were advised to use the restroom before embarking on the experiment. A number of the participants were worried about having to use a dictionary during translation, so the investigator reassured them by explaining that the material was designed and manipulated in a way that no dictionary would be required. The investigator also explained that there would be a practice session in which the practice material was equally easy as the material prepared for the rest of the experiment.

Afterwards, the participants were asked to read and signed the consent form. The Language and Translation Background Questionnaire was described to the participants. They were then asked to fill in the questionnaire. If the participants had, at any point, queries or doubts about any listed question that required clarifying, the investigator would give a detailed explanation until they fully understood the answer to the question.

They were told that they were allowed to look at the keyboard to search for the correct key(s) to press for text production. During the practice session, if they wished to look downwardly to search for the correct key(s), they were asked to refrain from overt downward movement of their heads. They were instructed to downwardly move their eyeballs in a way the Pan-Tilt camera would still be able to capture data related to their eye

movement without too much information loss. A live demonstration was performed by the investigator when necessary.

The instruction regarding the task type was given verbally by the investigator when the participant was sitting with minimal movement on the chair and staring at point 5 of the calibration map. The investigator would ask whether or not the participant was ready. Once the participant said YES, the investigator would tell him/her to translate or type or read by saying PLEASE TRANSLATE or PLEASE TYPE or PLEASE READ. Upon hearing the instruction, the participant would firstly minimize the eye calibration map and then start the task as instructed in the split-window environment. At this moment, the investigator would initiate Eye-Trac 6000 for data collection/recording.

To reiterate, the participants were instructed to say DONE, once they thought they had finished any given task. The investigator would then stop data recording. This concludes how the participants were prepared for the experimental process.

3.14.3 Practice Session

Before the experiment, a practice session was conducted to help the participants familiarize themselves with the entire experimental process. For the first experiment, two short English sentences (Appendices M and N) and two short Mandarin sentences (Appendices O and P) were used for translation practice and typing practice. For the second experiment, the same English sentences were still utilized as practice material, and two short Spanish sentences were employed for translation practice, typing practice and reading practice (Appendices Q and R); typing and reading Spanish shared the same practice sentence. If anything went wrong, the researcher would verbally explain the problem and ask the participant to practice again with the same practice material until s/he was fully aware of and familiar with all the details involved in the experiment. Should any questions arise on the part of a participant during the practice session, the researcher would answer them to the best of his ability until s/he was happy to embark on the experiment.

3.15 Conclusion

This chapter has described basic eye biology, principles of eye-tracking technology, eye-tracking laboratory setup, respective experimental designs of the two eye-tracking experiments, data collection and experimental procedures involved. This has established a methodological backdrop against which the study was carried out.

Chapter Four Mandarin (A) – English (B) Eye-tracking Experiment

4.1 Introduction

This chapter reports results and findings of the Mandarin (A) – English (B) eye-tracking experiment, based on the dependent variables introduced in **Chapter Three**. The first part of the reported results is based on *t* tests conducted to determine if there was any significant difference in cognitive loading caused by different translation directions, and the concepts behind the test are introduced in **Section 4.3**.

Additionally, in a factorial experiment design where more than two factors are involved, 'ANOVA' (analysis of variance) is generally carried out to determine (1) if any given *factor* has a *main effect* on the dependent variables; and (2) if an *interaction* exists between the *factors*. Based on the results generated by ANOVAs, *post-hoc* analyses are carried out for further results with an *exploratory* purpose. An introduction of ANOVA, 'main effect', 'interaction', and 'post-hoc analysis' is included in **Section 4.6**, and the results based on these analyses are reported.

In addition, results of two exploratory statistical analyses, Principal Component Analysis (PCA, which is explained in **Section 4.10.2**) and Longitudinal Analysis (see **Section 4.10.3**) are described. As previously described in **Section 2.8**, Balling (2008) pointed out the potential application of PCA to seek correlations between the various dependent variables in an eye-tracking study. Since the study presented here collected a variety of eye-tracking data, a correlation analysis (**Section 4.10.1**) that is a precursor to a PCA was carried out to explore the correlation structure of the dependent variables, and then the PCA was conducted and the results of the analysis are reported in this chapter. Also indicated in **Section 2.8**, Hyönä *et al.* (1995) suggested a 'familiarity effect' during 'interpreting'; therefore, longitudinal analysis (using pupillometric data) was carried out to test if such an effect existed also during 'written translation'. The chapter concludes with a discussion regarding whether or

not eye-tracking has a broader applicability to the field of T&I Studies and regarding whether or not the predictions suggested by the Revised Hierarchical Model were valid at a textual level.

4.2 Behavioral Results

It is crucial to ascertain that the participants' behavior during the experiment produced what was instructed by the investigator. Departure from the investigator's instructions could produce (during typing or translating) inaccurate outputs, which would invalidate the physiological data. Therefore, behavioral results on both typing and translating had to be evaluated. Once the behavioral results had been demonstrated to reflect the instructions given, physiological data could then be validated and entered for statistical analyses. This was an improvement from the study by Pavlović and Jensen (2009). Therefore, two Mandarin language/translation teachers with English as their second language graded the translation and typing outputs, based on a standard scoring scheme (100 = full mark or 100% accuracy; 60 = pass or 60% accuracy). Table 12 shows the scoring results for typing Mandarin Chinese, typing English, translating from Mandarin into English and translating from English into Mandarin Chinese:

	N	Minimum	Maximum	Mean	Std. Deviation
Type Mandarin %	15	95	100	99.27	1.387
Type English %	15	98	100	99.60	.737
Translation Mandarin → English %	15	87	95	91.40	2.772
Translation English → Mandarin %	15	90	99	94.93	2.434
Valid Number of Subjects	15				

Table 12. Descriptive Statistics on Scorings of the Outputs (Mandarin - English)

The behavioral results indicated that the subjects did conduct the tasks as instructed by the investigator; their translation and typing outputs scored at a fairly high accuracy level. This validated the physiological/visual-attention data, which were then deemed usable for further statistical analyses. Translations into English scored lower than into Chinese. One possible interpretation for this is that translating into a second language generates more cognitive loading than into a first language. Additionally, the translations were clearly of lower quality, as is evident from the grades; the question is whether this is a necessary result of it being a more difficult task, which is quite likely. However, it has to be acknowledged that another arguable interpretation is presumably that a translation could be

of lower quality, independently of the cognitive loading.

4.3 Key Concepts behind *t* Test in Statistics

Chapter Three specified the dependent variables measured as data in the study: pupil size, fixation count, task time, fixation frequency, blink frequency and fixation duration/time. A set of data collected is called a *dataset*. For instance, in this study, the time spent by 'each' participant on translating from Mandarin Chinese into English was collected as a single data point. The temporal durations of completing the task by 'all' participants were individually collected, and then jointly formed a set of data points or 'dataset' of task time.

One of the primary aims of the study is to determine whether or not there is any difference in the cognitive loading between A→B and B→A translation. To select an appropriate statistical test to determine the extent to which the respective datasets of the two tasks are different from each other, such concepts as 'normal distribution', and 'standard deviation' (SD) are used.

A 'distribution' refers to how a dataset is distributed and whether or not a distribution is 'normal' has an effect on selection and use of the type of statistical test for analysis. Figure 7 is a graphic representation of a 'normal distribution'.

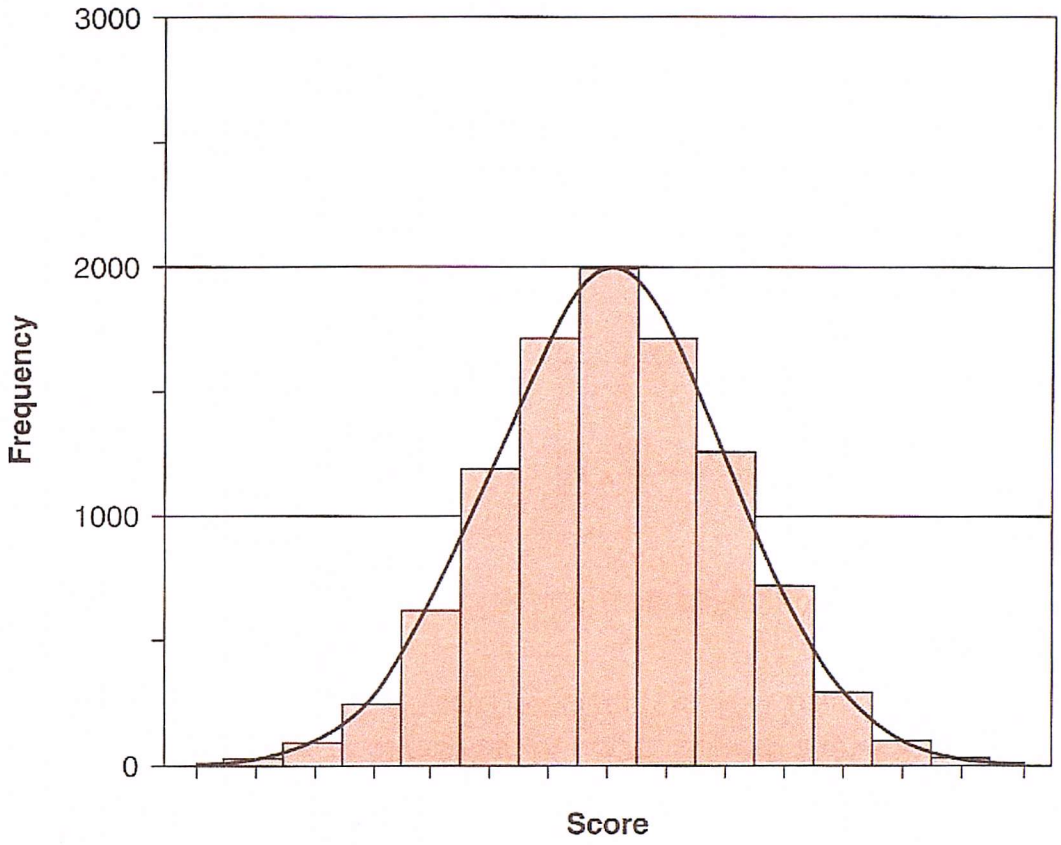


Figure 7. A Bell-shaped Normal Distribution

Source: (Ward, 2005: 9)

Field (2005) explained in a clear fashion what a normal distribution is:

> 'Once you've collected some data, a very useful thing to do is to plot a graph of how many times each score occurs. This is known as a frequency distribution, or histogram, which is simply a graph plotting values of observations on the horizontal axis, with a bar showing how many times each value occurred in the data set. Frequency distributions can be very useful for assessing properties of the distribution of scores. For one thing, by looking at which score that occurs most frequently in the data set. Frequency distributions come in many shapes and sizes...In an ideal world, our data would be distributed symmetrically around the centre of all scores. As such,

> if we drew a vertical line through the centre of the distribution, then it should look the same on both sides. This is known as a normal distribution, and is characterized by the bell-shaped curve...This shape basically implies that the majority of scores lie around the centre of the distribution (so that the largest bars on the histogram are all around the central value). Also, as we get further away from the centre, the bars get smaller, implying that as scores start to deviate from the centre, their frequency is decreasing. As we move further away from the centre, our scores become very infrequent (the bars are very short). An example of a normal distribution is shown in Figure [1]' (ibid.: 8).

The concept of 'standard deviation' helps us to understand the shape of a distribution:

> '[I]f the mean represents the data well, then most of the scores [sample values] will cluster close to the mean, and the resulting standard deviation is small relative to the mean. When the mean is a worse representation of the data, the scores cluster more widely around the mean...and the standard deviation is larger' (Field, 2005: 10).

In statistics, a *t* test is adopted to statistically compare the means of various measured variables. The use of a *t* test assumes that the distributions of the variables are normal. What a *t* test does is to reject the null hypothesis (H_0) and therefore see if the alternative hypothesis (H_A) can be established.

- **H_0: $\mu_A = \mu_B$ (H_0 denotes null hypothesis)**
- **H_A: $\mu_A \neq \mu_B$ (H_A denotes alternative hypothesis)**

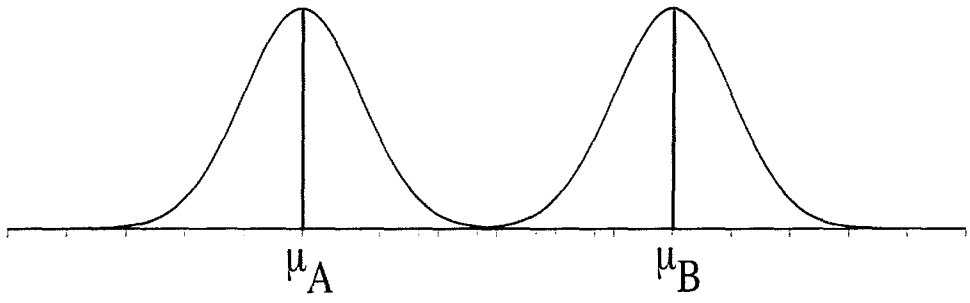


Figure 8. Two Normal Distributions with Different Means

μ denotes the mean value of a distribution, and μ_A denotes the mean of distribution A and μ_B denotes the mean of distribution B. Source: (CSIC, Online)

To illustrate, a null hypothesis is that *there is no significant difference between the means of the two variables in question,* whereas an alternative hypothesis is that *there is a significant difference between the means of the two variables.* Figure 8 shows a typical example of rejected null hypothesis, revealing that the alternative hypothesis has been established, i.e. the means of the two sample distributions are significantly different. However, Figure 9 shows that, even though the two sample distributions (one in green and the other red) are normal, they have the same mean value, and a *t* test would not be able to reject the null hypothesis, which means that there is no significance between the means of the two sample distributions.

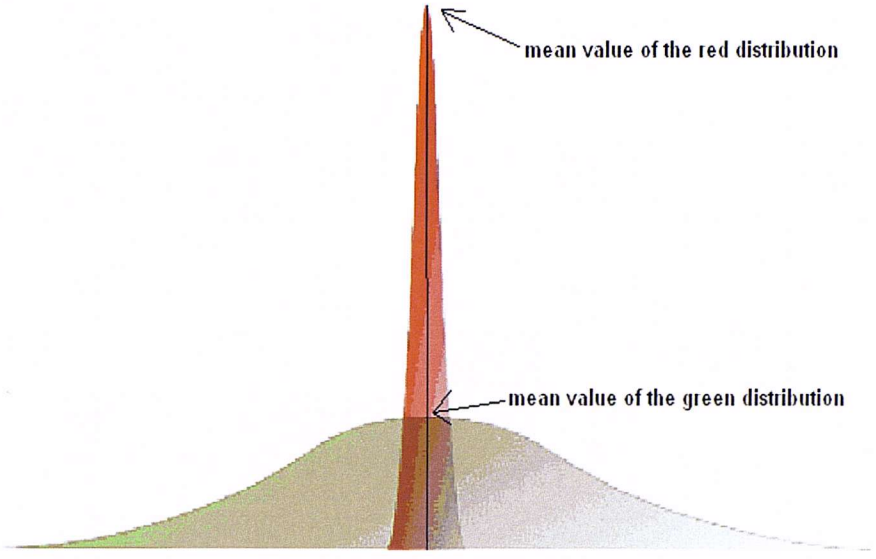


Figure 9. Two Normal Distributions with the Same Mean Value

Source: Modified from Herbie (Online)

Once a *t* test is conducted and the null hypothesis is rejected with a statistical significance (*p* value) of 0.05 or less than 0.05, one can then safely say that there is a statistically significant difference between the means of the two distributions. Statistical significance is often thresholded at 0.05, which indicates that 95 out of 100 times the null hypothesis will be rejected. Anything above the value of 0.05 can be marginally significant (e.g. 0.07) or simply non-significant (e.g. 0.1).

The shapes of the distributions of all the following datasets of the dependent variables measured in the study were explored with SPSS 15, and it was revealed that they were all normally distributed. Therefore, a *t* test was employed to determine whether or not there was a significant difference between A→B and B→A translations.

Further, all the datasets of the measured variables presented below were obtained from a 'repeated measures' experiment; in other words, all the data points were collected from the same person within a particular timeframe, which makes all the data points dependent on each other. Therefore, a 'paired samples *t* test' was selected to compare the

means of various variables that indicated global cognitive loading between translation tasks of different directions. A paired samples *t* test is also called a dependent samples *t* test, or a related scores *t* test, indicating that this specific *t* test is conducted to compare means of two variables whose data points are related to each other.

A paired samples *t* test is carried out in a two-tailed fashion. Statistical significance (*p* value) of a two-tailed *t* test is always reported in a study, unless a researcher has a very clear and definite hypothesis (based on *a priori* knowledge) to test, in which case one-tailed *p* value can be reported. In the study, all of the statistical tests were two-tailed. This is because the Revised Hierarchical Model could not help to generate a definitive hypothesis regarding cognitive loading of translation at a *sentential, textual* level (see **Section 2.4**). Field (2005) gives a very helpful analogy of the concept of one-tailed vs. two-tailed tests as follows:

> 'When using statistical tests, we either have a specific prediction about what will happen, such as "the more someone reads this book, the more they want to kill its author", or we don't really know what will happen, such as "reading more of this book increase or decrease the reader's desire to kill its author". The former example is directional: we've explicitly said that people will want to kill [the author] more, as they read more of this book. If we tested this hypothesis statistically, the test is called a one-tailed test. The second hypothesis is non-directional: we stated that the desire to kill [the author] will change because of reading more of this book, but we haven't said whether that desire will increase or decrease. If we tested this hypothesis statistically, the test would be a two-tailed test' (ibid.: 29).

In this doctoral study, the aim was to quantitatively measure if there is any difference in cognitive loading between A→B and B→A textual translations, without any definitive prediction of the outcome, so the two-tailed test was required. Other statistical terms involved in reporting *t* test analyses are introduced immediately after results of pupil size are presented.

4.4 Difference in Global Cognitive Loading between Mandarin Chinese → English vs. English → Mandarin Chinese Translations

4.4.1 Pupil Size(mm)

A paired samples *t* test reveals that the pupil size of translating from Mandarin Chinese into English (M = 5.09, SE = 0.17) was significantly larger than that of translating from English into Mandarin Chinese (M = 4.87, SE = 0.15, $t(14) = 2.33$, $p = 0.036$, $r = 0.53$). M denotes the 'mean value' of the samples in a study. SE refers to 'standard error of the mean'. 'Standard error' helps us understand how well the samples represent the population in question (for a detailed discussion, see Field, 2005: 14-15). The italicized letter r means 'effect size', and this 'is simply an objective and standardized measure of the magnitude of the observed effect' (Field, 2005: 32). Simply put, it numerically represents the strength of an effect on the dependent variable, if and when there is an effect. Cohen (1992) has suggested the 'r' values should be interpreted as follows: $r = 0.10$ (small effect); $r = 0.30$ (medium effect); and $r = 0.50$ (large effect). Regarding degrees of freedom, they represent the number of estimates or samples that are allowed to vary during statistical calculations (for a more detailed discussion, see Field, 2005; 320). The result therefore suggests that A→B translation was more cognitively demanding than B→A.

4.4.2 Overall Fixation Count across Texts on the Screen

A paired samples *t* test reveals that the fixation count during translating from Mandarin Chinese into English (M = 123.47, SE = 12.33) was significantly larger than that during translating from English into Mandarin Chinese (M = 79.80, SE = 12.65, $t(14) = 5.24$, $p < 0.001$, $r = 0.81$). This suggests that the former task caused more cognitive loading than the latter task. The fixation count result has suggested that, across the novice translators, A→B translation was more cognitively demanding than B→A translation.

4.4.3 Task Time (measured in milliseconds)

A paired samples *t* test reveals that the task time for translating from Mandarin Chinese into English (M = 3489.33.47, SE = 622.91) was significantly longer than that for translating from English into Mandarin Chinese (M = 2290.73, SE = 441, $t(14) = 3.29$, $p = 0.005$, $r = 0.66$). This suggests that, for novice translators, A→B translation was more cognitively demanding than B→A translation.

However, as task time can be measured with or without the aid of an eye-tracker, even though the result was significant, it is not relevant to the question of whether eye-tracking has a broader applicability in T&I Studies.

4.4.4 Fixation Frequency across Texts on the Screen (decimal numbers)

A paired samples *t* test reveals that the fixation frequency during translating from Mandarin Chinese into English (M = 0.62, SE = 0.03) was significantly higher than that during translating from English into Mandarin Chinese (M = 0.54, SE = 0.04, $t(14) = 2.39$, p = 0.031, r = 0.54). This suggests that, for the novice translators, A→B translation was more cognitively demanding than B→A translation.

4.4.5 Blink Frequency (decimal numbers)

A paired samples *t* test showed no significance (p = 0.419). This indicates that, unless participants in any eye-tracking study are not allowed to look down, blink frequency is not a very reliable observation.

The results presented between **Sections 4.4.1** and **4.4.4** have suggested that A→B translation was more cognitively demanding than B→A translation. This indicates that the predictions suggested by the Revised Hierarchical Model were valid at a textual level. In addition, the usefulness of the above eye-tracking measurements has demonstrated a

broader applicability of eye-tracking to the field of T&I Studies.

4.5 Secondary Results: Attention Allocation to Target Text (TT) and Source Text (ST)

- **Fixation Count on ST or TT during English → Mandarin Translation**

A paired samples *t* test reveals that, during translating from English into Mandarin Chinese, the fixation count on the TT (M = 58.67, SE = 9.47) was significantly larger than that on the ST (M = 11.40, SE = 2.18, $t(14)$ = -5.48, $p < 0.001$, r = 0.83).

- **Fixation Time (sec) on ST or TT during English → Mandarin Translation**

A paired samples *t* test reveals that, during translating from English into Mandarin Chinese, the fixation time on the TT (M = 53.99, SE = 9.42) was significantly larger than that on the ST (M = 5.75, SE = 1.1, $t(14)$ = -5.25, $p < 0.001$, r = 0.81).

- **Fixation Frequency on ST or TT during English → Mandarin Translation**

A paired samples *t* test reveals that, during translating from English into Mandarin Chinese, the fixation frequency on the TT (M = 0.7, SE = 0.04) was significantly higher than that on the ST (M = 0.5, SE = 1.1, $t(14)$ = 2.35, $p < 0.001$, r = 0.75).

- **Fixation Count on ST or TT during Mandarin → English Translation**

A paired samples *t* test reveals that, during translating from Mandarin Chinese into English, the fixation count on the TT (M = 102.07, SE = 11.41) was significantly higher than that on the ST (M = 15.33, SE = 4.18, $t(14)$ = -7.04, $p < 0.001$, r = 0.88).

- **Fixation Time (sec) on ST or TT during Mandarin → English Translation**

A paired samples *t* test reveals that, during translating from Mandarin Chinese into English, the fixation time on the TT (M = 90.65, SE = 12.04) was significantly longer than that on the ST (M = 8.39, SE = 2.4, $t(14)$ = -6.66, $p < 0.001$, r = 0.87).

- **Fixation Frequency on ST or TT during Mandarin → English Translation**

A paired samples *t* test reveals that, during translating from Mandarin Chinese into

English, the fixation frequency on the TT ($M = 0.6$, $SE = 0.09$) was significantly higher than that on the ST ($M = 0.1$, $SE = 0.01$, $t(14) = 2.25$, $p < 0.001$, $r = 0.76$).

The results presented in **Section 4.5** indicate that, irrespective of the directionality of the translation task, the subjects' total number of fixations on TT, fixation duration within TT and the probability for the subjects to fixate their gaze on TT were higher than that within or on ST. This finding agrees with the results of the studies conducted by Sharmin *et al.* (2008) and Pavlović and Jensen (2009), which also demonstrated that translators paid more visual attention to a target text, based on their fixation data. Even so, whether or not more attention is allocated to a target text during translation with its various modalities and different language combinations still warrants further research.

A possible explanation is that during translation, translators tend to focus more on the output in order to ensure accuracy. During translation, the 'cognitive components' required for ST (static text display) involve ST reading and comprehension, whereas the cognitive components required for TT (dynamic text production) involve keyboard search, typing, reading, TT comprehension, mid-task editing, mid-task proofreading, etc. This could be why more attention was paid to the TT, and this was revealed by Sharmin *et al.* (2008) and Pavlović and Jensen (2009). Once again, this possible explanation merits further research (see **Chapter Seven**). Similar to the results reported in **Section 4.4**, those presented in **Section 4.5** have also indicated a broader viability of eye-tracking in the field of T&I Studies. In particular, the results in **Section 4.5** can form a basis for formulating a more concrete hypothesis to be tested by future research.

4.6 Repeated Measure 2-way ANOVA on Pupillometry

ANOVA is short for 'analysis of variance', and 'way' in a 2-way ANOVA simply means 'factor' in an experiment. Therefore, a 2-way ANOVA is analysis of variance for an experiment with two factors. As the study was designed as a repeated measures experiment, a repeated measures 2-way ANOVA was carried out on the data of pupil size, fixation count, fixation frequency, and task time. It has to be taken into account that, even though 'task

time' does not require an eye-tracker for collection, it was still measured with the aim of furnishing further evidence and strengthening potential results. Again, in terms of ANOVA, Field (2005: 309) explains,

> 'ANOVA has the advantage that it can be used to analyse situations in which there are several independent variables. In these situations, ANOVA tells us how these independent variables interact with each other and what effects these interactions have on the dependent variable'.

An interaction effect is 'the combination effect of two or more *predictor variables* on an outcome *variable*' (Field, 2005; 734). In the Mandarin - English eye-tracking experiment, an 'interaction effect' refers to a scenario where the two independent variables (Task Type and Input Language) interact to generate an effect on the dependent variable(s). ANOVA analysis can help to reveal if there is an interaction between the independent variables; however, in order to interpret the interaction effect, a researcher needs to examine the interaction graph(s).

A repeated measures 2-way ANOVA was carried out to determine (1) if the two factors (Input Language vs. Task Type) had an effect on pupil size; and (2) if there was an interaction between the two factors.

Results revealed that only Task Type had a significant effect on pupil size, $F = 6.34$, $df = 1, 14$, $p = 0.024$, $r = 0.72$, whereas Input Language did not show significant effect, $p = 0.83$. There was a significant interaction between Task Type and Input Language, $F = 14.94$, $df = 1, 14$, $p = 0.02$, $r = 0.72$ (Figure 10).

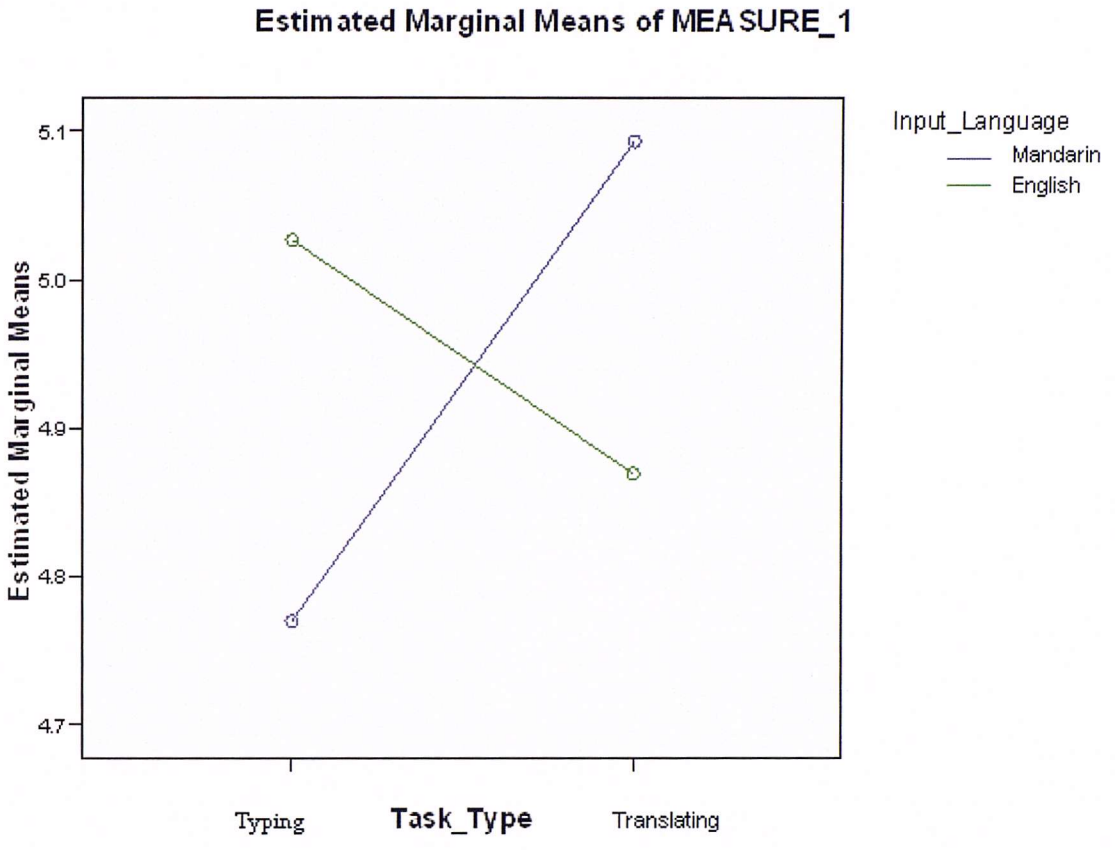


Figure 10. Interaction of Input Language and Task Type (Pupillometry, mm)

It has been shown that Task Type had a main effect on pupil size, and this means that, when the task type changes from typing to translating, pupil size generally increased significantly. However, the factor of Input Language did not show a significant main effect on pupil size. This suggests that whether or not an input language was an ideogram-based (Mandarin Chinese) or alphabet-based language (English) did not seem to have a significant effect on the dependent variable of pupil size.

On the other hand, there was a significant interaction effect between the two factors, i.e. Task Type and Input Language. As shown in Figure 10, this indicates that task type had a different effect on pupil size depending on which input language was associated with the

task involved. In other words, when it comes to typing or translating, in combination with the input language being Mandarin Chinese, the pupil size tended to increase, whereas when the input Language was English, the pupil size tended to be relatively smaller.

4.6.1 Post-hoc Bonferroni-corrected Paired Samples t Test

A post-hoc analysis is an analysis conducted *after* a number of analyses has been carried out and has yielded results. When the existing results present a *potential to generate further, secondary results,* the post-hoc analysis is then employed with an *exploratory* purpose. Take the pupillometric data in the Mandarin – English experiment for instance: (1) It has been revealed that translating from Mandarin Chinese into English was more cognitively demanding than translating from English into Mandarin Chinese (see **Section 4.4.1**); and (2) the ANOVA has shown that Task Type, not Input Language, had a significant effect on pupil size, and that there was an interaction between the two factors (see **Section 4.6**). However, in the same experiment, cognitive loading of 'typing' was measured without any *a priori* assumption; therefore, it would be interesting to *explore* the potential of pupillometric data to yield further, ancillary results by comparing cognitive loading of typing English and typing Mandarin. This is also useful in demonstrating the broader applicability of eye-tracking by exploring further potential results that may form the basis of more concrete hypotheses for future research.

According to Craig (2007), carrying out one post-hoc *t* test is fine; however,

> [D]oing more than one post-hoc *t* test requires a correction to the maximum *p* value allowed called a *Bonferroni* correction[15]. This involves dividing the normal maximum significance level ($p = 0.05$) by the number of comparisons to be made, e.g. if the number of comparisons is two, the maximum *p* value for a significant effect is $0.05 / 2 = 0.025$, [and] if three, then $0.05 / 3 = 0.016$, and so on (ibid.: 5).

[15] Bonferroni correction is '[a] stringent correction for multiple comparisons in which the alpha value is decreased proportionally to the number of independent statistical tests' (Huettel *et al.*: 346).

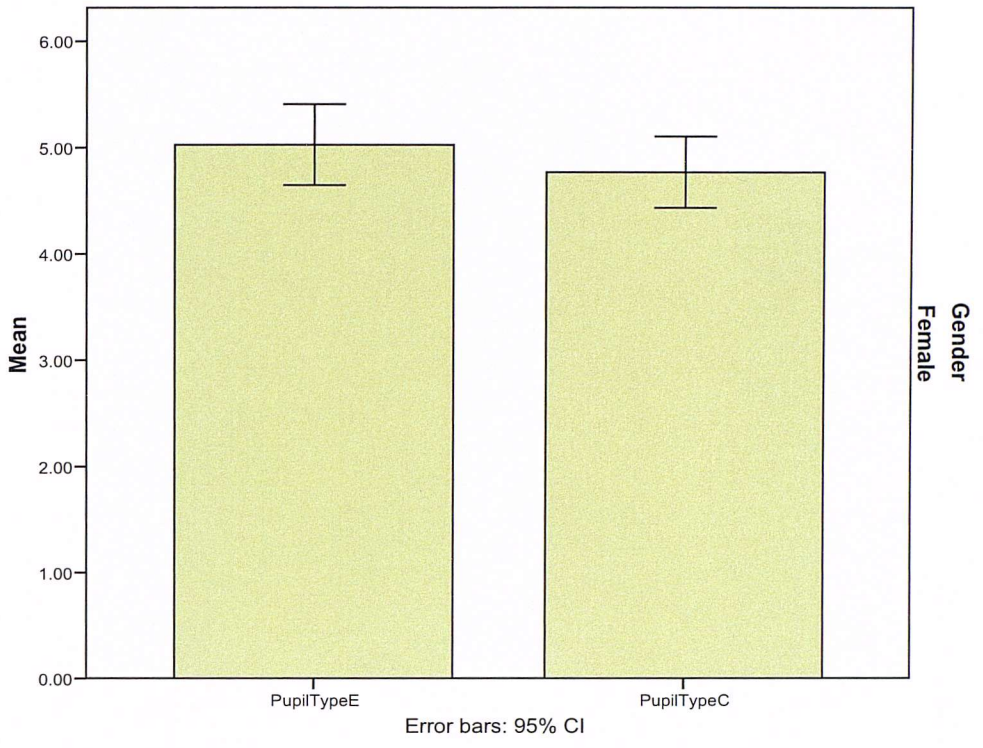


Figure 11. Pupil Sizes (mm) during Typing English and Typing Mandarin Chinese

An eye-tracking study (Wang *et al.*, 2001) compartmentalized the cognitive process of typing Mandarin Chinese into four major steps: 'Pinyin generation', 'pinyin typing', 'multiple choice reaction', and 'selector typing' (ibid.: 355), and suggested that typing Mandarin Chinese was cognitively more demanding than typing English, using an English keyboard. A post-hoc Bonferroni-corrected paired samples t test was conducted (p of 0.05 divided by 1 = 0.05) to determine if there was a significant pupillometric difference between the two typing tasks, and the result showed that pupillometric data during typing English was significantly larger than during typing Mandarin, $M = 0.27$, $SE = 0.1$, $t(14) = 2.58$, $p < 0.022$, $r = 0.57$ (Figure 11). This suggests that typing English was more cognitively demanding than typing Mandarin. However, the result did not support the study of Wang *et al.* (2001).

4.6.2 Fixation Data for Typing Mandarin Chinese and Typing English

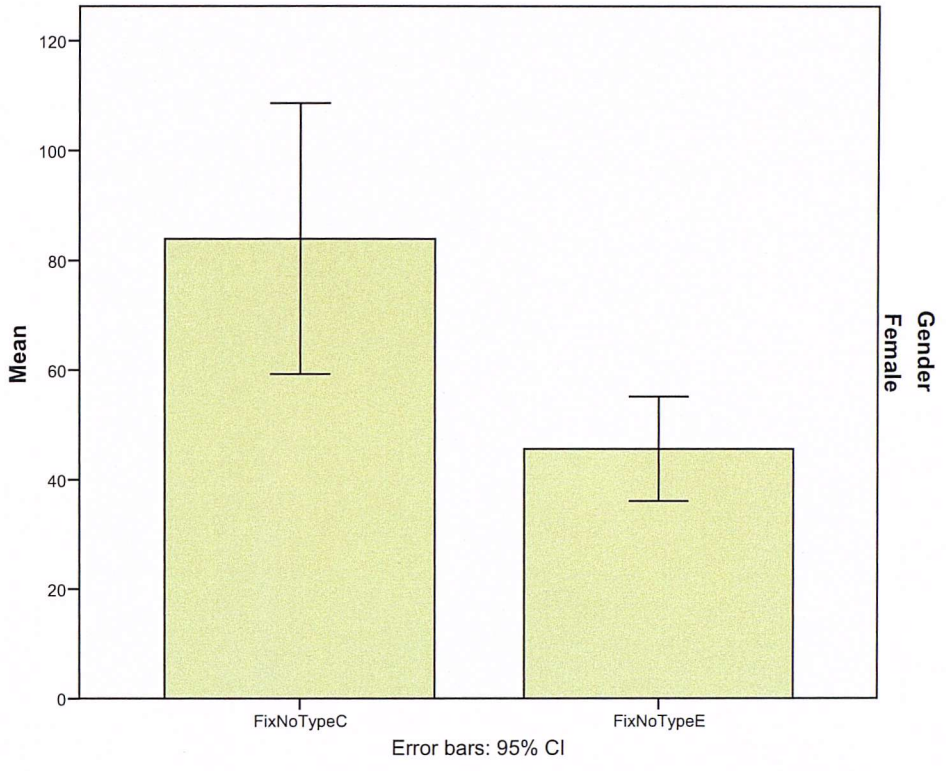


Figure 12. Fixation Counts (integers) during Typing English and Typing Mandarin Chinese

The mean fixation count for 'typing' Mandarin Chinese across the fifteen subjects was 83.93, and the mean fixation count for 'typing' English was 45.67 (Figure 12). A post-hoc paired samples *t* test was carried out and revealed that there was a significant difference between the fixation count during Mandarin typing and during English typing , $M = 38.27$, $SE = 9.27$, $t(14) = 4.13$, $p = 0.001$, $r = 0.74$.

Based on fixation counts, the statistical result has indicated an interesting point that requires interpreting: Typing Mandarin was cognitively *more* demanding than typing English. This result is *inconsistent* with the result yielded by the pupillometric data. However, the fixation results were consistent with the self-reported results from the

questionnaire which revealed that twelve out of the fifteen Chinese participants considered it more difficult to type Mandarin and the reasons of the twelve participants are listed in Table 13. In the questionnaire, the subjects self-reportedly wrote down the reason why they considered typing Mandarin more difficult than typing English, and they indicated that typing Mandarin would require extra steps. These steps were described by Wang *et al.* (2001). This suggests that typing Mandarin required more eye fixations to appear on a computer screen for selection of a correct character to produce, a phenomenon inherent in typing Mandarin replete with homophones. The fixation-based result lends further support to Wang *et al.* (2001).

Reasons Why Typing Mandarin Chinese Was Considered More Difficult Than Typing English	
Participant No.	**Reason(s)**
1	*I am not used to its input system.*
2	*I need to think of the radicals that make up a word before typing.*
3	*I need to select the right character, as I use the Pinyin input system.*
4	*I need to choose the right character from a list of homophones.*
5	*Because, in Chinese, words with the same pronunciation have the same alphabetical spellings, selecting the right word is important.*
6	*First, there are many homophones in Chinese, so selecting the correct word is crucial. Second, sometimes, I am uncertain as to how a Chinese word can be decomposed into English alphabet, using the Pinyin input system.*

7	*Many homophones appear at the same time, so I need to choose the right word.*
8	*I tend to misspell a Chinese character on an English keyboard, using the Pinyin system.*
9	*Word selection is needed, due to simultaneous generation of a variety of homophones.*
10	*With the same pronunciation for many words, I need to select the right words from a list of words.*
11	*Chinese characters have many strokes and radicals and are therefore complicated, as opposed to English typing that is more straightforward.*
12	*I need to think of the English spellings of a Chinese character before typing.*
Reasons Why Typing English Was Considered More Difficult Than Typing Chinese	
14	*The frequency of typing English is lower in comparison.*
15	*Not sure of the spellings sometimes.*
15	*Not always sure where the letters are position on the keyboard.*

Table 13. Reasons Why Typing Mandarin Chinese or Typing English Was Considered More Difficult

In addition, in the study by Pavlović and Jensen (2009) investigating cognitive loading of different translation directions with the language combination of Danish –

English (both alphabet-based), both their pupillometric and fixation results were consistent and suggested that A→B translation was more cognitively demanding than B→A translation. In the Mandarin – English eye-tracking study presented here, the language combination involved Mandarin, an ideogram-based language; the production of such a language using an English keyboard may have involved more instances of gazing on the screen and hence more fixations, due to the different cognitive components pointed out by Wang *et al.* (2001). A potential explanation for the inconsistency between pupillometric and fixation results is that both types of data are more likely to yield consistent results regarding cognitive loading of 'typing' if the language combination involves only alphabet-based languages. However, if the language combination involves an ideogram-based language, inconsistency between pupillometric and fixation results are likely to appear. In light of this suggestion, the issue of whether or not pupillometric and fixation data are more probable to generate consistent results accompanying an experiment with the language combination that is only alphabet-based will be addressed in the second Spanish – English eye-tracking experiment described in **Chapter Five**. Also, it is worth mentioning that it will be interesting to see if consistency between the two types of data and results exists when the two languages in the language pair are both ideogram-based, e.g. Mandarin – Japanese.

To summarize, Mandarin → English (A→B) and English → Mandarin (B→A) translations were considered *holistic* tasks, and **Sections 4.4.1** to **4.4.4** have shown that the former created more cognitive loading on the subjects involved. In terms of secondary findings, the result of the fixation count post-hoc analysis regarding the difference in the cognitive loadings of typing Mandarin vs. typing English has supported the self-reported questionnaire results, i.e. typing Mandarin was considered more difficult than typing English. However, this point was not agreed by the result of the pupil size post-hoc analysis. As previously suggested, this may well be related to the ideogram-based nature of the Mandarin language.

4.7 Repeated Measure 2-way ANOVA on Fixation Count

Repeated measures 2-way ANOVA was carried out to determine (1) if the two factors (Input Language vs. Task Type) had an effect on fixation count; and (2) if there was an interaction between the two factors.

Results revealed that Task Type had a significant effect on fixation count, $F = 13.075$, $df = 1, 14$, $p = 0.03$, $r = 0.69$, and Input Language also had a significant effect on fixation count, $F = 58.285$, $df = 1, 14$, $p < 0.001$, $r = 0.90$. However, there was no significant interaction between Task Type and Input Language, $p = 0.705$. Figure 13 shows there was no interaction between the two factors.

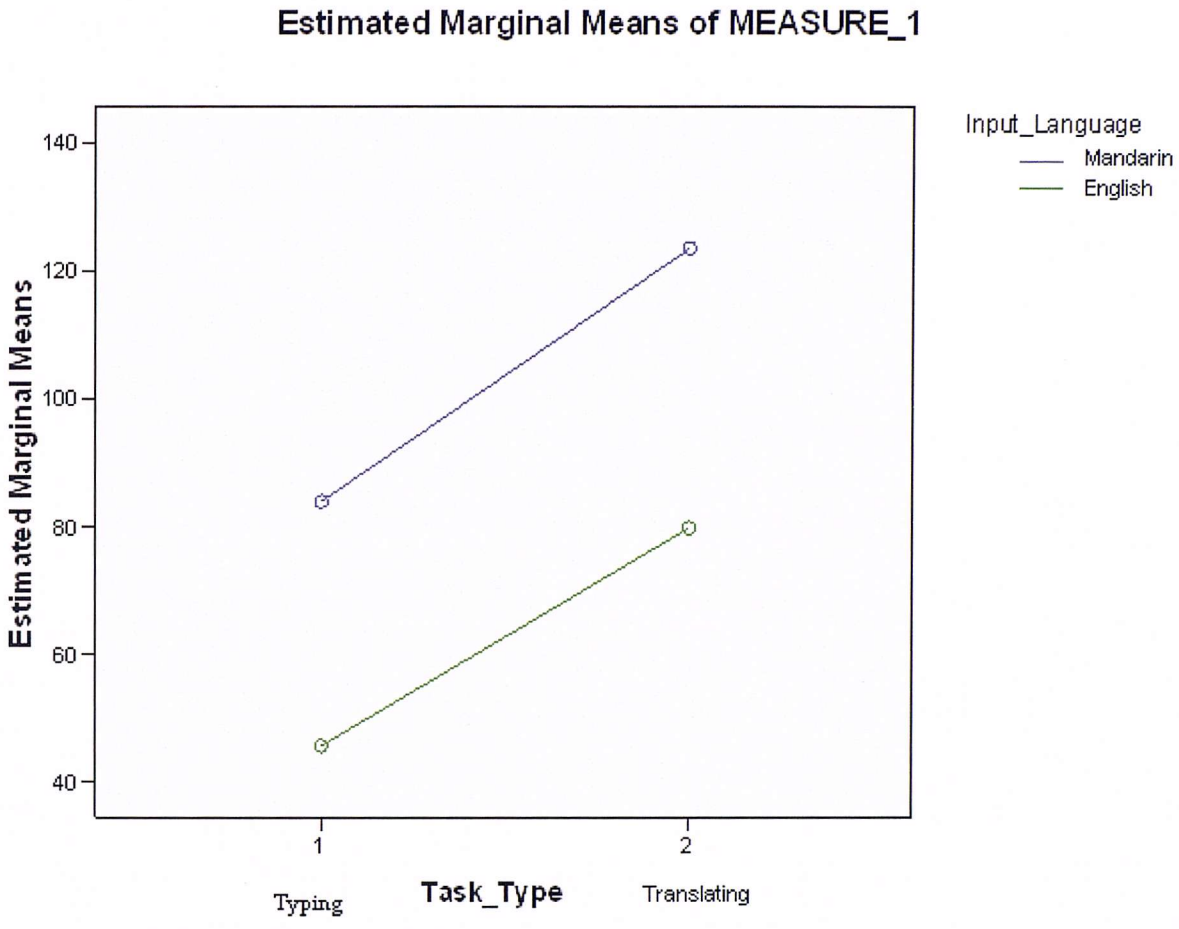


Figure 13. Interaction between Input Language and Task Type Based on Fixation Count

The main effect of Input Language on fixation count means that the number of fixations that appeared was affected by the type of Input Language. Therefore, based on the chart, when the Input Language was Mandarin Chinese, regardless of whether the subjects were instructed to carry out typing or translating into English, there were always significantly more fixations. The main effect of Task Type on fixation count means that the number of fixations that appeared was affected by the type of task conducted (typing vs. translating). Again, based on the chart, it is noticeable that when comparing the fixation counts of the two task types, irrespective of input language or translation directionality, there were invariably more fixations during translating than during typing.

4.8 Repeated Measure 2-way ANOVA on Fixation Frequency

Results revealed that only Input Language had a significant effect on fixation frequency, $F = 0.53$, $df = 1, 14$, $p < 0.001$, $r = 0.75$., whereas Task Type did not show a significant effect, $p = 0.119$. There was a significant interaction between Task Type and Input Language, $F = 0.42$, $df = 1, 14$, $p = 0.002$, $r = 0.70$. The significant interaction means the type of input language influenced the fixation frequency during both typing and translating tasks.

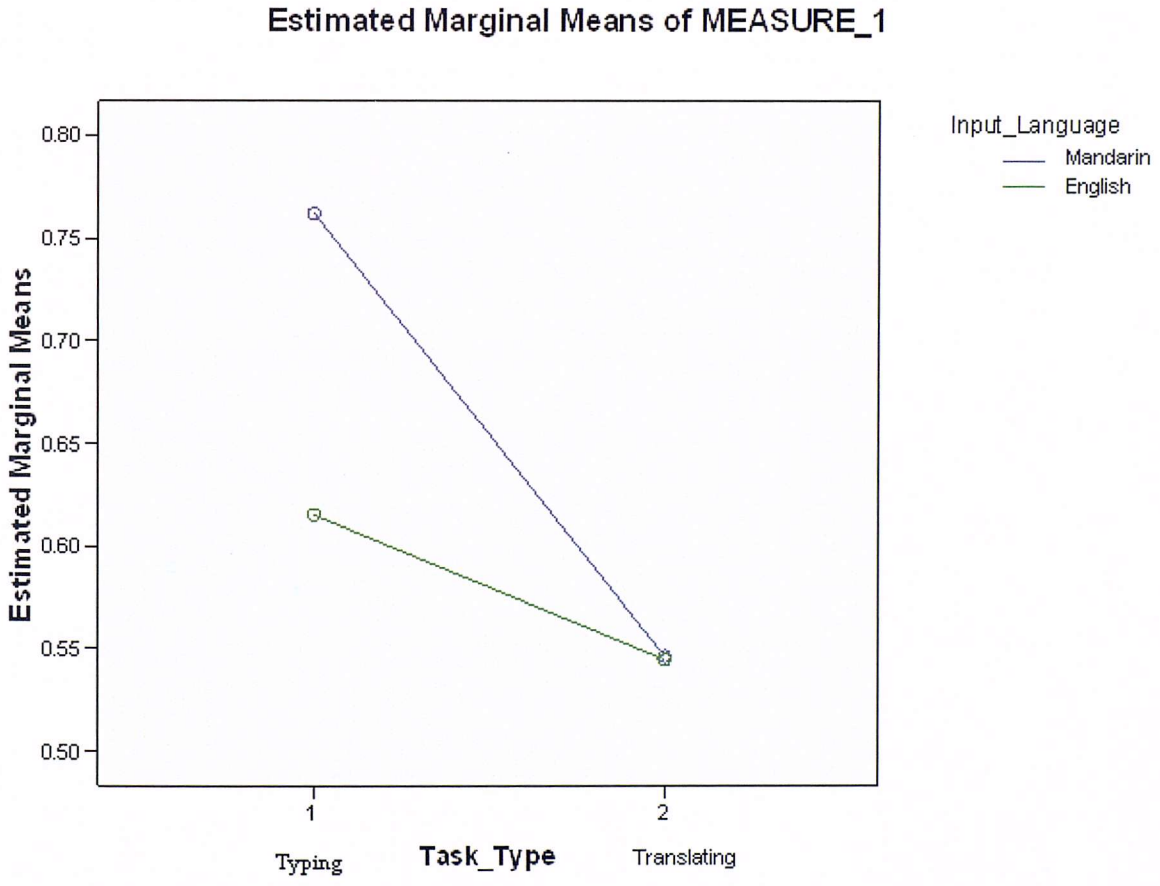


Figure 14. Interaction between Input Language and Task Type Based on Fixation Frequency (decimal numbers)

In Figure 14, it is suggested that there might be a significant difference between the fixation frequencies during typing Mandarin Chinese and typing English. Therefore, a post-hoc Bonferroni-corrected paired samples *t* test was administered. Results show that, on average, the fixation frequency was higher during Mandarin Chinese typing ($M = 0.76$, $SE = 0.04$) than English typing ($M = 0.55$, $SE = 0.06$, $t(14) = 5.165$, $p < 0.001$, $r = 81$). This is consistent with the result of the fixation post-hoc analysis described in **Section 4.6.2**.

4.9 Repeated Measure 2-way ANOVA on Task Time (sec)

Results revealed that the factor of Task Type had a significant effect on task time, $F = 8.843$, $df = 1, 14$, $p = 0.01$, $r = 0.62$, while Input Language also had a significant effect on task time, $F = 18.555$, $df = 1, 14$, $p = 0.001$, $r = 0.75$. However, there was only a very marginally significant interaction between Task Type and Input Language, $F = 3.375$, $df = 1, 14$, $p = 0.088$, $r = 0.44$.

The effect of Task Type indicates that, if we ignore the type of input language involved, the time spent was different according to the type of task involved. In this case, translating took longer to complete than typing.

In the same vein, the effect of Input Language indicates that, if we ignore the type of task involved, the time spent was different according to the type of input language involved. In this case, when the input language was Mandarin Chinese, it took a longer time to complete the task.

In terms of the marginally significant interaction effect of Task Type and Input Language, Figure 15 shows that the type of task had a marginally significantly different effect on task time, depending on the input language involved. In this case, translating from Mandarin Chinese into English took the longest time, followed by translating from English into Mandarin Chinese, and typing Chinese and typing English.

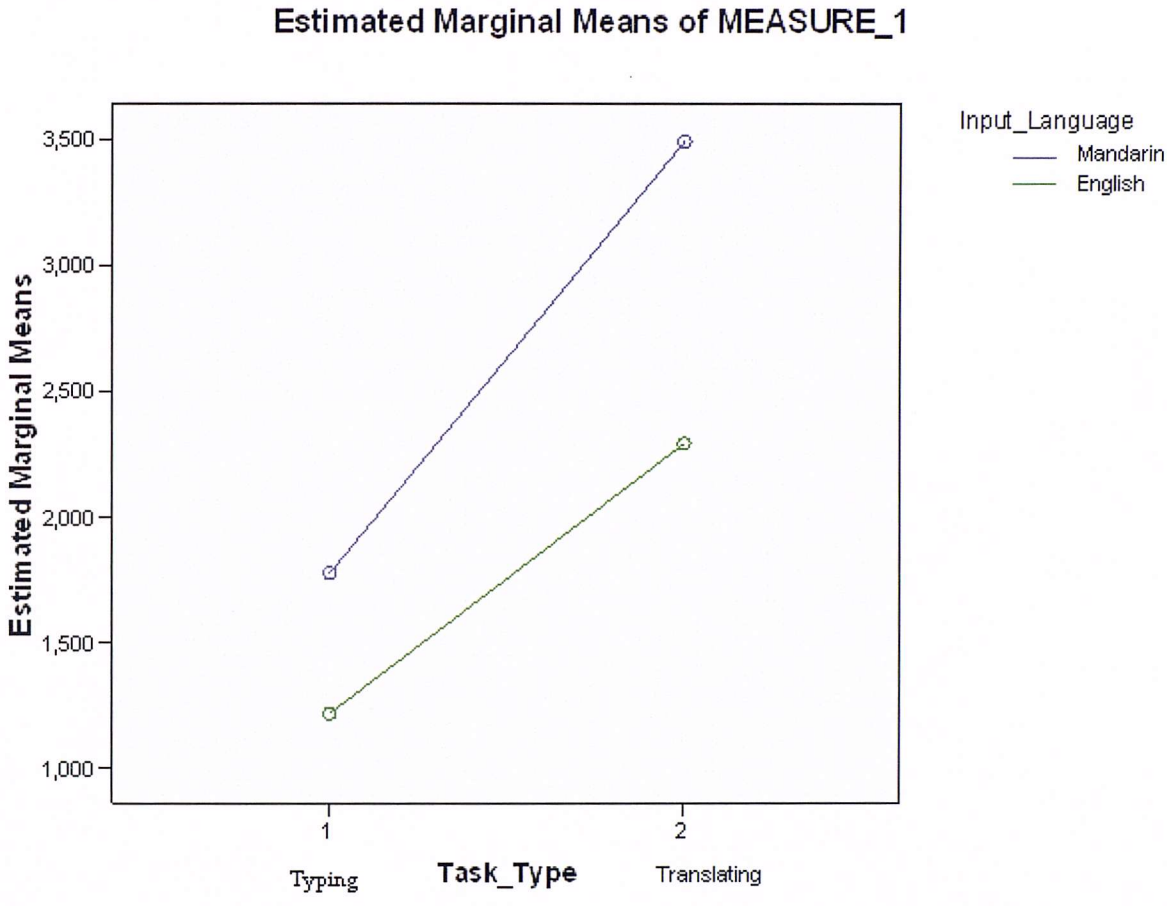


Figure 15. Interaction between Input Language and Task Type Based on Task Time (sec)

It has to be pointed out that no repeated measures 2-way ANOVA was carried out on blink frequency because some of the zero values in the data that were meant to represent blink values could have been 'look-down' data. This is why no ANOVA was conducted on blink frequency.

To summarize, additional, interesting results regarding a number of main effects of the independent variables on the dependent variables and the interactions have been shown by the ANOVAs reported above. Further, apart from the pupillometric result, the results of fixation count, fixation frequency, and task time have all suggested that typing Mandarin

Chinese was cognitively more demanding than typing English, which supported the study by Wang *et al.* (2001) and the questionnaire data. These secondary results may help devise more concrete hypotheses to be tested by future research, and this supports the idea that eye-tracking has a broader applicability to T&I Studies.

4.10 Further Exploratory Statistical Analyses

Further statistical analyses were carried out to further clarify whether eye-tracking technology has a broader viability to T&I Studies

4.10.1 Correlation Analysis

A correlation analysis was carried out to investigate the **correlation structure** of the following dependent variables: Pupil Size, Task Time, Blink Frequency, Fixation Count and Fix Rate[16] for the two tasks, i.e. translating from Mandarin Chinese into English and translating from English into Mandarin Chinese. A correlation analysis is a descriptive tool that gives some *preliminary* information on how variables are correlated. The Spearman correlation coefficient is computed on only 'one pair of variables at a time' (Bartholomew *et al.*, 2002). A correlation usually refers to the strength of the relationship between two variables, and the strength of a correlation is usually numerically represented by a 'correlation coefficient'. The value of a correlation coefficient ranges between 1 and -1, with 1 referring to total positive correlation and -1 represents a total inverse correlation. The + or – value of the correlation coefficient only indicates the direction of the relationship of the two variables. Therefore, for example, when a variable x increases in value, another variable y increases accordingly, and this represents a positive correlation. Conversely, a negative correlation refers to a scenario where when x increases in value, y decreases accordingly. To summarize, the closer a correlation coefficient is to 1 or -1, the stronger the correlation, e.g.

[16]Due to practical limitations of SPSS, in the analysis where figures and tables are included, 'Pupil' refers to 'Pupil Size'; 'Time' means 'Task Time'; 'Blink' represents 'Blink Frequency'; 'Fixation' refers to 'Fixation Count'; and 'fixation rate' and 'fixation frequency' are used interchangeably.

0.8, -0.8 etc. On the contrary, the closer a correlation coefficient is to 0, the weaker the correlation.

The correlation analysis is a precursor to the Principal Component Analysis (PCA) that ensues in **Section 4.10.2**. The difference between the two analyses is that the correlation analysis computes only 'one pair of variables at a time', and the PCA is a 'multivariate' technique that is able to analyze all variables 'simultaneously'. Therefore, statistically, the PCA strengthens the results generated by the correlation analysis. In addition, it has been shown that the result of blink frequency did not show significance. This suggests that this measurement was not reliable in the experiment, a conclusion reached *after* the analysis. However, conducting a correlation analysis and a PCA must include ALL dependent variables a study 'sets out' to measure *before* an experiment, without excluding any of the dependent variables after analyses (Bartholomew *et al.*, 2002). Therefore, blink frequency was included in the analyses.

For each task, a scatter plot for pair wise combinations of the five variables and a table containing the pair wise correlation coefficients are presented

➢ *Mandarin → English Translation*[17]

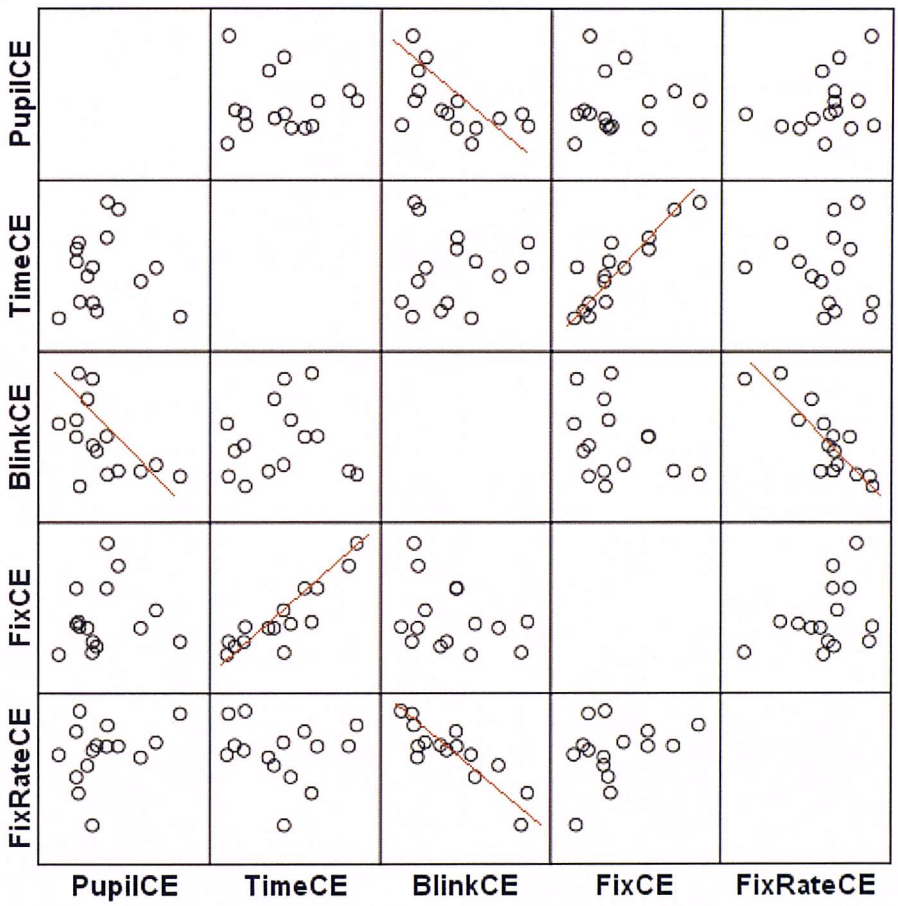


Figure 16. Scatter Plot for Pair Wise Correlations of the five Variables during Mandarin → English Translation

[17] In the figures and tables, CE means Mandarin → English translation, whereas EC represents English → Mandarin translation.

Correlations

			PupilCE	TimeCE	BlinkCE	FixCE	FixRateCE
Spearman's rho	PupilCE	Correlation Coefficient	1.000	.046	-.600*	.190	.354
		Sig. (2-tailed)	.	.869	.018	.498	.196
		N	15	15	15	15	15
	TimeCE	Correlation Coefficient	.046	1.000	.164	.854**	-.096
		Sig. (2-tailed)	.869	.	.558	.000	.732
		N	15	15	15	15	15
	BlinkCE	Correlation Coefficient	-.600*	.164	1.000	-.215	-.839**
		Sig. (2-tailed)	.018	.558	.	.442	.000
		N	15	15	15	15	15
	FixCE	Correlation Coefficient	.190	.854**	-.215	1.000	.322
		Sig. (2-tailed)	.498	.000	.442	.	.241
		N	15	15	15	15	15
	FixRateCE	Correlation Coefficient	.354	-.096	-.839**	.322	1.000
		Sig. (2-tailed)	.196	.732	.000	.241	.
		N	15	15	15	15	15

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 14. Pair Wise Correlation Coefficients of the Five Variables during Mandarin → English Translation

From Figure 16 and Table 14, it can be seen that the highest correlations are **FixCE vs. TimeCE** and **FixRateCE vs. BlinkCE** followed by **BlinkCE vs. PupilCE**. A positive sign means that when both variables follow the same direction while a negative sign means that when one variable goes up the other goes down. The result shows preliminary 'pair-wise correlations of two variables' during Mandarin → English translation. To illustrate, in terms of FixCE vs. TimeCE, a higher number of fixation count indicated a larger amount of task time. As for FixRateCE vs. BlinkCE, when fixation frequency became low, so did blink frequency. In the case of BlinkCE vs. PupilCE, lower blink frequency indicated a smaller pupil size. This result was further strengthened by the PCA results reported in **Section 4.10.2**.

➢ *English → Mandarin Translation*

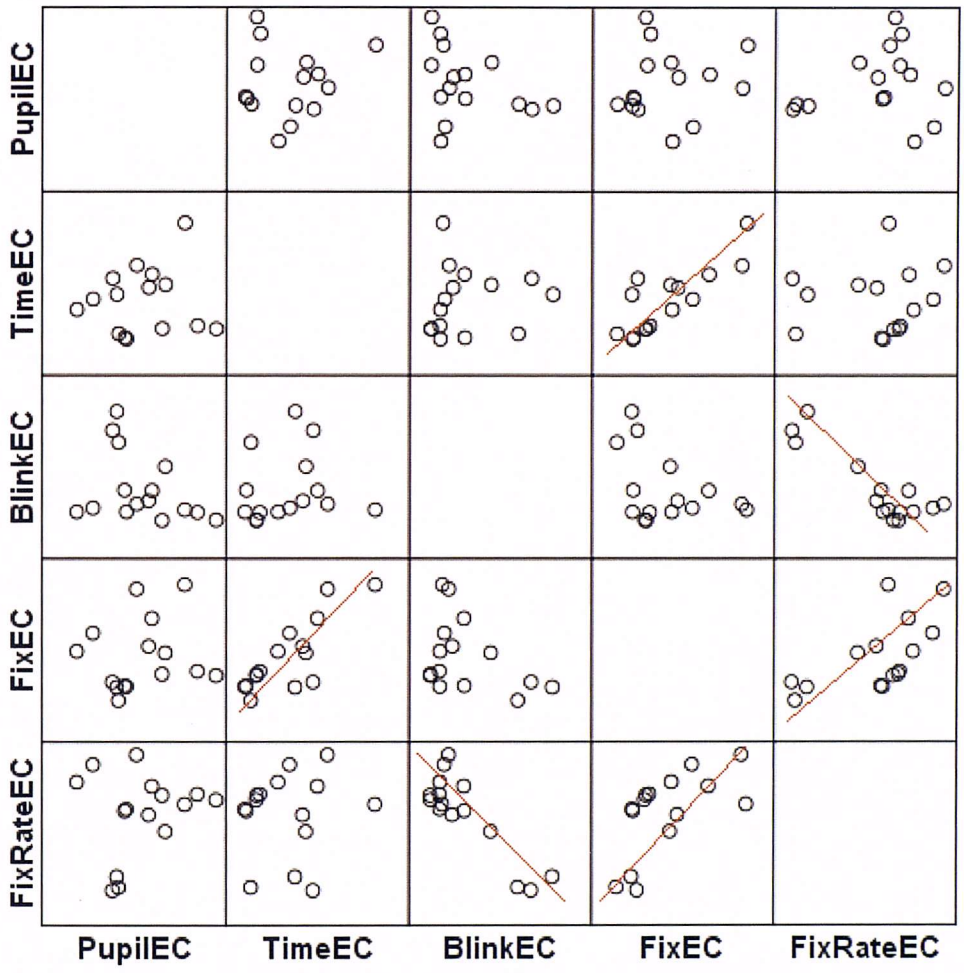


Figure 17. Scatter Plot for Pair Wise Correlations of the five Variables during English → Mandarin Translation

Correlations

			PupilEC	TimeEC	BlinkEC	FixEC	FixRateEC
Spearman's rho	PupilEC	Correlation Coefficient	1.000	.096	-.454	.250	.096
		Sig. (2-tailed)	.	.732	.089	.369	.732
		N	15	15	15	15	15
	TimeEC	Correlation Coefficient	.096	1.000	.290	.729**	.139
		Sig. (2-tailed)	.732	.	.295	.002	.621
		N	15	15	15	15	15
	BlinkEC	Correlation Coefficient	-.454	.290	1.000	-.281	-.640*
		Sig. (2-tailed)	.089	.295	.	.311	.010
		N	15	15	15	15	15
	FixEC	Correlation Coefficient	.250	.729**	-.281	1.000	.661**
		Sig. (2-tailed)	.369	.002	.311	.	.007
		N	15	15	15	15	15
	FixRateEC	Correlation Coefficient	.096	.139	-.640*	.661**	1.000
		Sig. (2-tailed)	.732	.621	.010	.007	.
		N	15	15	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 15. Pair Wise Correlation Coefficients of the Five Variables during English → Mandarin Translation

From Figure 17 and Table 15 above, it can be seen that the highest correlations are **FixEC vs. TimeEC** and are **FixEC vs. FixRateEC** followed by **FixRateEC vs. BlinkEC**. The result shows preliminary 'pair-wise correlations of two variables' during English → Mandarin translation. To elucidate, in terms of FixEC vs. TimeEC, a higher number of fixation count indicated a larger amount of task time. As for FixRateEC vs. BlinkEC, when fixation frequency became low, so did blink frequency. In the case of FixEC vs. FixRateEC, a larger number of fixation count indicated a higher fixation frequency. Again, this result was further strengthened by the PCA results reported in **Section 4.10.2**.

To summarize, these plots and tables are a descriptive tool that gives some preliminary information on how variables are correlated. The Spearman correlation coefficient is computed on only 'one pair of variables at a time', so it was better to use a 'multivariate' technique that is able to analyze the five variables 'simultaneously'. In this case, it was appropriate to use the **PCA**.

4.10.2 Principal Components Analysis (PCA)

As previously described in **Section 2.8**, Balling (2008: 184-185) has suggested the potentially useful application of the PCA to eye-tracking data. The study presented here sought to explore the extent to which the PCA could be usefully applied to the dataset generated from the Mandarin – English eye-tracking experiment. A PCA is used as a descriptive, exploratory method to summarize a dataset in a way which expresses its structure in a small number of dimensions without much loss of information from the original dataset.

A primary aim of a PCA is to identify underlying factors, or components, that explain the pattern of correlations within a set of observed variables. In essence, a PCA aims to reduce the dimensionality of the original dataset by seeking high correlations between a number of variables and condensing the highly-correlated variables into just a few uncorrelated *component(s)* or *score(s)* that can still explain the pattern of the dataset without much information loss. The PCA is a powerful way of identifying patterns in data, and expressing the data so as to highlight their similarities and differences. Using the PCA, the study explored how the variables were correlated and what the major uncorrelated components were. With these components determined, future researchers seeking to conduct similar studies or even replicate the study will be able to focus on the important components, as opposed to having to collect all the data. According to Bartholomew (2002),

> 'The main aim of principal component analysis (PCA) is to replace...correlated variables by a much smaller number of uncorrelated variables which contain most of the information in the original set. This greatly simplifies the task of understanding the structure of the data since it is much easier to interpret two or three uncorrelated variables than 20 or 30 that have a complicated pattern of interrelationships. In order to translate this objective into a practical method, we have to be more precise about what it is to retain "most of the information'" (ibid.: 115).

To illustrate, the data measured in the study shared the same underlying physiological or cognitive causes – the PCA could help identify potential independent components. In addition, if independent components can be revealed based on the correlation structure of the experimental data, future researchers conducting studies with a similar design can focus on collecting a selective number of measurements without losing much information on the dataset. For instance, a study designed to explore the overall academic ability of a high school's first-year students will require collecting all first-year students' scores on all subjects (English, mathematics, biology, history, geography, physical education, physics, and extracurricular activity). If a PCA reveals that their academic performance in English strongly correlates with those in history and geography (component one or Humanities component); that scores on mathematics strongly correlate with those on biology and physics (component two or Science component); and that grades in physical education strongly associate with those in extracurricular activity (component three or Activity Component), then the value of this is that the general academic ability of the first-year students of the school can be efficiently described by the three components. Based on this finding, if a future researcher hopes to predict the overall academic ability of a first-year student in this school or replicate a similar study, s/he does not have to collect scores on *all* subjects. Data on scores on English (from Humanities component), and mathematics (from Science component) and physical education (from Activity component) will be sufficient to describe the dimensionality of the dataset without much information loss. This is because the variables within any given component strongly correlate with one another, and the one variable collected is strongly predicative and indicative of the other variables within the same component (Bartholomew *et al.*, 2002).

In the study presented in this thesis, the PCA was carried out on both translation tasks. It should be kept in mind that the objective of this method is to explore the correlation structure of the five variables and not to compare cognitive loadings of Mandarin → English vs. English → Mandarin translations.

For Mandarin → English translation, three components were identified by the PCA,

and Table 16 shows the correlation of the five variables with the component.

Component Matrix

	Component		
	1	2	3
PupilCE	.244	.009	.967
TimeCE	-.161	.986	.003
BlinkCE	-.899	-.071	-.349
FixCE	.315	.947	.022
FixRateCE	.975	.038	.091

Extraction Method: Principal Component Analysis.

Table 16. Component Matrix with Correlation Coefficients (Mandarin → English)

For each component, there were five numbers corresponding to each of the five variables. These numbers are the correlation between the component and the variables. For example, for the first component (in column one), 0.975 is the correlation between FixRateCE and the component. The study was interested in the high correlations because they would be the ones that identify dataset patterns.

For component one (Fixation Frequency: FixRateCE vs. Blink Frequency: Blink CE), there were two correlations that stood out: FixRateCE and BlinkCE. Also noted was that their correlations with the other two components were low. Individuals who had higher values in FixRateCE would score highly on this component. In contrast, individuals with a low score on this component would tend to have higher BlinkCE values. This further strengthened the result of the correlation analysis (see **Section 4.10.1**). For the study, this component was labeled as '**Frequency Component**'.

For the second component (Task Time: TimeCE vs. Fixation Count: FixCE), there were two high correlations: one corresponding to TimeCE (0.986) and one for FixCE (0.947). This also further strengthened the result of the correlation analysis (see **Section 4.10.1**). Again, their correlations with the other two components were low. Both correlations were

positive, so this component can be interpreted as follows: an individual with a high score on TimeCE and FixCE would have a high score on this component, while an individual who has a low score on each of these two variables would have a low score on this component. For the study, this component was named '**Fixation Component**'.

The third component represents Pupil Size (PupilCE, correlation =0.967). This component was called '**Pupil Component**'. Using the PCA, the potential independent components contributing to the general difference between A→B and B→A translations were found, and these three components could be unpacked by three different groups of experimental data.

Further, Table 17 indicates that the three components explained 97.6% of the total variability. This means that instead of the five original variables, the study could explain the degree of difficulty of the task by using three components or factors, losing only 2.4% of information in the original dataset. In Table 17, under the column of 'Total', eigenvalues are shown; the eigenvalue is important in terms of identifying the strength of a component's ability to account for the variability in the dataset. The sum of eigenvalues is the same as the number of variables that have been entered into PCA. The larger the eigenvalue, the more powerful a component is in terms of describing the variability of the dataset. Therefore, for Mandarin → English translation, the first component (blink frequency and fixation frequency) is the most important component, followed by task time and fixation Count and finally pupil size.

Total Variance Explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	2.396	47.913	47.913
2	1.765	35.296	83.209
3	.722	14.440	97.649
4	.114	2.281	99.929
5	.004	.071	100.000

Extraction Method: Principal Component Analysis.

Table 17. Total Variance Explained by the Components (Mandarin → English)

The results shows that the dataset from the Mandarin – English eye-tracking study can be described by the three components, and that future studies with a similar design can select to collect only fixation frequency, fixation count, and pupil size as data which will sufficiently describe the variance of the dataset. This result will be applied to data collection in the second eye-tracking Spanish – English experiment in **Chapter Five**. This also indicates that eye-tracking has a broader applicability to T&I Studies.

As for English → Mandarin translation, Table 18 shows the correlations of the five variables with the components. Table 19 shows that three components explain 97.9% of the total variability.

Component Matrix

	Component		
	1	2	3
PupilEC	.136	.054	.988
TimeEC	-.112	.986	.052
BlinkEC	-.947	.032	-.240
FixEC	.451	.886	.029
FixRateEC	.962	.207	.001

Extraction Method: Principal Component Analysis.

Table 18. Component Matrix with Correlation Coefficients (English → Mandarin)

Total Variance Explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	2.506	50.129	50.129
2	1.476	29.517	79.646
3	.916	18.313	97.959
4	.085	1.703	99.662
5	.017	.338	100.000

Extraction Method: Principal Component Analysis.

Table 19. Total Variance Explained by the Components (English → Mandarin)

The results show that the difficulty of the task can be represented or explained by the same three components: BlinkEC & FixRateEC (**Frequency Component**); TimeEC & FixEC (**Fixation Component**); and Pupil (**Pupil Component**); this also further strengthened the results yielded by the correlation analysis. The same PCA results from both Mandarin → English and English → Mandarin translations have presented usefulness for future eye-tracking studies with a similar design where not all but a selective number of dependent variables can be collected to maintain the dimensionality of the dataset without much information loss. In terms of the Frequency Component, 'Fixation Frequency' can be selected as a dependent variable, if the study involves a non-head-mounted eye-tracker machine (such as in the studies presented in **Chapters Four & Five**); however, it has to be acknowledged that 'Blink Frequency' can be potentially usefully if the eye-tracking equipment is head-mounted. As for the Fixation Component, given that an eye-tracker is not absolutely required to measure Task Time, the dependent variable of 'Fixation Count' can be selected for future studies. Where the Pupil Component is concerned, it is apparent that pupillometric data are very useful.

To summarize, the PCA results have shown that, regardless of translation directions, the dataset from the Mandarin – English eye-tracking study can be described by the three components, and that future eye-tracking experiments with a similar experimental design can select to collect only 'fixation frequency', 'fixation count', and 'pupil size' as data which will sufficiently describe the dimensions of the dataset. The results have also demonstrated that PCA can be usefully applied to multiple eye-tracking data, as suggested by Balling (2008). The results also indicate that eye-tracking has a broader applicability to T&I Studies. The results were applied to 'data collection' in the second eye-tracking Spanish – English experiment reported in **Chapter Five**.

4.10.3 Longitudinal Analysis

In an experiment where numerical data over a specific period of time is collected and

where a researcher is interested in exploring whether or not a 'pattern' or 'trend' exists in the data over the time period, a longitudinal analysis will be conducted. For instance, if a researcher is interested in exploring a potential trend in the birth rate over the past decade in Japan, data on birth rates will be collected and entered into a longitudinal analysis. In the study presented in this thesis, to explore whether or not the 'familiarity effect' identified during 'interpreting' (Hyönä *et al.*, 1995) also existed during translations in different directions required a longitudinal analysis on the pupillometric data collected.

Before the longitudinal analysis was carried out, the datasets of pupil size over time during translating from Mandarin Chinese into English and translating from English into Mandarin Chinese were tidied up. As a value of 'zero' in the dataset could mean a 'blink' data point or a 'look-down' data point, all the zeros were removed from the datasets. By doing do, the datasets were no longer continuous and this had to be taken into account during the analysis. It was then decided that all the data points in the datasets should be divided into three equal temporal *windows* or *intervals* and that comparisons between the means of all the three intervals would be carried out statistically to see if they were significantly different from each other.

The time points were categorized into three intervals: < 4,000 (interval one), 4,000-8,000 (interval two) and > 8,000 (interval three). The intervals were categorized this way because, based on Figures 12 and 13 of the data points over time, many data points were lost after the 12,000th data point. This is because some of the subjects completed their tasks before the 12,000th data point. In order to retain as many data points as possible within a particular task, every 4,000 data points was categorized as an interval. Figure 18 shows pupil size data over time during Mandarin → English translation, and Figure 19 shows pupil size data over time during English → Mandarin translation.

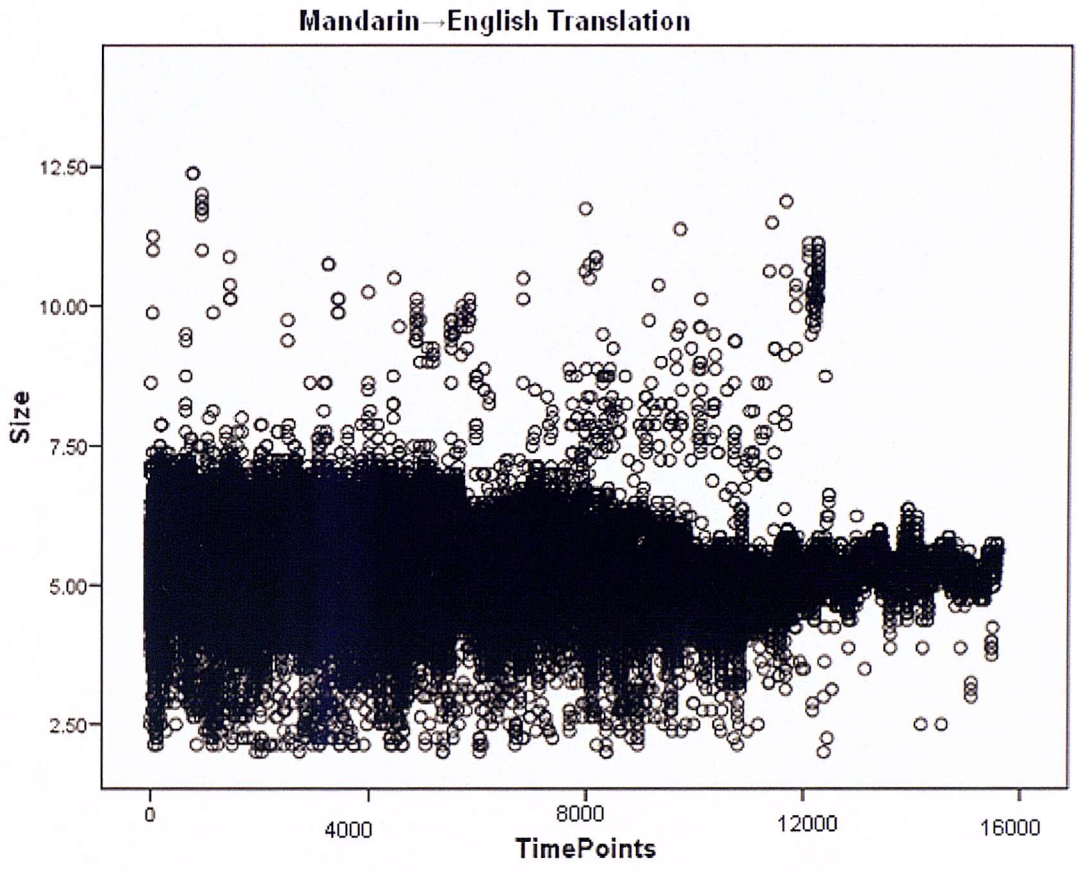


Figure 18. Pupillometric Data (mm) during Mandarin → English Translation

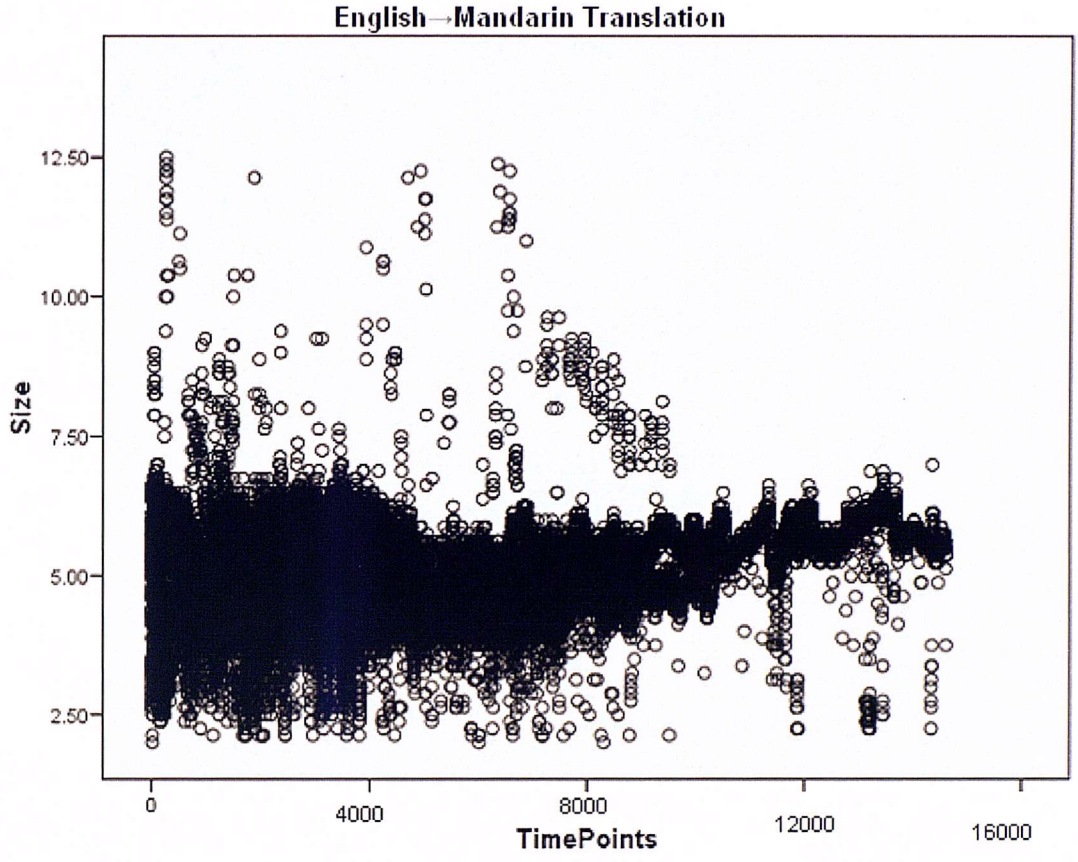


Figure 19. Pupillometric Data (mm) during English → Mandarin Translation

- **Mandarin → English Translation**

Table 20 shows the estimated means of the pupil size data for the three time intervals (named as Time Categories in the table). The table shows that the pupillometric value during Mandarin → English translation went down from the first interval to the second, and yet went up from the second to the third. Table 21 shows that the differences between all the three intervals were significant. To facilitate a sufficient understanding of Table 21, a number of elucidations are described as follows: (1) in the first column of 'Time Category', 1, 2 and 3 refer to the three separate time intervals; (2) each interval was statistically compared with the other two intervals, so, for instance, as shown in the second row of the column, interval 1 was compared with interval 2 and 3; and (3)

all comparisons passed a standard p value of 0.05, as indicated in the 'p-value' column.

Estimates

Time Categories	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1.622	.031	14.003	1.555	1.689
2	1.612	.031	14.003	1.545	1.679
3	1.620	.031	14.004	1.553	1.687

Table 20. Mean Pupillometric Values (mm) of the Three Intervals during Mandarin → English Translation

Pairwise Comparisons

(I) Time Category	(J) Time Category	Mean Difference (I-J)	Std. Error	df	p-value	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	2	.010*	.000	165036.1	.000	.009	.011
	3	.002*	.001	165038.4	.000	.001	.003
2	1	-.010*	.000	165036.1	.000	-.011	-.009
	3	-.008*	.001	165037.9	.000	-.009	-.007
3	1	-.002*	.001	165038.4	.000	-.003	-.001
	2	.008*	.001	165037.9	.000	.007	.009

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

Table 21. Pair-wise Comparisons between the Three Intervals during Mandarin → English Translation

- **English → Mandarin Translation**

The trend regarding pupil size variations over time during English → Mandarin translation was similar to that during Mandarin → English translation. Table 22 shows the estimated means of the pupillometric data for the three time intervals. The table shows that

the mean pupil size during Mandarin → English translation went down from the first interval to the second, and yet went up from the second to the third. Table 23 shows that the differences between all the three intervals were statistically significant.

Estimates

Time Categories	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1.575	.030	14.002	1.511	1.638
2	1.567	.030	14.005	1.504	1.631
3	1.580	.030	14.010	1.517	1.644

Table 22. Mean Pupillometric Values (mm) of the Three Intervals during English → Mandarin Translation

Pairwise Comparisons

(I) Time Category	(J) Time Category	Mean Difference (I-J)	Std. Error	df	p-value	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	2	.007*	.001	116252.7	.000	.006	.009
	3	-.006*	.001	116254.7	.000	-.007	-.004
2	1	-.007*	.001	116252.7	.000	-.009	-.006
	3	-.013*	.001	116251.8	.000	-.015	-.011
3	1	.006*	.001	116254.7	.000	.004	.007
	2	.013*	.001	116251.8	.000	.011	.015

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

Table 23. Pair-wise Comparisons between the Three Intervals during English → Mandarin Translation

The results of the longitudinal analysis have shown that, regardless of directionality, there was a significant difference between all three time intervals (*ps* < 0.001). Interestingly, during both A→B and B→A translations, there was a significant dip in the pupil size, suggesting that less cognitive loading was caused in the middle of the tasks. However, the results have also revealed that there was a significant rise in pupil size during the third

interval, i.e. toward the end of the tasks. This suggests that more cognitive loading was induced toward the end of both tasks, and the results were not supportive of Hyönä *et al.*'s findings of 'familiarity effect' (1995). 'Familiarity effect' (Hyönä *et al.*, 1995) means that gradually less cognitive loading is caused toward the end of both A→B and B→A interpreting. The interpretation by Hyönä *et al.* of this effect was that their interpreter subjects, throughout the course of the experimental procedure, were getting all the more familiar and comfortable with the interpreting tasks in question and that therefore the cognitive loading was significantly lower toward the end of the tasks.

However, during A→B and B→A written translation tasks, if there was a 'familiarity effect', this appeared only transiently halfway through the tasks. A possible interpretation of this disappearance of 'familiarity effect' might be that trained translators have the urge to carry out post-editing or proofreading upon completion of a translation in order to ensure that the translation output is free of errors. This is a possible cause of extra cognitive loading toward the end of a translation task, and therefore of the observed increase in pupil diameter in the third interval. Another possible explanation is that the participants got excited toward the end as they were aware that they were about to complete the task, causing pupil dilation (Cummings, under review). This interpretation warrants further investigation by future researchers.

4.11 Conclusion

To sum up, the results and findings have demonstrated that the eye-tracking technology has a broader applicability to the field of T&I Studies. In addition, they have shown that A→B translation was more cognitively demanding than B→A translation at a textual level. This goes to indicate that the predictions suggested by the Revised Hierarchical Model were valid at a *textual* level.

Chapter Five Eye-tracking Experiment Two: Spanish (A) – English (B)

5.1 Introduction

A second eye-tracking experiment, using eight Spanish native speakers with English as their second language, was conducted. This second eye-tracking experiment incorporated two main merits. Firstly, the results of the second experiment with the language combination of Spanish (A)-English (B) could be used to confirm whether the results from the first Mandarin (A)-English (B) eye-tracking study were generalizable to a language combination involving languages that are both alphabet-based, as the first eye-tracking experiment involved both an ideogram-based language and an alphabet-based language, since the writing systems were different. Secondly, the experimental design was modified, based on the experience of the first experiment, to render it more statistically rigorous.

As previously explained in **Section 3.5.1.2**, 'reading' was introduced as an extra task, alongside typing and translating in the Spanish – English eye-tracking experiment. 'Reading' was incorporated as another level within the factor of Task Type. In this case, six cognitive tasks were involved, i.e. (1) Read Spanish; (2) Read English; (3) Type Spanish; (4) Type English; (5) Translate from English into Spanish; and (6) Translate from Spanish into English. Since there were three levels of Task Type and two levels of Input Language (Spanish vs. English), the experiment used a 3 x 2 within-subject repeated measures factorial design. Further, this second study focused mainly on three dependent variables, i.e. pupil size, fixation count and fixation frequency. This is because the PCA results presented in **Chapter Four** revealed that these three measurements could explain most of the variability in the entire dataset from an experiment with a similar factorial design. Further, as previously stated, even though 'task time' does not require an eye-tracker for

collection, it was still measured with the aim of furnishing further evidence in support of potential findings. ANOVAs were also conducted to determine (1) if Task Type and Input Language had an effect on the dependent variables; and (2) if there was an interaction between the two factors involved. In light of the ANOVA results, post-hoc analyses were carried out for further results with an exploratory purpose. Lastly, a longitudinal analysis was once again carried out to explore the variation in pupil size over time during translation in both directions.

5.2 Behavioral Results

Two Spanish language/translation teachers with English as their second language graded the translation and typing outputs, based on their standard scoring scheme (100=full marks; 60=pass). Table 24 shows the scoring results of typing Spanish, typing English, translating from Spanish into English and translating from English into Spanish:

	N	Minimum	Maximum	Mean	Std. Deviation
Typing Spanish %	8	98	100	99.00	.926
Typing English %	8	98	100	99.50	.926
Translation Spanish → English %	8	90	95	91.13	1.808
Translation English →	8	90	99	93.75	3.105

Spanish %					
Valid Number of Subjects	8				

Table 24. Descriptive Statistics on Scorings (Spanish - English)

The behavioral results revealed that the subjects did conduct the tasks as instructed by the investigator, validating the physiological/visual-attention data, which were then deemed usable for further statistical analyses. Translations into English scored lower than those into Spanish. This suggests that A→B translation was cognitively more demanding than B→A translation. Again, it has to be acknowledged that a translation can be of lower quality for reasons beyond cognitive loading, an interpretation also presented in the first eye-tracking experiment (see also **Section 4.2**).

5.3 Difference in Global Cognitive Loading between Spanish → English and English → Spanish Translation

Paired samples *t* tests were carried out to determine whether or not there was a statistically significant difference between A→B and B→A translations, and the results were as follows:

- Pupillometry (mm):

  It can be inferred that more cognitive loading was induced during A→B translation ($M = 5.33$, $SE = 0.14$) than B→A translation ($M = 4.99$, $SE = 0.18$, $t(7) = -2.62$, $p = 0.034$, $r = 0.70$).

- Fixation Count (integers):

  An inference can be made that more cognitive loading was caused during A→B translation ($M = 46.5$, $SE = 10.2$) than B→A translation ($M = 34.63$, $SE = 6.48$, $t(7) = -$

2.83, $p = 0.026$, $r = 0.73$).

- Fixation Frequency (decimal numbers):

The result suggests that more cognitive loading was created during A→B translation ($M = 0.74$, $SE = 0.036$) than B→A translation ($M = 0.56$, $SE = 0.06$, $t(7) = 5.962$, $p = 0.001$, $r = 0.91$).

To summarize, the above results show that Spanish → English translation was more cognitively demanding than English → Spanish translation. This indicates that the predictions suggested by the Revised Hierarchical Model were valid at a textual level.

5.4 Secondary Results: Attention Allocation to Target Text (TT) and Source Text (ST)

- **Fixation Count on ST or TT during English → Spanish Translation**

A paired samples *t* test reveals that, during translating from English into Spanish, the fixation count on the TT ($M = 68$, $SE = 12.58$) was significantly higher than that on the ST ($M = 8.88$, $SE = 1.88$, $t(7) = -5.24$, $p = 0.001$, $r = 0.82$).

- **Fixation Time (sec) on ST or TT during English → Spanish Translation**

A paired samples *t* test reveals that, during translating from English into Spanish, the fixation time on the TT ($M = 82.34$, $SE = 14.49$) was significantly longer than that on the ST ($M = 11.02$, $SE = 1.65$, $t(7) = -5.33$, $p = 0.001$, $r = 0.90$).

- **Fixation Frequency (decimal numbers) on ST or TT during English → Spanish Translation**

A paired samples *t* test reveals that, during translating from English into Spanish, the fixation frequency on the TT ($M = 0.8$, $SE = 0.03$) was significantly higher than that on the ST ($M = 0.4$, $SE = 0.02$, $t(7) = -6.05$, $p < 0.001$, $r = 0.65$).

- **Fixation Count on ST or TT during Spanish → English Translation**

A paired samples *t* test reveals that, during translating from Spanish into English, the

fixation count on the TT (M = 105.88, SE = 17.44) was significantly higher than that on the ST (M = 19.75, SE = 4.07, $t(7)$ = -5.64, p = 0.001, r = 0.91).

- **Fixation Time on ST or TT during Spanish → English Translation**

A paired samples t test reveals that, during translating from Spanish into English, the fixation time on the TT (M = 112.19, SE = 19.40) was significantly longer than that on the ST (M = 14.46, SE = 4.10, $t(7)$ = -5.64, p = 0.001, r = 0.91).

- **Fixation Frequency on ST or TT during Spanish → English Translation**

A paired samples t test reveals that, during translating from Spanish into English, the fixation frequency on the TT (M = 0.8, SE = 0.02) was significantly higher than that on the ST (M = 0.5, SE = 0.09, $t(7)$ = -5.72, $p < 0.001$, r = 0.9).

This section has presented results regarding attention allocation to ST or TT during A→B and B→A translation. Consistent with the findings reported in **Section 4.5**, the results here have demonstrated that, regardless of directionality, more attention was paid to the TT. To reiterate, a possible explanation is that during translation, translators tend to focus more on the output because the 'cognitive components' required for ST (static text display) involve ST reading and comprehension, whereas the cognitive components required for TT (dynamic text production) involve keyboard search, typing, reading, TT comprehension, editing, and proofreading. This could be why more attention was paid to the TT, and this was demonstrated by Sharmin *et al.* (2008) and Pavlović and Jensen (2009). Once again, this possible explanation merits further research (see **Chapter Seven**). The findings suggest that eye-tracking has a broader applicability to T&I Studies, because, apart from being able to measure cognitive loading, the results and findings from the eye-tracking data can be used to yield hypotheses to be tested by future research in the discipline.

5.5 Repeated measures 2-way ANOVA on Pupillometry

As stated above, a repeated measures 2-way ANOVA was conducted to see (1) if Task Type and Input Language had an effect on pupil size; and (2) if there was an

interaction between Task Type and Input Language (Figure 20).

- Both effects were reported as significant at $p < 0.05$. There was a main effect of Task Type on pupil size, $F = 7.163$, $df = 1.363, 9.540$, $p = 0.018$, $r = 0.65$. Input Language also had a significant effect on pupil size $F = 18.335$, $df = 1, 7$, $p = 0.004$, $r = 0.85$, and there was a significant interaction between Task Type and Input Language, $F = 7.064$, $df = 1.349, 9.442$, $p = 0.019$, $r = 0.65$.

- The results have revealed that (1) the task of translating had a more significant effect on the dilation of the pupil than typing; (2) the input language of Spanish (A→B) had a more significant effect on pupil dilation than English (B→A); and (3) the interaction between Task Type and Input Language showed that the mean pupil sizes for typing Spanish and English were comparable, whereas a major difference was observed when the task was translation. In addition, the pupil size value was at its largest when the task was switched to translation and when the input language was Spanish.

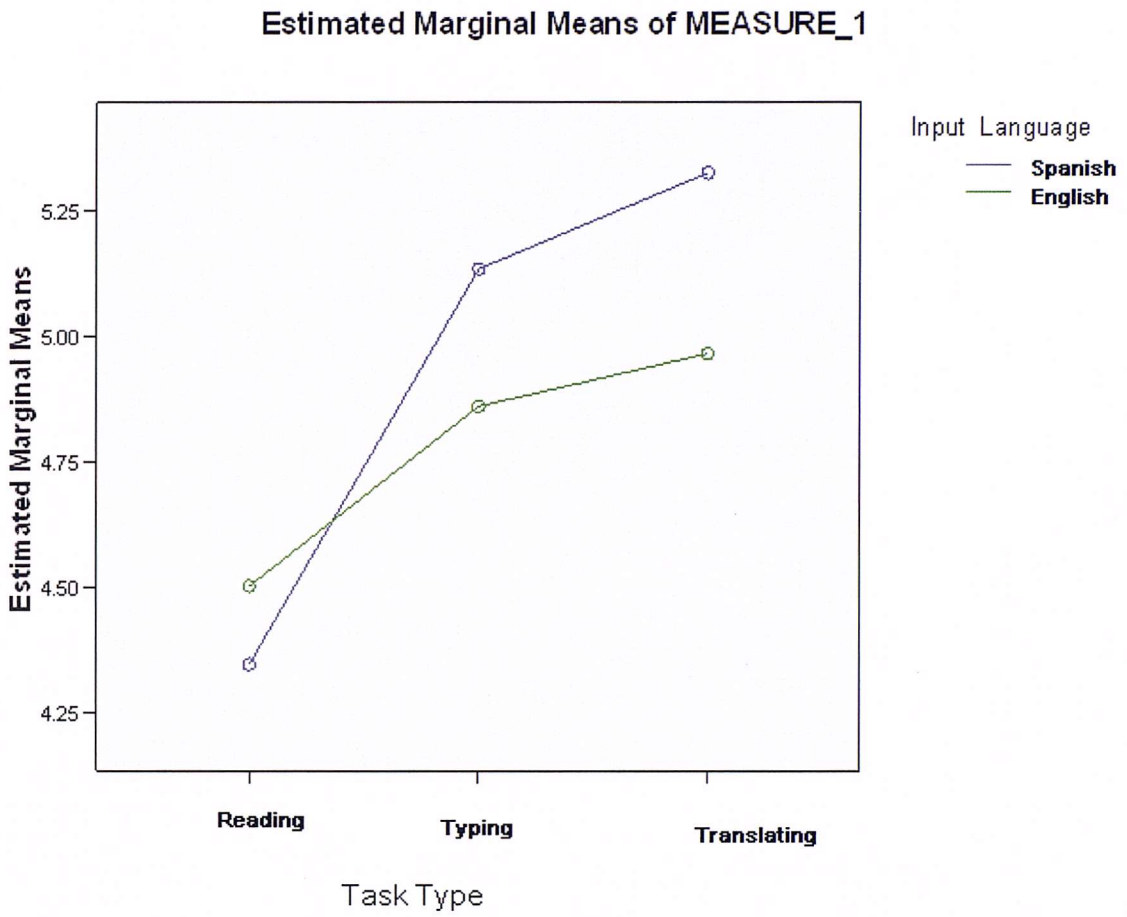


Figure 20. Interaction of Input Language and Task Type (Pupil Size, mm)

5.5.1 Bonferroni-corrected Post-hoc Paired Samples *t* Tests

Bonferroni-corrected post-hoc paired samples *t* tests were conducted to determine if there was a significant difference between the pupil sizes during different tasks. A number of interesting findings are presented as follows:

- **Reading**: Results show a significant difference between reading English and reading Spanish, M = -0.16, SE = 0.04, $t(7)$ = -3.96, p = 0.005, r = 0.83. This suggests that reading English (the second language) was more cognitively demanding than reading Spanish (the first language), which is consistent with the findings reported by Grabe (1991) that the cognitive

loading of reading an individual's second language was higher than that of reading the first language.

- **Typing**: Results show a marginally significant difference between typing English and typing Spanish, $M = 0.27$, $SE = 0.09$, $t(7) = 3.01$, $p = 0.02$, $r = 0.75$. This suggests that typing Spanish was more cognitively demanding than typing English, based on pupillometric data.

The result of the analysis on 'typing' was consistent with the self-reported questionnaire results supplied by seven out of the eight Spanish-English participants (Table 25). The pupil size result suggests that typing an individual's first language was more cognitively demanding than typing the second language.

Reasons Why It Is More Difficult to Type Spanish	
Participant No.	**Reason(s)**
1	*Because of work, I tend to normally use an English keyboard for English outputs.*
2	*Lack of practice.*
3	*Spanish letters are spread out on the keyboard.*
4	*Lack of practice.*
5	*I am more used to typing in English.*
6	*There are more punctuations and symbols in Spanish.*
7	*Spanish accents are not available in a UK keyboard, and out of practice with a Spanish keyboard.*

Reason Why It Is More Difficult to Type English	
8	*I am more experienced with keying in Spanish than English.*

Table 25. Reasons Why It Is More Difficult to Type Spanish or Type English.

However, the observation (based on pupillometric data) that typing an individual's first language seems to be more difficult than typing their second language was not reflected in the Mandarin-English eye-tracking experiments. This is because the 'pupillometric' results yielded from the Mandarin-English eye-tracking experiment indicated that typing English (the second language) was more cognitively demanding than typing Mandarin (the first language). Interestingly, the pupillometric results confirmed participants' own opinions on the relative difficulty of typing their first language using an English keyboard in one experiment, but not the other. A possible reason behind the result inconsistency is that the language combinations (Mandarin - English vs. Spanish - English) were different. The former involved an ideogram-based language, whereas the latter involved two alphabet-based languages. A potential interpretation is that, if an individual's first language is an ideogram-based language (e.g. Chinese Mandarin) that is vastly different from an alphabet-based language (e.g. English), pupil size during typing the second language on an 'English keyboard' would be generally larger, which is indicative of a higher level of cognitive loading. Moreover, if an individual's first language is an alphabet-based language (e.g. Spanish), the level of cognitive loading regarding typing on an English keyboard tends to be associated with the alphabetical complexity of the language (e.g. typing Spanish sometimes require two keystrokes to produce an *accented* letter), or with habit and practice.

In summary, the results presented above go to demonstrate that A→B translation was cognitively more demanding than B→A translation, even with an alphabet-based language combination, and that the predictions suggested by the Revised Hierarchical Model were valid at a textual level. Further, the post-hoc analyses have yielded interesting

findings that warrant future testing. This also indicates that eye-tracking has a broader viability to T&I Studies

5.6 Repeated Measures 2-way ANOVA on Fixation Count

Results reveal that Task Type had a major effect on fixation count, $F = 10.836$, $df = 1.340, 9.381$, $p = 0.006$, $r = 0.73$. Input Language had a marginally significant effect on pupil size $F = 5.294$, $df = 1, 7$, $p = 0.055$, $r = 0.66$. There was a significant interaction between Task Type and Input Language, $F = 10.903$, $df = 1.648, 11.537$, $p = 0.03$, $r = 0.95$. Figure 21 shows the relationship between Input Language, Task Type and fixation counts.

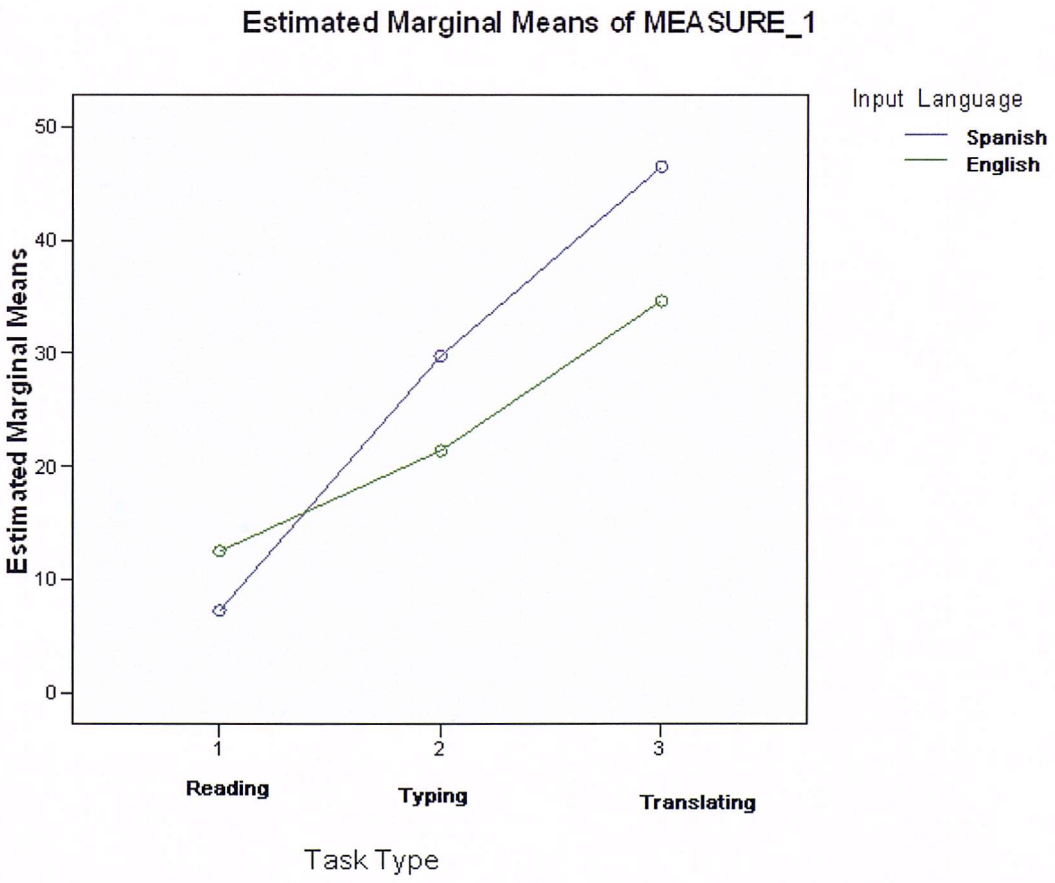


Figure 21. Interaction of Input Language and Task Type (Fixation Count, integers)

The main effect of Task Type (reading, typing and translating) on fixation count means that fixation counts varied significantly when the task type changed. In addition, regardless of translation direction, the number of fixations was a function of the Input language. Thus, in the data that involved Spanish as the input language, Spanish → English translation had the highest number of fixations, followed by typing Spanish and then reading Spanish; similarly, in the data with English as the input language, English → Spanish translation had the highest number of fixations, followed by typing English and then reading English. Bonferroni-corrected post-hoc paired samples *t* tests were conducted to see if these differences were significant.

- **Reading**: Fixation-based results show a marginally significant difference between reading English and reading Spanish, M = -5.25, SE = 1.85, $t(7)$ = -2.84, p = 0.025, r = 0.73. This suggests that reading Spanish (the first language) was less cognitively demanding than reading English (the second language). This is consistent with the pupillometric results in the study, and also supports the findings reported by Grabe (1991).

- **Typing**: Results show a marginally significant difference between typing English and typing Spanish, M = 14.625, SE = 3.78, $t(7)$ = 4.09, p = 0.005, r = 0.84. This suggests that typing Spanish on an English keyboard was cognitively more demanding than typing English. The result is consistent with the result yielded by the pupil size data. In addition, the result is consistent with the fixation-based result from the Mandarin – English eye-tracking experiment, i.e. typing Mandarin on an English keyboard was more cognitively demanding than typing English, based on the fixation data.

In **Sections 4.6.1** and **4.6.2**, a discussion regarding the inconsistency between pupillometric and fixation results in the Mandarin – English eye-tracking experiment suggested that such a inconsistency was more likely to occur when an ideogram-based language was involved, whereas the two types of results tended to be consistent when the languages involved were both alphabet-based, as indicated by the study of Pavlović and

Jensen (2009). The pupil and fixation results reported from the Spanish – English experiment also seemed to support this suggestion. Intuitively, typing Mandarin as an ideogram-based language using an English keyboard should be more cognitively demanding than typing English. This was confirmed by the results of fixation count, fixation frequency, and task time, but not by pupil size (see **Section 4.6.2** and **Sections 4.7** to **4.9**). The discussion presented in **Section 4.6.2**, in combination with the results yielded from the Spanish – English experiment as well as the study by Pavlović and Jensen (2009), suggests that, in terms of measuring cognitive loading of translation directions involving an ideogram-based language, pupillometric data could be relatively less reliable. Again, whether or not this idea holds true has yet to be further verified.

In spite of the above suggestion, the results of fixation count, fixation frequency and task time in relation to typing a person's first language and second language in the Mandarin – English eye-tracking experiment, as well as in the Spanish – English experiment (see also **Sections 5.7** and **5.8**), suggested that typing an individual's first language that was orthographically more complex, using an English keyboard, was more cognitively demanding than typing English. However, it has to be acknowledged that typing one's first language could be less cognitively demanding if an appropriate keyboard was provided.

5.7 Repeated measures 2-way ANOVA on Fixation Frequency

Results show that Input Language has an effect on fixation frequency ($F = 048$, $df = 1, 7$, $p < 0.001$, $r = 0.87$). However, Task Type did not have a significant effect on fixation frequency ($p = 0.097$). Further, there was a significant interaction between Task Type and Input Language ($F = 0.39$, $df = 1.675, 11.728$, $p = 0.011$, $r = 0.64$). Figure 22 shows the interaction between the two factors.

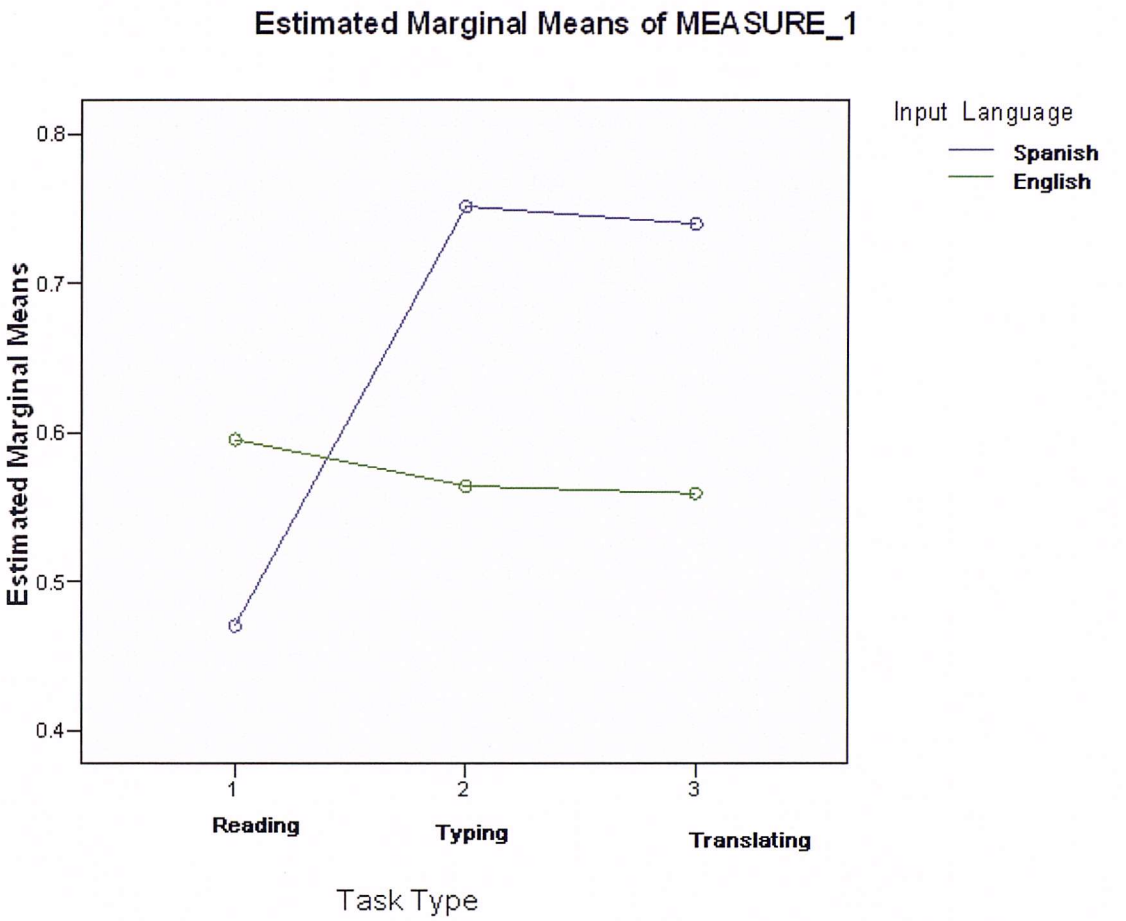


Figure 22. Interaction of Input Language and Task Type (Fixation Frequency, decimal numbers)

The figure suggests that there might be a significant difference between fixation frequencies during typing Spanish and typing English, as well as between reading Spanish and reading English. Post-hoc Bonferroni-corrected paired samples *t* tests were carried out, and the results are described as follows:

- **Reading**: Results generated by the fixation frequency data have shown that, on average, fixations were more likely and more frequent while reading English (M = 0.47, SE = 0.08) than while reading Spanish (M = 0.6, SE = 0.09, $t(7)$ = -3.666, p = 0.008, r = 0.81). Again, this supports the suggestion that reading an individual's first language is less cognitively demanding than

reading the second language (Grabe, 1991). This result based on the fixation frequency data is also consistent with the results yielded by data on both pupil size and fixation count.

- **Typing**: Results have shown that, on average, fixations were more likely and more frequent to appear while typing Spanish (M = 0.75, SE = 0.05) than while typing English (M = 0.56, SE = 0.83, $t(7)$ = 3.472, p = 0.01, r = 0.8). This suggests that typing Spanish was cognitively more demanding than typing English. This result is consistent with the pupil- and fixation-based results reported above. In addition, the result is also consistent with the results from the Mandarin – English eye-tracking experiment.

5.8 Repeated measures 2-way ANOVA on Task Time

As stated in **Section 5.1**, even though task time was not intended as one of the key dependent variables to be collected at the outset of the Spanish – English eye-tracking experiment, the data was collected and analyzed with the aim of furnishing further evidence regarding the potential effect of the factors. The results show that Task Type had a significant effect on task time (F = 5.215, df = 1.021, 7.144, p = 0.05, r = 0.64), and that Input Language also had a significant effect on task time (F = 18.555, df = 1, 14, p = 0.001, r = 0.75). Further, there was a significant interaction between Task Type and Input Language (F = 11.481, df = 1.228, 8.597, p = 0.007, r = 0.76). Irrespective of the input language involved, translating took longer to complete than typing or reading. Similarly, irrespective of whether the task was typing or translating it took a longer time to complete the task when the input language was Spanish than when it was English. However, when the task was reading, it took longer in English than in Spanish.

In terms of the interaction effect of Task Type and Input Language, Figure 23 shows that the type of task had a different effect on task time, depending on the input language involved. Translating from Spanish into English took the longest time, followed by translating from English into Spanish, then typing Spanish, typing English, reading English

and reading Spanish.

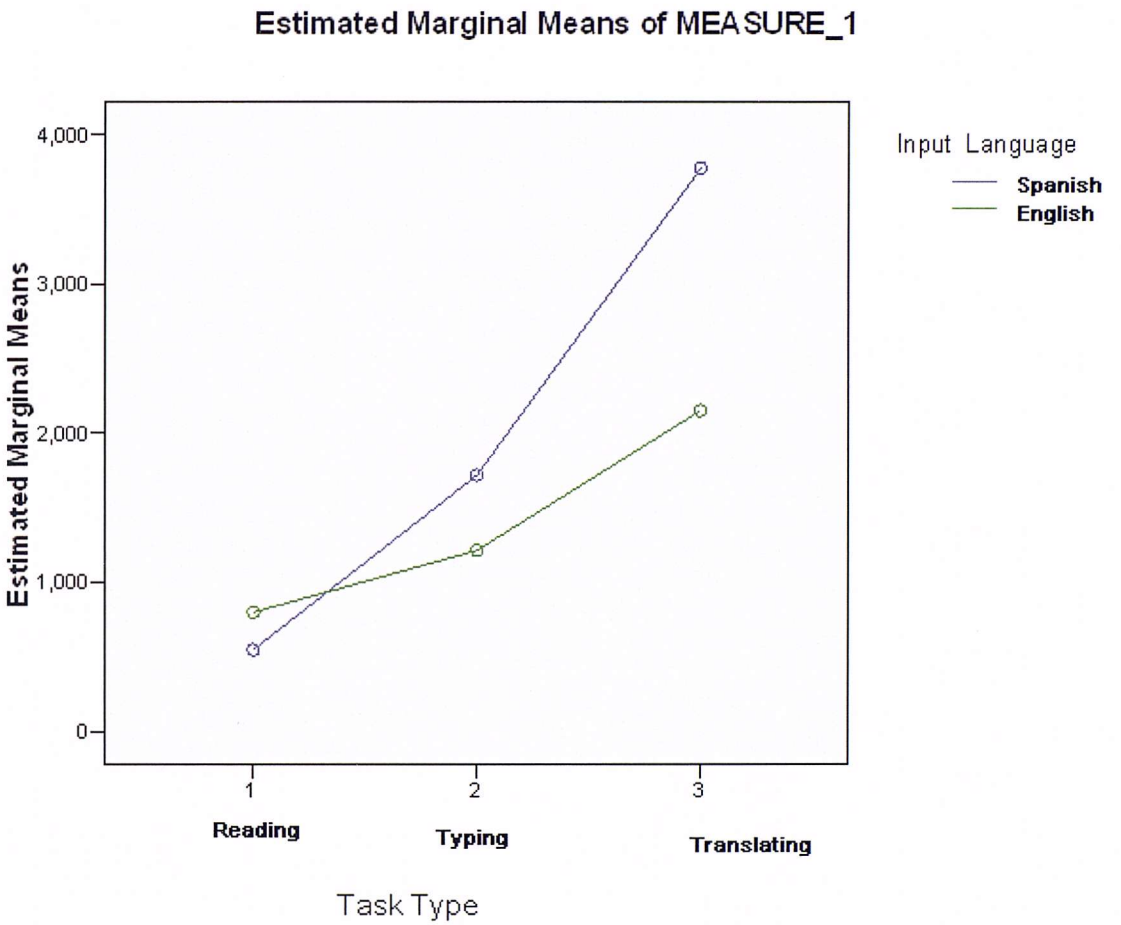


Figure 23. Interaction of Input Language and Task Type (Task Time, sec)

- **Reading**: Results have shown that, on average, task time for reading Spanish (M = 547.25, SE = 70.617) was shorter than that for reading English (M = 797.63.56, SE = 47.57, $t(7)$ = -4.101, p = 0.05, r = 0.82). Again, this is in support of the suggestion by Grabe (1991).
- **Typing**: Results have shown that, on average, task time required for typing Spanish (M = 1717.75, SE = 210.896) was longer than that for typing English

(M = 1212.88, SE = 143.161, $t(7)$ = 2.883, p = 0.024, r = 0.74). The result further strengthens the findings generated by the data on pupil size, fixation count and fixation frequency: typing Spanish was cognitively more demanding than typing English. The result is also consistent with the result from the Mandarin – English eye-tracking experiment.

To summarize, the above results (based on pupil size, fixation count, fixation frequency and task time) have all demonstrated that, across all the participants, reading English yielded more cognitive loading than reading Spanish. This suggests that reading English, whether it is cognate with the first language or not, generates more cognitive loading. Two possible reasons for this suggestion are as follows: (1) reading a second language, i.e., English, was generally more cognitively demanding than reading one's first language (Grabe, 1991); and (2) the time or years spent on using a particular language can have an effect on cognitive loading when reading the language, while as all the participants were Spanish native speakers with more years spent using the language practically, the level of cognitive loading caused by reading in Spanish would be lower than that caused by reading a second language. In addition, the above results (based on pupil size, fixation count, fixation frequency and task time) have all shown that, across all the participants, typing English yielded less cognitive loading than typing Spanish. Two potential explanations are as follows: (1) the relative complexity of the orthographies of the languages may affect a subject's cognitive loading while typing her/his first and second language; and (2) the type of keyboard used may have an effect on the cognitive loading of typing; therefore, for the subjects to generate Spanish words or Mandarin characters using an English keyboard would be more difficult than generating English characters on an English keyboard, despite the fact that Spanish or Mandarin is the subject's native language. Again, it has to be taken into account that typing one's first language could be less cognitively demanding if an appropriate keyboard was provided.

5.9 Longitudinal Analysis

All procedures of the longitudinal analysis presented in this section are the same as those of the longitudinal analysis conducted for the first eye-tracking experiment (see **Section 4.10.3**). To reiterate, all the zeros were removed from the dataset. Afterwards, all the data points in the dataset were divided into three equal temporal *windows* or *intervals* and comparisons between the means of all the three intervals were carried out statistically to see if they were significantly different from each other.

Figure 24 shows pupil size (unit: mm) data points over time during Spanish → English translation, and Figure 25 shows pupil size data over time during English → Spanish translation.

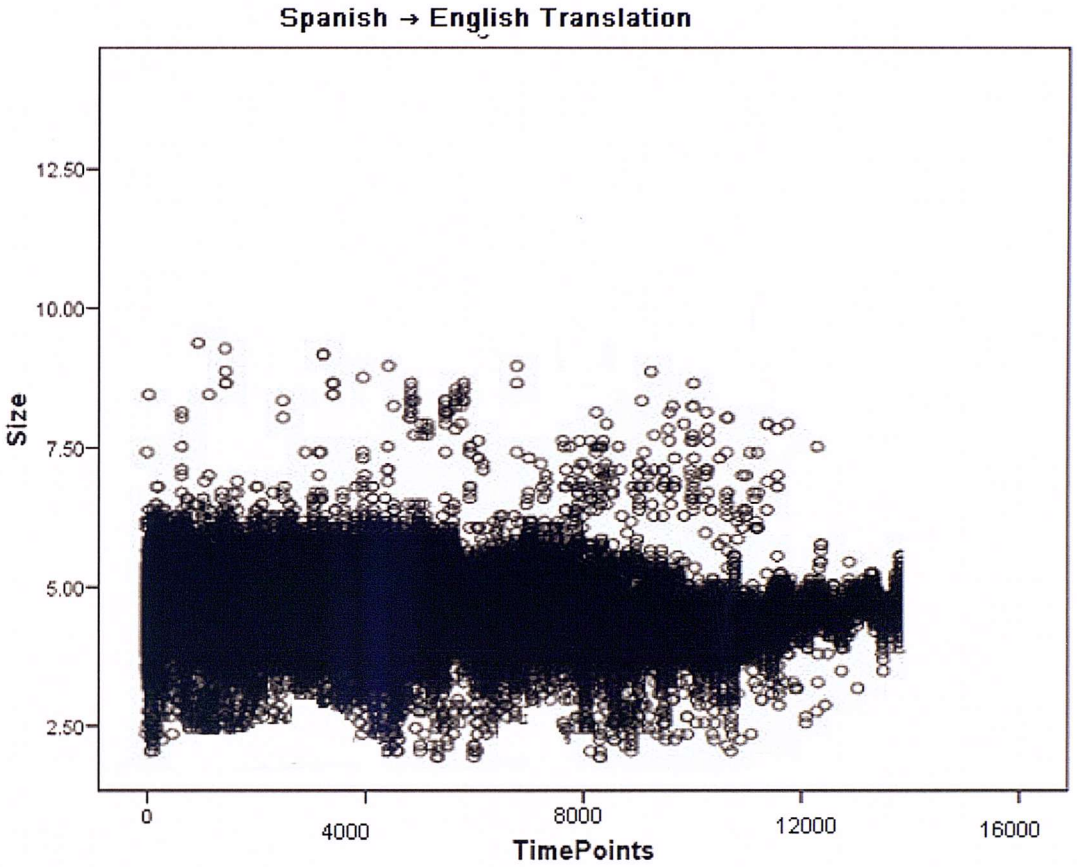


Figure 24. Pupillometric Data (mm) during Spanish → English Translation

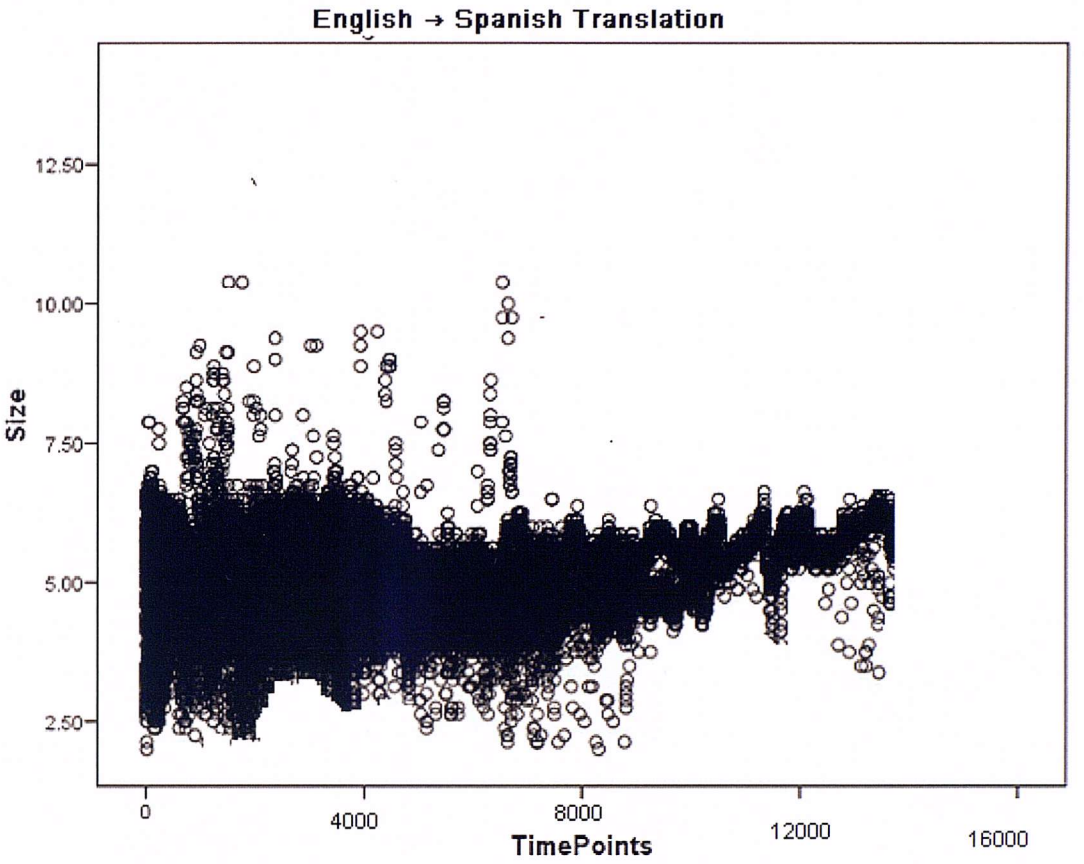


Figure 25. Pupillometric Data (mm) during English → Spanish Translation

➢ **Spanish → English Translation**

Table 26 shows the estimated means of the pupil size data for the three time intervals (named as Time Categories in the table). The table shows that the pupillometric value during Spanish → English translation went down from the first interval to the second, and yet went up from the second to the third. Table 27 shows that the differences between all the three intervals were significant.

Estimates

Time Categories	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1.711	.029	7.003	1.495	1.589
2	1.701	.029	7.003	1.434	1.579
3	1.719	.029	7.003	1.483	1.597

Table 26. Mean Pupillometric Values of the Three Intervals during Spanish → English Translation

Pairwise Comparisons

(I) Time Category	(J) Time Category	Mean Difference (I-J)	Std. Error	df	p-value	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	2	.009*	.000	158766.2	.000	.009	.011
	3	.003*	.000	158760.4	.000	.002	.002
2	1	-.009*	.000	158766.2	.000	-.009	-.009
	3	-.008*	.000	158764.3	.000	-.009	-.007
3	1	-.003*	.000	158760.4	.000	-.002	-.001
	2	.009*	.000	158764.3	.000	.009	.009

Based on estimated marginal means

*. The mean difference is significant at the .05

Table 27. Pair-wise Comparisons between the Three Intervals during Spanish → English Translation

➢ **English → Spanish Translation**

The trend regarding pupil size variations over time during English → Spanish translation is similar to that during Spanish → English translation. Table 28 shows the estimated means of the pupillometric data for the three time intervals. The table shows that the mean pupil size during Mandarin → English translation went down from the first interval to the second, and yet went up from the second to the third. Table 29 shows that the differences between all the three intervals were statistically significant.

Estimates

Time Categories	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1.675	.030	7.004	1.4781	1.462
2	1.666	.030	7.004	1.464	1.442
3	1.679	.030	7.004	1.472	1.465

Table 28. Mean Pupillometric Values of the Three Intervals during English → Spanish Translation

Pairwise Comparisons

(I) Time Category	(J) Time Category	Mean Difference (I-J)	Std. Error	df	p-value	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	2	.005*	.000	105434.8	.000	.006	.009
	3	-.005*	.000	105587.9	.000	-.005	-.005
2	1	-.005*	.000	105434.9	.000	-.005	-.005
	3	-.010*	.000	105666.8	.000	-.010	-.011
3	1	.005*	.000	105587.9	.000	.005	.007
	2	.010*	.000	105666.6	.000	.010	.010

Based on estimated marginal means

*. The mean difference is significant at the .05

Table 29. Pair-wise Comparisons between the Three Intervals during English → Spanish Translation

Like the results from the first Mandarin – English eye-tracking experiment, the results of the longitudinal analysis presented in this section have shown that, regardless of directionality, there was a significant difference between all three time intervals (*ps* < 0.001). In particular, during both A→B and B→A translations, there was a significant decrease in the pupil size, and this suggests that the level of cognitive loading was lower in the middle of the tasks. However, the results have also revealed that there was a significant hike in

pupil size during the third interval, i.e. toward the end of the tasks. This suggests that more cognitive loading was induced toward the end of both tasks, and the results were not supportive of Hyönä *et al.*'s findings regarding a 'familiarity effect' (1995), as was the case with the Mandarin – English study.

To elaborate again, the interpretation by Hyönä *et al.* (1995) of this effect was that their *interpreter* subjects were over time getting all the more familiar and comfortable with the interpreting tasks in question and that therefore significantly less cognitive loading was caused toward the end of the tasks. However, during A→B and B→A *written translation* tasks, a 'familiarity effect' appeared only transiently halfway through the tasks for both eye-tracking experiments. Again, a possible interpretation of this disappearance of any 'familiarity effect' might be that trained translators have the urge to carry out post-editing or proofreading upon completion of a translation in order to ensure that the translation output is free of errors. This is a possible cause of extra cognitive loading toward the end of a translation task, and therefore of the observed increase in pupil diameter in the third interval. Another possible explanation is that the participants got excited toward the end as they were aware that they were about to complete the task, causing pupil dilation (Cummings, under review). This explanation warrants further investigation by future researchers.

5.10 Conclusion

The conclusion of this chapter shares similarities with that of **Chapter Four**. To reiterate, the results and findings have demonstrated that the eye-tracking technology has a broader applicability to the field of T&I Studies, because it has been shown that other types of eye-tracking data (e.g. fixation freqeuncy) were useful in measuring cognitive loading, and because eye-tracking can generate results and findings as the basis of more concrete hypotheses to be tested by future research in the discipline. In addition, the results have shown that A→B translation was more cognitively demanding than B→A translation, and this indicates that the predictions suggested by the Revised Hierarchical Model were valid

at a textual level.

Chapter Six Exploratory fMRI Experiment: Mandarin-English

6.1 Introduction

Functional magnetic resonance imaging (fMRI) is employed to investigate neurological activities accompanying a variety of cognitive tasks. To date, the use of fMRI to study T&I has been minimal. This raises the question of whether or not fMRI is a viable research method for the field of T&I Studies. Providing a potential answer to the question will be a useful supplement to the eye-tracking experiments described in the previous chapters, i.e. demonstrating that both fMRI and eye-tracking technologies can be adequately applied as methods for T&I Studies.

In this chapter, gross brain anatomy will be presented to formulate a planar description of the brain as well as its fundamental cognitive functions. Physical principles of magnetic resonance imaging and functional magnetic resonance imaging are also introduced (for a detailed introduction, see Abutalebi *et al.*), alongside a description of how a brain image is formed and acquired. Further, the fMRI experiment design, data analysis, results, and findings are described.

6.2 Physical Principles of MRI vs. Functional MRI (fMRI)

Magnetic resonance imaging (MRI) is one of the most influential developments in the field of medicine in the 20th century, with the first MRI image published in 1973 (Lauterbur, 1973). Its inventors, Peter Mansfield and Paul Lauterbur, were awarded the Nobel Prize in Physiology or Medicine in 2003, which goes to demonstrate the significance of the invention of MRI.

It is important to understand the distinction between MRI (structural imaging) and functional MRI (functional imaging). Structural imaging is employed to yield three-

dimensional images of body tissues, e.g. brain, lung, liver, etc., based on the fact that different tissues have varied physical properties. These properties can be utilized to map detailed, 'static' images of the brain. Functional imaging relies on the assumption that neuronal activity leads to moment-to-moment regional physiological variations in different areas of the brain. This makes it possible to map moment-to-moment, 'dynamic' images of the brain during a cognitive task. When the static and dynamic brain maps are superimposed during analysis, precise localization of activated brain regions accompanying the cognitive task can then be determined. The next two sections further outline the basic principles of these twin techniques: (1) the physical principles of MRI, and (2) the assumptions behind functional MRI.

6.2.1 MRI

MRI is applied to image soft issues of the body to generate structural images. Most of the soft tissues, e.g. the brain, are water-based, and different tissues have different amounts of water. This causes different tissues to behave differently when stimulated by magnetization. In MRI, this is the fundamental concept behind generating a three-dimensional brain volume or image. An MRI brain volume is made up of multiple 'slices' or 'planes', with each slice representing an approximately two mm thickness of brain tissue scanned. A slice is made up of 'voxels', where 'voxel' means a three-dimensional image unit generally of 2 mm x 2mm x 2mm, just like the term 'pixel' presenting a two-dimensional image unit. Simply put, a three-dimensional whole brain volume consists of slices of brain image presented in voxels, and can be obtained within the order of two to three seconds by rapidly acquiring slice after slice of scanned brain tissue (see Savoy, 2002).

Seventy-eight percent of a brain is made up of water molecules (McIlwain and Bachelard, 1985), and it is the variations of magnetic moments of single protons found in water molecules in the brain (the hydrogen nuclei of H_2O) that form the source of the signal for an MRI scanner (Gazzaniga *et al.*, 2002). Based on the principle outlined above, acquiring a structural image of a whole brain requires the following sequential events.

Firstly, a strong, external magnetic field is applied to the whole brain. Before this, the magnetic fields of the protons, depending on the different tissue (e.g. white matter, gray matter or cerebral fluids), are randomly oriented. However, once the external magnetic field is applied, the protons will randomly align parallel or anti-parallel to the applied field (Figure 26 & Figure 27). The external magnetization remains constant in strength and is consistently applied during scanning[18].

[18] The unit used to measure the strength of the magnetization is called Tesla (or T for short). A typical MRI laboratory will have a MRI scanner of 1.5 T or 3 T magnetization field strength. By comparison, the Earth's magnetic field is approximately 0.0001 T (Merrill and McElhinny, 1983). Although 7 T scanners are now being manufactured for use with an even higher imaging spatial resolution, whilst maintaining arguably the highest level of safe magnetization.

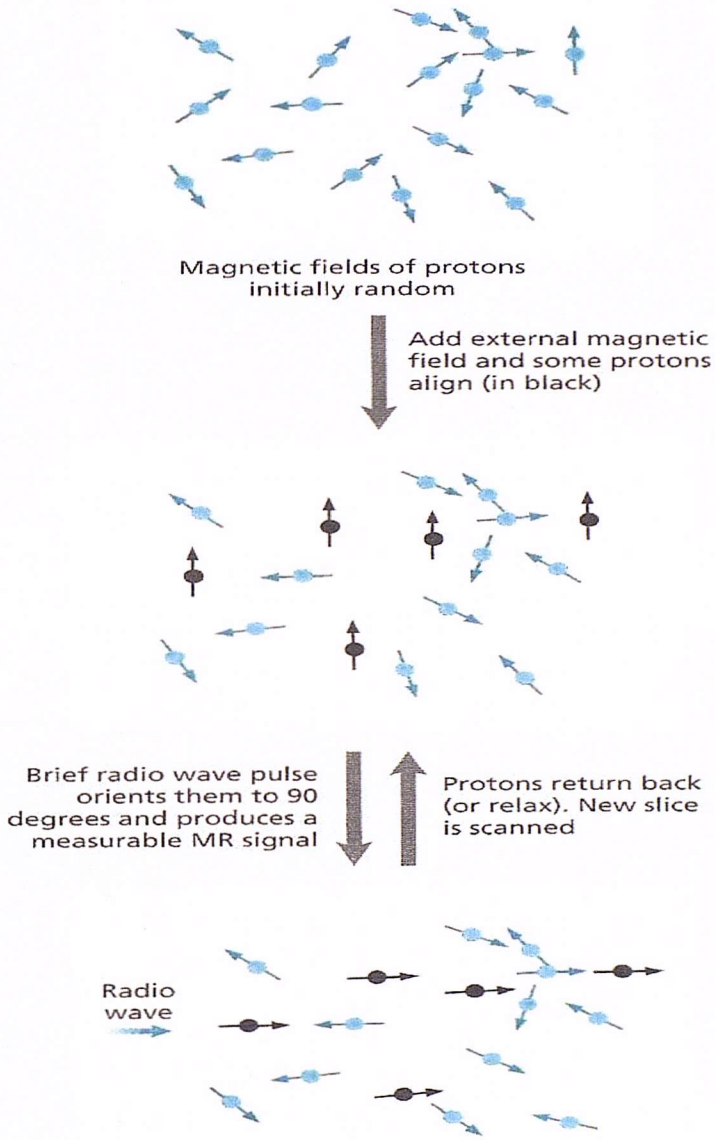


Figure 26. Alignment of Protons to Magnetization

Source: (Ward, 2006: 51)

When the protons are forced to align with the external magnetization to reach a magnetic equilibrium, a brief radio frequency pulse is applied to excite them and make them deflect from the aligned orientation by 90 degrees (Figure 27).

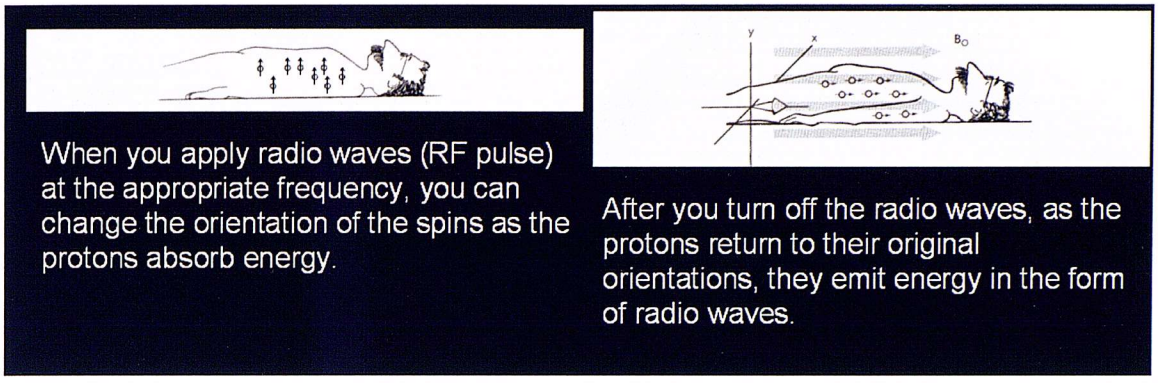


Figure 27. States of Protons When Radio Frequency is Applied vs. Off

Source: (Culham, Online: 8)

During a *spin* of the protons (reorientation or in a technical expression *precess*) to the new post-90°-rotation position, the protons yield variations in the magnetic field and the signals are picked up by an MRI scanner. After the radio wave excitation, the protons will be pulled back to their original equilibrium alignment by the same external magnetization; this process is called *relaxation* (Figure 26). The MRI scanner repeats this process by applying a radio frequency pulse to the brain, and measures signals from slices of the brain sequentially.

The scanner uses sophisticated electronic technology to detect these minute changes in nuclear magnetic responses to distinguish different kinds of tissue. In the case of the brain, for instance, it is able to distinguish grey matter from white matter. Different MRI signals generate different kinds of image. 'Signal differences' in how quickly protons in different tissues manage to return to their original magnetic equilibrium orientation after the sending of a radio frequency pulse are referred to as 'T_1 (Tesla 1) relaxation time'. The signal differences are used to distinguish how images of different tissues are displayed. Generally, in a T_1-weighted image (Figure 28), gray matter appears gray and white matter appears white in color. Further, there is the other signal component called T_2 (Tesla 2), generated due to the fact that, during the post-90°-pulse state, MR signal decays as a result of molecular interaction (local transverse magnetic field inhomogeneities) and also that the

magnetic field changes. This forms the source of signal for functional imaging (explained in the next section), as deoxyhaemoglobins and oxyhaemoglobins create distortions in this component and help form a 'T_2 image' (Figure 29) (Thulborn *et al.*, 1982; for a more detailed intoduction, see Buxton, 2002; Huettel *et al.*, 2003).



Figure 28. T_1-weighted Image

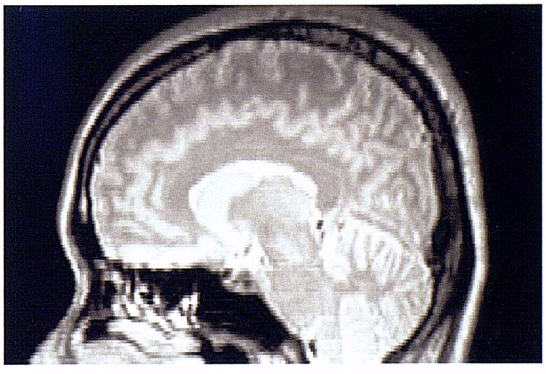


Figure 29. T_2 Image.

6.2.2 Functional Magnetic Resonance Imaging (fMRI)

A human brain is active at all times, requiring metabolites, e.g. oxygen and glucose, for it to be fully operational. Blood is responsible for supplying these metabolites to the

brain. Where functional MR imaging is concerned, oxygenated haemoglobins and deoxygenated haemoglobins are the source of MR signals called BOLD (blood oxygenation level dependent) signals. When neuronal activity increases in a particular brain region, more oxygen is supplied by blood flow rich with oxygenated haemoglobins. As deoxygenated haemoglobins are paramagnetic, the ratio between oxyhaemoglobins and deoxyhaemoglobins is measured by an MR scanner during a cognitive task when the brain region is turning oxyhaemoglobins into deoxyhaemoglobins. The magnetic inhomogeneities, forming the BOLD signal, are detected by the scanner. In other words, the paramagnetic properties of deoxygenated haemoglobins act as markers of the locality of an activated region for a cognitive task, as well as the intensity of neuronal activity accompanying the task. Figure 30 shows the linear relationship between a stimulus and BOLD signal. When a stimulus is presented to a subject, neurons required to process the stimulus will be excited or maybe inhibited, and this requires, as previously described, metabolites and oxygen. For this, extra blood flow will flux to the area where a population of neurons has been activated for the stimulus. The proportion between oxyhaemoglobins and deoxyhaemoglobins generates changes in the strength of the local magnetic field, and this forms the BOLD signals picked up by the scanner.

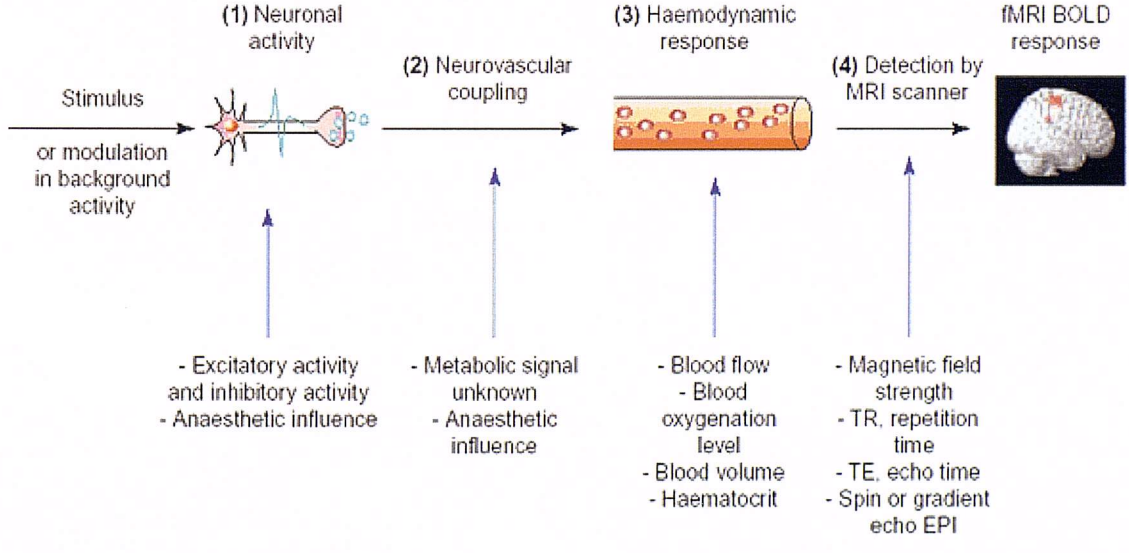


Figure 30. Linearity of a Stimulus and BOLD Signal

Source: (Arthurs and Boniface, 2002: 28)

6.2.3 Brain Volume Formation

MRI results are presented as a series of planar images of brain slices. Figure 31 shows the sequential planar image acquisition of an fMRI scanner. However, it is an example of a brain volume consisting of fifty planar images. In the study, fifty slices were acquired to form a brain volume. In the figure, from left to right and from the upper left image to the lower right image, it represents a series of horizontal planar slices through the head moving gradually upwards.

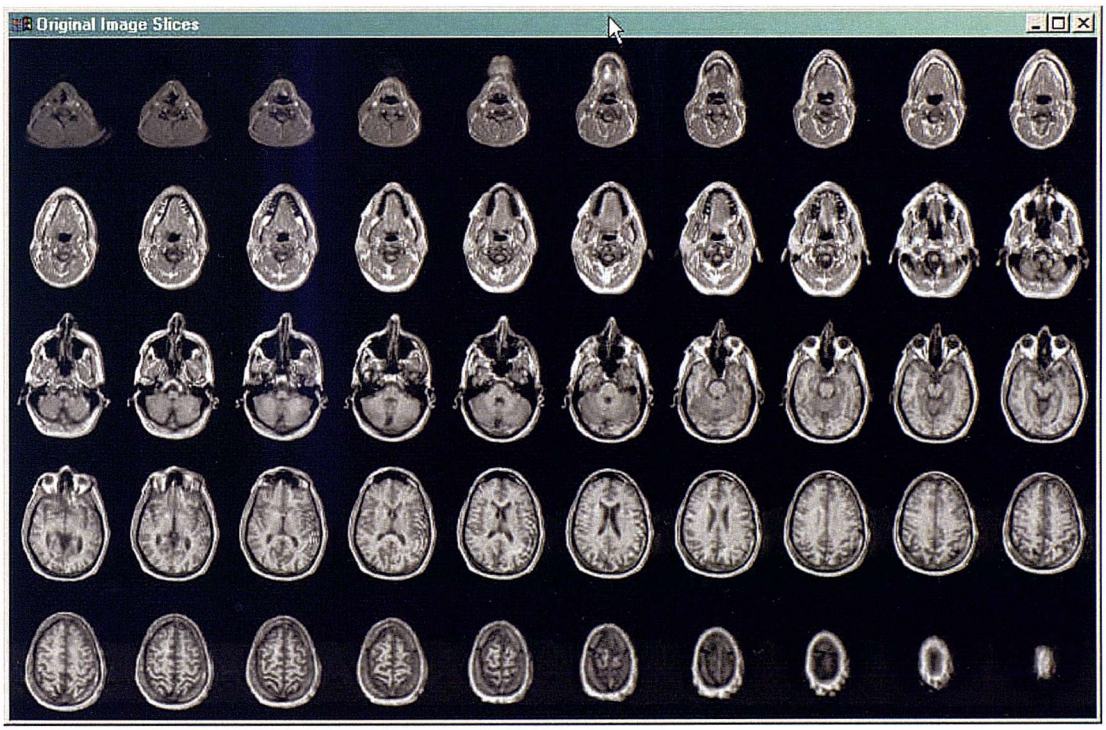


Figure 31. PLANAR Image Sequential Acquisition

Source : (Fanning, Online)

6.3 Brain Anatomy and Cognitive Functions

To reiterate what was mentioned in **Section 1.1,** interpreting requires the ability of 'concurrent' comprehension of a text in an input language (presented either in an auditory or visual format) and verbal production of the same message in an output language. Interpreters need to process a number of cognitive components: (1) listen to or read large chunks of information in a source language; (2) extract and temporarily store the information; and (3) reproduce it in a target language (Gile, 1995). Therefore, interpreting involves processing language-related multi-sensory information, except olfactory, tactile and taste faculties. In this section, a general, gross brain anatomy is introduced, followed by a description of specific brain areas that have been demonstrated to subserve language-related multi-sensory processing. Because of the experimental paradigm, i.e. *'sight* interpreting' (see **Section 6.6**) employed in the study, additional anatomical information regarding visual processing is included.

The cerebral cortex consists of folded gray matter organized in two hemispheres (right and left). The raised folds of the cortex are called 'gyri' (or gyrus in its singular form) and the grooves are called 'sulci' (or sulcus in its singular form). The cortex is divided into four lobes: frontal, temporal, parietal and occipital lobe. Figure 32 shows a planar description of the four lobes that make up the cerebrum or the forebrain. The rest of the observable brain includes the midbrain (superior and inferior colliculi and substantia nigra) and hindbrain (pons, cerebellum, and medulla oblongata).

Superior and inferior colliculi (its singular form is colliculus) are visual and auditory pathways that help humans to react very swiftly to sudden visual or aural stimuli, e.g. light flashes and explosions, before they are consciously processed (Harms and Melcher, 2002; Rinne *et al.*, 2007).

The cerebellum (little brain) consists of folds of gray matter much more convoluted than those of the cerebrum. It is associated with reflexes and the smooth execution of movement by integrating motor activities with sensory inputs (Gazzaniga *et al.*, 2002). The

pons serves as the communication bridge between the cerebrum and the cerebellum, receiving visual information to control movements of the eye and body (Park *et al.*, 2008). The medulla oblongata converges with the spinal nerves. It is responsible for such vital functions as heartbeat, breathing, swallowing, etc. (Bear *et al.*, 2007).

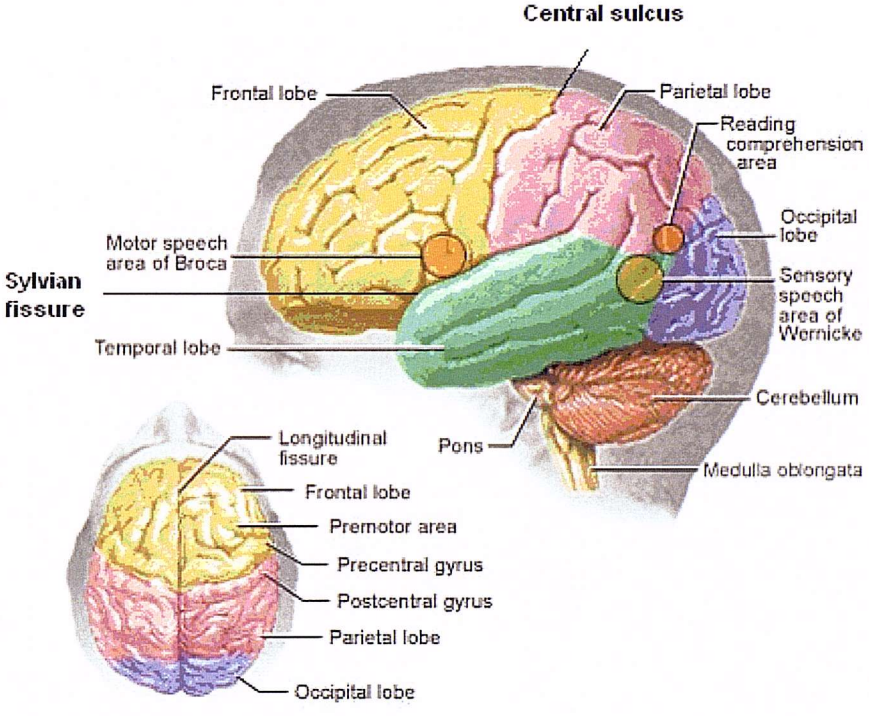


Figure 32. Brain Anatomy

Source: Modified from (HowStuffWorks, Online)

Wrapped up by the four lobes and buried deep within a brain is the limbic system (Figure 33). Language-related areas in the limbic system include the thalamus, cingulate gryus, mamillary body, hippocampus, amigdala, putamen, and caudate nucleus. The thalamus is the main control center for all sensory information (except for smell subserved by a separate entity called 'olfactory bulb') and is connected with almost all cortical areas. The cingulate gyrus regulates alertness, agitation, and various other emotion and cognitive

functions. Further, cingulate gyrus is also associated with language tasks, e.g. distinguishing language ambiguity (Ketteler *et al.*, 2008). The mamillary bodies are implicated in memory; small volume mamillary bodies are correlated with memory loss in Alzheimer's disease patients (Copenhaver *et al.*, 2006). The hippocampus is associated with spatial memory and consolidation of knowledge. The amygdala is associated with detecting stimuli that are posed as potential threats and fear (Phan *et al.*, 2008). The putamen and caudate nucleus are associated with motor control (activity execution and termination) and skill acquisition (Ward, 2006).

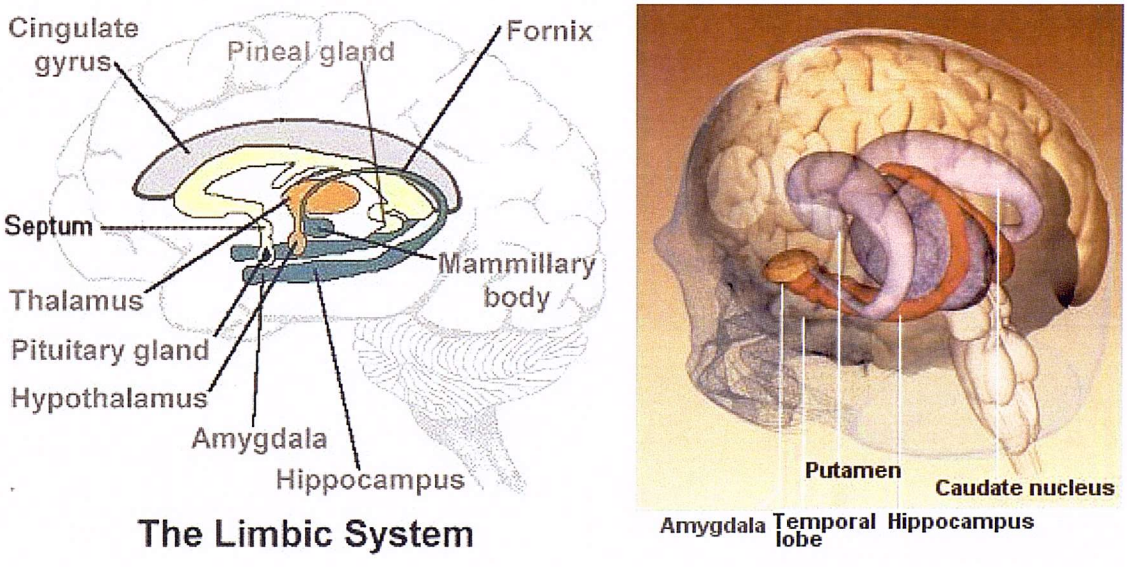


Figure 33. Limbic System

Source: (UNVERSE-REVIEW, Online)

The demarcation lines between lobes are not readily observable, apart from the central sulcus that serves as a dividing line between frontal lobe and parietal lobe as well as the sylvian fissure (Figure 32) that marks the division between frontal lobe and temporal lobe. Further, some cortical areas are not observable on the brain surface, as they are medially located in the brain. Two salient examples are the cingulate cortex that medially spans across the frontal lobe and parietal lobe, as well as the insula (an island-shaped

cortical region) buried underneath and across the temporal lobe; these two areas strongly associate with language processing (Gazzaniga, 2002).

Different lobes are associated with different cognitive functions (for an overview see Diamond *et al.*, 1985; Huettel *et al.*, 2003; Martin, 2003; Ward, 2006; Bear *et al.*, 2007). Figure 34 shows the general cognitive functions subserved by different lobes. The frontal lobe is mostly associated with language production (the Broca's area), planning and execution of body and eye movements, cognition and emotions. The primary motor area (precentral gyrus) in the frontal lobe controls the mechanical actions of movements, whilst the premotor cortex (anterior to the primary motor cortex) is involved in decision-making and planning of motor activities. The prefrontal cortex is related to executive function, decision-making, cognition, emotions and memories (Xiao *et al.*, 2008).

The temporal lobe, situated below the frontal lobe and parietal lobe, has various cognitive functions. The superior temporal gyrus, where the primary auditory cortex is located, subserves perception and recognition of auditory input and memories. The temporal lobe also plays a crucial part for language processing. The Wernicke's area, located in the upper part of the temporal lobe, is involved in language perception and comprehension (Andoh *et al.*, 2008). Adjacent to this area are the cortices often associated with long-term lexical and semantic memory (Venneri *et al.*, 2008).

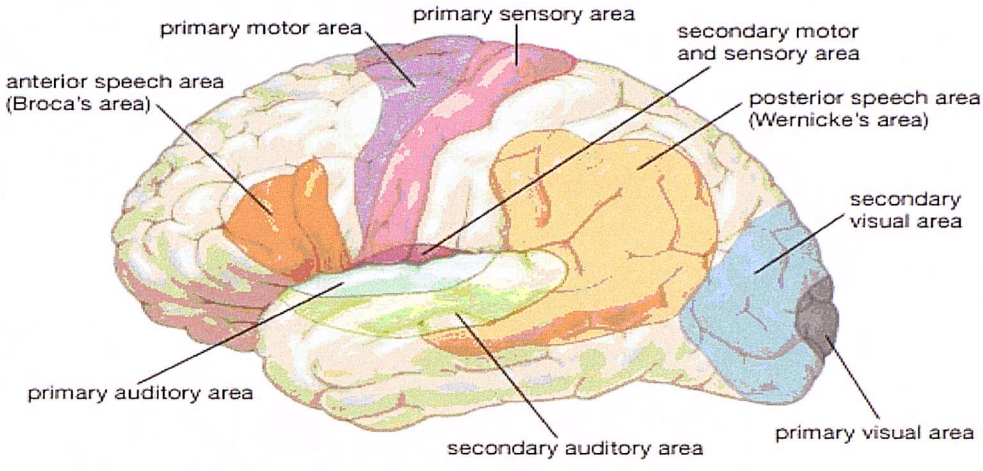


Figure 34. General Cognitive Functions of the Four Lobes of a Brain

Source: (Britannica, Online)

The parietal lobe is posterior to the frontal lobe, and the central sulcus is the observable dividing line between the two lobes. The primary sensory cortex (postcentral gyrus) is located in the parietal lobe and is responsible for perception of bodily and limb sensations, e.g. pain, touch, heat etc. (Ward, 2006). The inferior parietal cortex is often associated with integration of speech-related perception stimuli (Ward, 2006). The superior parietal cortex is associated with mediating behavioral interactions with the outside world, as well as basic arithmetic abilities (Butterworth, 1999).

The occipital lobe is the most posterior lobe of the brain. Primary and secondary visual cortices are located here. This lobe only deals with the processing and perception of visual stimuli (Huettel *et al.*, 2003).

6.4 fMRI Experimental Design

The fMRI study was carried out in the Imaging Sciences Department, Hammersmith Hospital of the Imperial College London. The model of the MRI scanner used was a Philips 3T Intera Scanner (Figure 35).

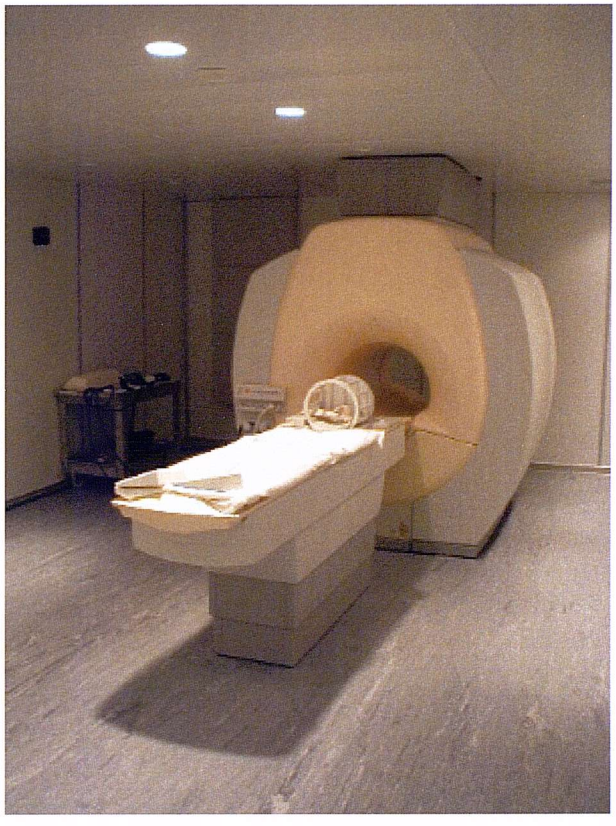


Figure 35. Philips 3T Intera Scanner Used at the Robert Steiner MR Unit at Imaging Sciences Department, Hammersmith Hospital Campus, Imperial College London, UK.

6.4.1 Participants

Five female novice interpreters from Mainland China[19] with the language pair of Mandarin as the first language and English as the second language were recruited from the University of Westminster. All were from Mainland China. The five subjects had successfully finished a one-year postgraduate course in interpreting and translation in the UK and had gained a Master's degree. The study was approved by the Imperial College Research Ethics Committee and conducted with the aims: 'Optimization of a new 3T magnetic resonance imaging (fMRI) scanner and comparison with a 1.5T scanner'. All the participants were rewarded with an honorarium of 20 pounds. Before the participants left

[19] As all the five participants were from Mainland China, simplified Chinese characters were used in the textual stimuli introduced in **Section 6.6.2**.

the laboratory, all of them were also given a printed copy of their own brain images.

6.4.2 Psychological State

Subjects who are emotionally vulnerable can be adversely affected by fMRI experimentation and psychological instability can be a confounding factor in the interpretation of fMRI data. To prevent any complication of this kind, participants completed a questionnaire named the Hospital Anxiety and Depression Scale or HADS (Appendix S) on arrival. This is commonly used by clinicians to determine the psychological suitability of subjects for such studies.

HADS scores were calculated on site and results showed that all subjects were psychologically stable and normal in a hospital setting and therefore suitable for the study (Table 30).

	Participant	Anxiety	Depression
1	1	3	2
2	2	3	4
3	3	7	6
4	4	0	0
5	5	5	1
Total Number of Subjects	5	5	5

Table 30. Anxiety and Depression Scoring for Each Participant (0 - 7 = normal, 8 - 10 = mild, 11 - 14 = moderate, 15 - 21 severe)

6.4.3 Handedness

Results based on the on-line Edinburgh Handedness Test (CSE, Online) revealed that all the participants were right-handed, consistent with left hemispheric dominance for languages (Knecht *et al.*, 2000). Table 31 shows the descriptive statistics of handedness of the subjects. In terms of the scoring, 100 means 100% right-handedness, whereas -100 represents 100% left-handedness.

	N	Minimum	Maximum	Mean	Std. Deviation
Handedness %	5	90.0	100.0	97.00	4.4721
Valid Number of Subjects	5				

Table 31. Handedness Scores

'Minimum' and 'Maximum' refer to the lowest score and the highest scores amongst the participants.

6.4.4 Informed Consent

The participants were asked to read through a Participant Consent Form (Appendix T) and sign it to show that they were willing to participate in the study with informed consent. When asked about the risk of staying in a strong magnetic field for approximately an hour, the investigator explained that there was no evidence to suggest that an MRI scan could cause any adverse effect. The Imagery Safety Check List or 'metal check' was then performed to ensure that the participants were free from any metallic objects inside or outside their bodies, e.g. heart pacemaker, and metallic joint. If a subject was wearing glasses with metallic frames or contact lenses, s/he would be asked to take them off and

replace them with plastic glasses (made up of plastic frames and lenses) already prepared in the laboratory. Further, to exclude the possibility of there being any metallic material in the participants' clothing, all the participants were asked to take off all their clothes (underwear included) and accessories, and to change into specially provided clothing that was completely metal-free. It is known that some make-up consists of a variety of minerals or metallic substances (e.g., Yabe *et al.*, 2001), and, therefore, all of the participants were asked not to put on any make-up when coming to the fMRI laboratory. The metal check was done meticulously as research has shown that ferromagnetic objects or metal during MRI scanning can suffer from over-heating (Shellock and Shellock, 1998), which is dangerous to participants. Moreover, any metallic object, regardless of its size, present in the strong magnetic field generated by an MR scanner will be sucked toward the scanner, which can also be dangerous to participants. Research has shown that tooth fillings are not affected by MRI scanning (Eggers *et al.*, 2005), as they are non-ferromagnetic, e.g. pure silver alloyed with pure gold or titanium. The metal check was conducted with great care under the supervision of Prof. Basant K. Puri. No problems arose and the scanning went smoothly.

All the participants were asked to provide their GPs' or general practitioners' contact information, so the GPs could be notified, should the experimenter find any brain anatomical abnormality during data analysis.

In addition to the fMRI scanning, two of the five subjects underwent DTI (diffusion tensor imaging) and spectroscopy imaging (exploring the concentration of chemical compounds in a particular brain area) in accordance with the full experimental protocol of the study. As these two elements are outside the scope of this doctoral study, they will not be discussed in detail here. However, a brief introduction to their research potential is presented in **Section 7.4.3** in the final chapter.

6.5 fMRI Sparse Paradigm

Hall *et al.* (Hall *et al.*, 1999) originally developed 'sparse' temporal sampling for acoustic fMRI scans with the aim of reducing the effect of the scanner noise; subsequent research has shown that its usefulness is wider than this. fMRI is sensitive to orofacial movement; ensuing studies have discovered the paradigm is useful for experiments that involve orofacial movement, and demonstrated that minimal orofacial movement does not significantly affect accurate brain localization (Heim *et al.*, 2006; Dresel *et al.*, 2005; Gracco *et al.*, 2005). The sparse sampling paradigm, which allows subjects to articulate with minimal orofacial movement and which includes time windows for scanning that is free from a potential effect of orofacial movement (see **Section 6.7.1**), was therefore considered useful in fMRI scanning in the study. Orofacial movement should be avoided during dynamic scans; the sparse paradigm offers a time window where a volume of brain image is being acquired by the MRI scanner with no effect of orofacial movement. Further, imaging studies, regardless of modalities, need behavioral data to validate the neurological data; therefore, sparse sampling provides a useful experimental paradigm where interpreters' behavioral data, i.e. speech output, can be recorded for analysis and for validating the neurological data.

6.6 Experimental Paradigm

As the experimenter's laboratory did not have the facility to present auditory stimuli, visual stimuli were chosen to formulate a naturalistic task that would simulate simultaneous interpreting. Therefore, 'sight translation' (also known as 'sight interpreting') was chosen as an experimental task. Sight translation is a special modality of simultaneous interpreting and it requires verbally translating a written text in a source language into a target language . As the task is usually performed in real time for a conference audience, it should be called more correctly 'sight interpreting' (Pöchhacker, 2004: 19). During sight interpreting, the interpreter has to concurrently read the written text and produce the translation; therefore, its 'simultaneity' is similar to that of utterance-based simultaneous interpreting. It is one of the interpreting techniques that an interpreter has to master. Sight interpreting was

selected as the experimental task, as it requires concurrent visual input in a source language and verbal production/translation in a target language. Therefore, 'Simultaneity of the interpreting' was retained.

6.6.1 Pure Insertion

One of the fundamental concepts behind the design of many fMRI and PET studies is *pure insertion*. To illustrate, if an investigator wishes to explore the neural correlates of semantic processing, s/he may choose to use the word - nonword paradigm. As previously introduced, a nonword is a word designed by a researcher for it to seem/sound like a word and yet it is NOT a word, such as 'lipump' or 'mopol'. In this case, Task A is reading/understanding a word, which requires morphological, phonological, orthographical and semantic processing, whereas Task B is looking at a nonword, which only requires morphological, phonological and orthographical processing without semantic processing, as the nonword contains no meaning. In this way, the two tasks are decomposed into several components, and the component the experimenter wishes to examine, i.e. semantic processing, is *inserted* into Task A. After collecting the imaging data regarding the two tasks, results of word recognition, with the results of nonword recognition *subtracted,* will yield results of neural substrates subserving semantic processing. This is also often called the 'subtraction paradigm'. Table 32 shows the idea behind pure insertion or cognitive subtraction.

Word recognition = (1) phonological, (2) morphological and (3) orthographical processing and 4) semantic processing
MINUS
Nonword recognition = (1) phonological, (2) morphological and (3) orthographical processing
EQUALS
Semantic processing

Table 32. Cognitive Subtraction for Semantic Processing

The above example elucidates the concept of pure insertion or cognitive subtraction. For the study presented here, 'interpreting' was the task in question and 'reading' was selected as a baseline task. Table 33 shows the cognitive subtraction paradigm of the study.

Sight Interpreting = (1) input language reading, (2) input language comprehension, (3) simultaneous interpreting or internal language conversion, (4) output language production and 5) language monitoring
MINUS
Reading = (1) input language reading, (2) input language comprehension, (3) output language production and (4) language monitoring
EQUALS
Simultaneous interpreting or internal language conversion

Table 33. Cognitive Subtraction for Simultaneous Interpreting

6.6.2 *Material*

Grade Level: Two Mandarin short stories were selected and modified from the texts produced by Lin (Online) and Xiao (Online) and used in two individual sessions or blocks where the tasks of 'interpreting' and 'reading' were mixed (see **Section 6.8**). Two English short stories were selected and modified from two children's story books (Carew, 1978; Milne, 1980). Due to the extremely large file sizes of the modified texts used as the stimuli, the texts have been made available online. The two Mandarin and the two English texts in the form of visually presented textual stimuli for the fMRI experiment are all retrievable on the CD included in this thesis. All of the four stories were segmented into short sentences to be visually presented as a .bmp file on the projector screen within the MRI scanner. The Flesch-Kincaid Grade Level test available on Microsoft Office Word 9.0 was used to establish that the grade level of the two English texts was 4, which means that a fourth-grader in the US educational system could read the texts without difficulty.

Comprehensibility: Five non-translator bilinguals[20] with Mandarin as their first language and English as their second were asked to read and rate all the four texts in terms of their comprehensibility level, based on the scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. Table 34 shows that all the four texts were rated at 1.6 or 1.8. This showed that the four stories were fairly easy to comprehend and minimized the potential effect of a text-related confounding factor of 'comprehensibility'.

[20] Due to time and practical constraints, as many as recruitable volunteers were asked to carry out the tests. Therefore, in terms of the number of people volunteering to conduct the comparability tests, the situation presented in this chapter was similar to those reported in **Chapter Three**.

	N	Minimum	Maximum	Mean	Std. Deviation
English Story 1 (Interpreting)	5	1	2	1.60	.548
English Story 2 (Reading)	5	1	2	1.60	.548
Mandarin Story 1 (Interpreting)	5	1	2	1.80	.447
Mandarin Story 2 (Reading)	5	1	2	1.60	.548
Valid Number of Subjects	5				

Table 34. Comprehensibility Test.

Readability: The same five non-interpreter bilinguals were asked to read and rate all the four texts in terms of readability level, based on the scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. Table 35 shows that all the four texts were rated at 1.4 or 1.6. This showed that the four texts were fairly easy to read and minimized the potential effect of the text-related 'readability' variable.

	N	Minimum	Maximum	Mean	Std. Deviation
English Story 1 (Interpreting)	5	1	2	1.40	.548
English Story 2 (Reading)	5	1	2	1.40	.548
Mandarin Story 1 (Interpreting)	5	1	2	1.60	.548
Mandarin Story 2 (Reading)	5	1	2	1.40	.548
Valid Number of Subjects	5				

Table 35. Readability Test

Interpretability: Two professional teachers of interpretation with Mandarin as their first language and English as their second were asked to read and rate all the four texts in terms of interpretability level, based on the scoring system of 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult. This is because the fact that the texts were rated as fairly easy to comprehend and read did not mean that they were easy to interpret; hence, an interpretability test was introduced. Both professional translation teachers gave 1 to all the four texts, and Cohen Kappa coefficient that represents inter-rater agreement was 1. This indicated almost perfect agreement (Landis and Koch, 1977: 167-171) and suggested that the texts were very easy to interpret, eliminating the potential confounding factor of 'interpretability'.

The four texts were segmented and turned into picture files as visual stimuli. Each stimulus carried approximately three and a half lines of words (font size 40, justified). Before deciding on the 3.5-line stimulus format, material testing was carried out on one translator and one interpreter. They were asked to read or interpret the two lines or 3.5 lines of words on the picture file within 7 seconds (the reason behind the 7-second duration will be explained later in section on *Sparse Scanning*). In terms of the two-line stimuli, both of them finished reading and sometimes finished interpreting within the 7-second timeframe, which would run the risk of a declining hemodynamic response (see **Section 6.7.1**) of the brain before the dynamic scan. In terms of the 3.5-line stimuli, none of the participants managed to finish reading or interpreting all of the words, a way to ensure the hemodynamic response would still be at its peak. This is the reason why 3.5 lines were displayed in each stimulus.

All four stories were presented to the participants to read out loud once to the researcher. This was done because interpreters are generally given the texts or at least relevant material to study before embarking on an interpreting assignment in reality. By doing so, the participants would be able to understand the gist of all the four stories to be visually presented later in the experiment. Each scanning block included a story, and there were four blocks with four stories in total. A block could take up to ten minutes to complete, and before the next block began, there was a rest session of two minutes. This created the possibility that a participant would forget what the ensuing stories would be about. Therefore, before each block started, the researcher would remind the participant of the gist of the story via the intercom until the participant verbally verified with the researcher that she knew which story was going to be presented in the upcoming block.

6.7 Experiment Setup on E-Prime

E-Prime (Psychology Software Inc.) is a Windows-based environment for inputting the experimental text stimuli; E-Studio is a component of this software where an experimenter can enter and store all the parameters and details of an experiment.

An important caveat is appropriate for any who wish to replicate this fMRI study in the future. For each trial, a stimulus (either visual or auditory) has to be fed into a particular stimulus grid within the TrialList entry. However, there is a buffer memory limit (800 KB) for each stimulus entry, and if this limit is exceeded, the entire E-Prime system will crash and the experiment design will fail. This information was not included in any E-Prime manual and hence was learned through trial and error. Any future researcher who wishes to carry out such a naturalistic fMRI experiment using texts as stimuli should definitely take this into account. Normally, a single word presented aurally (i.e. as a sound file) requires more buffer memory than a single word presented visually (i.e. an image file), and such a single word approach should not exceed the buffer memory limit pre-set in E-Prime.

As previously stated, the stimulus used for each trial in this study had an average of 3.5 lines of words. Initially it was planned to use aural stimuli for this series of experiments. Considerable effort was invested in getting the four stories recorded by native Chinese and English speakers and then editing the audio-files to produce the required series of stimuli. However, these files proved too large to be handled by E-Prime, and despite many attempts to compress them, the idea of aural stimuli had to be abandoned. After consultations with Dr. Alex Dresner[21], 'sight interpreting' was selected as the experimental task instead. It was found that image files of the stimuli texts conformed satisfactorily to the memory limits imposed by E-Prime and 'simultaneity' of interpreting was retained.

All the stories were later edited, split into bitmap files and fed into E-Prime. On average, each visual stimulus in a bitmap file was less than 200 KB, and this fitted nicely into the buffer memory limit for each stimulus. By doing so, subsequent test runs were all successful and 'simultaneity' of interpreting was once again retained.

The instruction (**READ** or **INTERPRET**) was on top of the file and the 3.5 lines of words were displayed underneath the instruction. Each text was divided into sixty bitmap

[21] Dr. Alex Dresner is an MRI physicist. His responsibilities are maintaining the MRI machine, as well as assisting setting up MRI parameters for experiments at the Department of Imaging Sciences, Imperial College London.

image stimuli for the fMRI scanning.

To reiterate, before the scanning, the four texts were given to the subjects to 'read just once'. The reasons were as follows:

- To replicate real-life situations. Sight interpreting is conducted this way, i.e. the text to be interpreted is given to interpreters in advance for them to read.
- To avoid memory loss of the story line.

6.7.1 Sparse Scanning

This fMRI study used two tasks, i.e. READ and INTERPRET, where READ was the baseline. After subtraction (INTERPRET - READ), light would be shed on the different cortical areas subserving 'interpreting' in different directions.

Based on the sparse paradigm (Hall *et al.*, 1999), the fMRI experimental design regarding sight interpreting was as follows:

For each volume acquisition (Figure 36), one single visual stimulus lasted seven seconds (while the subject was concurrently verbally conducting on-line reading or interpreting) to ensure the hemodynamic response or BOLD (blood oxygenated level dependent) signal would peak or plateau. This duration was chosen because Pu *et al.* (Pu *et al.*, 2001) demonstrated that the average peak of hemodynamic response for both Chinese and English oral verb-generation was nearly eight seconds. Immediately after this, there was a three-second dynamic scan during which the subject stopped speaking (to avoid orofacial movement) and a volume of brain images were obtained in an ascending order, i.e. from the neck to the crown of the head.

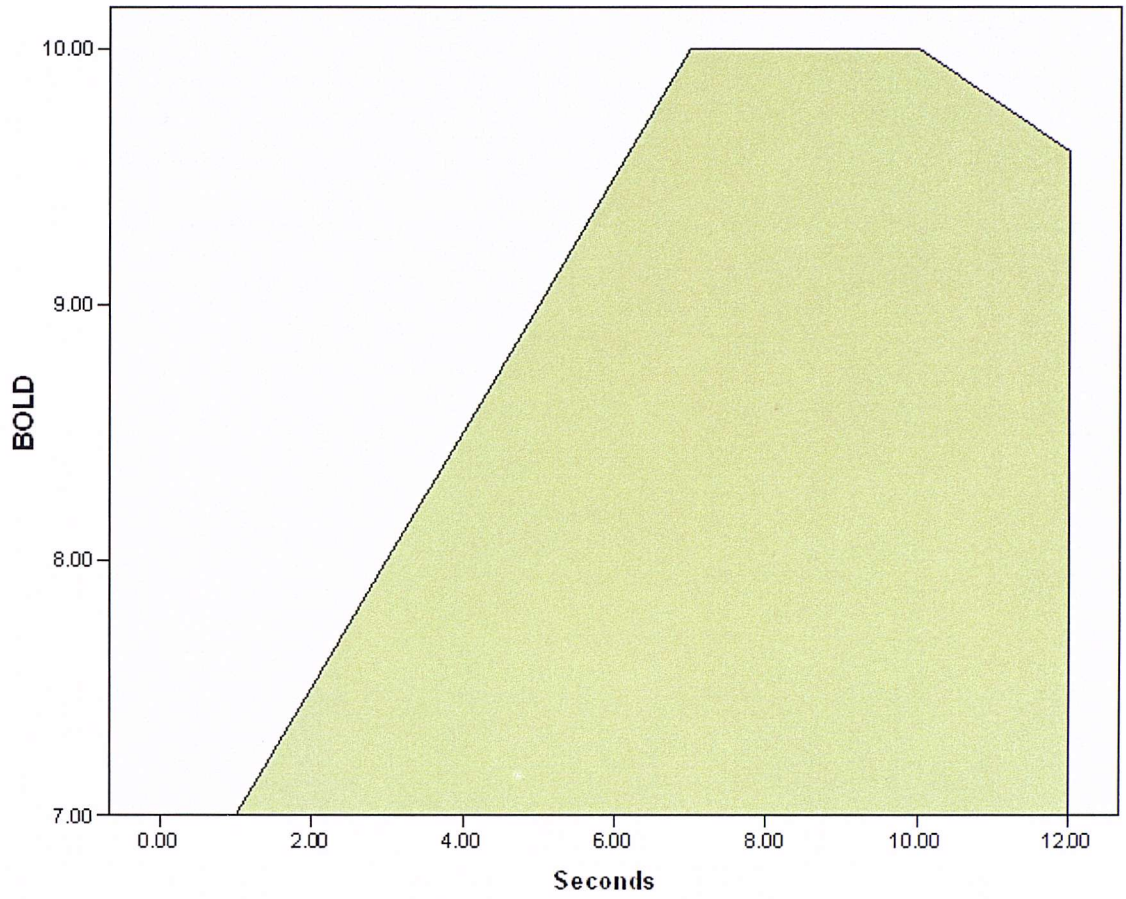


Figure 36. One Volume Acquisition

The unit of the x axis is seconds, whereas the numbers along the y axis represent the incremental strength of the BOLD signal (the ratio between deoxyhaemoglobins and oxyhaemoglobins).

As one acquisition required ten seconds and each text had sixty stimuli, a whole sparse sampling block lasted ten minutes and sixty brain volumes were obtained. The four texts produced four sparse blocks, so the total sparse sampling time was forty minutes, not including the time needed for surveys, dummy scans and structural scans. Two hundred and forty volumes were acquired in all: sixty volumes for English-into-English READ; sixty volumes for English-into-Chinese INTERPRET; sixty volumes for Chinese-into-Chinese READ; and sixty volumes for Chinese-into-English INTERPRET.

6.8 Trial Arrangement

In this study, there were two conditions: READ and INTERPRET. During each of the four blocks, trials of different conditions were included: (1) sight interpreting from English into Chinese (SIEC); (2) sight interpreting from Chinese into English (SICE); (3) reading Chinese (REC); (4) and reading English (REE). Within each block, trials of conditions with the same source language were mixed. For instance, during the first block, trials of SIEC were mixed with those of REE in such a way that six trials of SIEC were followed by six trials of REE, followed by another six trials of SIEC, followed by another six trials of REE, and so on. By the same token, in the second block, six trials of SICE were followed by six trials of REC, followed by another six trials of SICE, followed by another six trials of REC and so on. A similar sequence was applied to the other sparse sampling blocks.

Before and after the experiment, there was a structural scan that lasted for five minutes. In between two successive blocks, as already stated, there was a two-minute rest period where no scans were involved. This was mainly to allow the subject to rest and prepare for the upcoming scanning, and also to enable the investigator to set up the experimental stimuli for the next block. However, one subject was allowed to rest for ten minutes at her request.

6.9 MRI Scanner and Exam Card

In the study, 240 T_2 functional volumes (each consisting of thirty-two slices acquired) were acquired from each participant's scanning. T_1 anatomical volumes (150 slices) were also acquired for each participant. Slices to form a full brain volume are acquired in a certain order. In this study, pre-set parameters dictated that all slices were acquired in an ascending order, which means that slices were obtained in a bottom-up order, with the first slice acquired at the very bottom of the brain and the very last slice obtained at the very top of the brain.

In order for the MRI scanner to synchronize the displaying of stimuli with the scanning, a sparse exam card that included all the parameters for an experiment was set up on the MRI computer console.

6.10 Piloting Using Phantom Scanning

Once the sparse paradigm had been set up, the scanning process was piloted. First, all the stimuli were successfully synchronized with the MRI scanner control console using the corresponding exam card. Then, a phantom scan was successfully carried out. This is a standard procedure prior to an fMRI experiment. A non-ferromagnetic object representing a human being is scanned to check the entire scanning process; in the study, a cylinder-shaped plastic container full of water (Figure 37) was used for phantom scanning. The phantom was placed within the coil cage of the scanning table, as if the phantom were the brain to be scanned. The researcher fed the MRI sliding table into the scanner, and started a structural scan. The first phantom scan was a success, and a clear structural image of the phantom was generated. Then, a second phantom scan was also successfully carried out. This enabled the researcher to move on to human piloting.

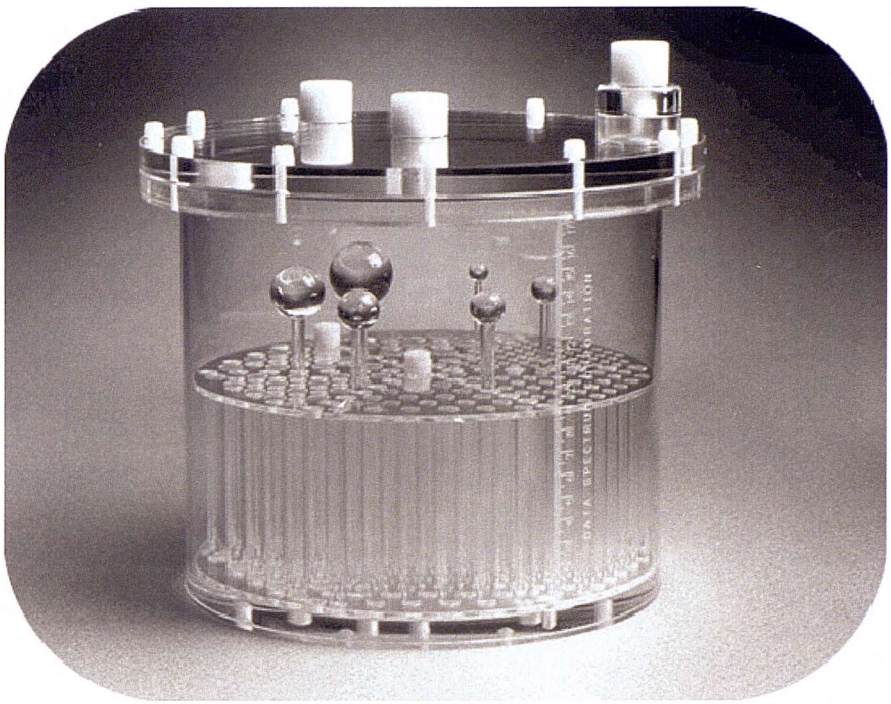


Figure 37. A Phantom Used for Phantom Scanning

Source: (DSC, Online)

6.11 Human Piloting

After phantom scanning, one non-interpreter bilingual male with the language combination of Mandarin and English volunteered for a human pilot run of a scanning session, in order to ensure that the entire experimental flow was free from any problems. Even though this was merely a pilot run on a non-interpreter, all ethical procedures were followed based on the ICREC ethical approval. With the successful completion of this human pilot, the scanning of the actual experimental subjects (i.e. novice interpreters) could begin.

6.12 Pre-scanning Procedure

All details of the experimental procedure had to be followed meticulously to ensure that (1) the full ethical protocol was followed; (2) participants' safety was maintained; and (3) the data collection would go smoothly. Any neglected detail in the process could have caused the entire experiment to fail. It is therefore important to report the detailed experimental procedure.

Upon arrival, to relieve potential anxiety generated in a hospital setting, the investigator would try and initiate a general conversation for a few minutes. Afterwards, preliminary details regarding the experiment were explained, still in a conversational setting. During the conversation, it was emphasized by the investigator that the participant could ask any question at any point in time before the experiment officially started and that she was allowed to withdraw from the experiment at any point in time without having to bear any consequences, should she feel uncomfortable or reluctant to go on.

Once the participant was led into the MRI laboratory, she was advised to put all of her personal belongings in a plastic bag already prepared in the laboratory, and it was the investigator's responsibility to take care of them. Afterwards, the documentation required by the ethical approval was presented to the participant. She was allowed as much time as she needed to read them, while the investigator sat next to her to explain any points unclear to her in the documentation.

Some of the issues could be potentially sensitive to the female participants in the study, so they were asked if they would prefer a female assistant to be present to discuss the issues. Two salient examples are as follows. First, during the process of metal check, the investigator has to ask the participant if she is using an IUCD (intrauterine contraceptive device); if she is, she would not be suitable for the study, due to the metallic property of the device. Second, also during the metal check, the investigator had to ask the participant to remove her bra for the scanning, especially if she was wearing a bra with metallic breast support. However, all of the five female participants were comfortable with discussing all the potentially sensitive issues with the investigator, and therefore the help of a female

assistant was not required.

After they had thoroughly read, fully understood and signed all documentation furnished, they were asked to change into patients' clothing in a changing room. As they also had to take off their shoes and socks, slippers made of paper were prepared for them to walk into the scanner room. Their personal clothes were kept in another plastic bag to be placed next to their personal belongings.

As the scanning must not be interrupted and the participants had to stay in the scanner for approximately an hour once the experiment started, they were encouraged to use the restroom before they were led into the scanner room. Before entering the scanner room, a small non-metallic object was taped onto the left temple of the subject, so the investigator would be able to correctly identify the left and right hemispheres during the stage of data analysis.

The participant was then led into the scanner room. The participant was asked to lie down on the scanner's sliding table which would later be used to transport the participant into the scanner. As the scanning procedure generates an uncomfortable amount of noise, the participant was asked to wear an earplug in each ear, followed by a pair of ear-defenders. A pyramid-shaped pillow was used as a leg rest to minimize body movement. The participant was instructed to rest her legs parallel to each other, rather than crossing them, to prevent a magnetic loop that could disturb the field inside the scanner. For the same reason, the participant was also instructed to rest her hands on either side of her body. During this process, because the participant's ears were covered, an intercom system facilitated communication. An alarm bell was then given to the participant in case she wished to communicate with the investigator or withdraw from the experiment during the scanning procedure.

Once the participant was securely and comfortably lying on the scanner table, the cage that carried a projector and mirror was moved slowly to enclose the participant's head. The investigator repeatedly reassured the participant, confirming that she was comfortable and could see a reflected image on the mirror used to present stimuli (Figure 38).

As the scanner room was well air-conditioned and the participant was only wearing patients' clothing that was very thin, a blanket was offered to cover her lower body in case she felt cold during the experiment. For the sake of cleanliness, the patient clothing and the blanket were used for one subject only and then replaced.

After all the above procedures were completed, the participant was then moved slowly on the table into the scanner, ready for scanning.

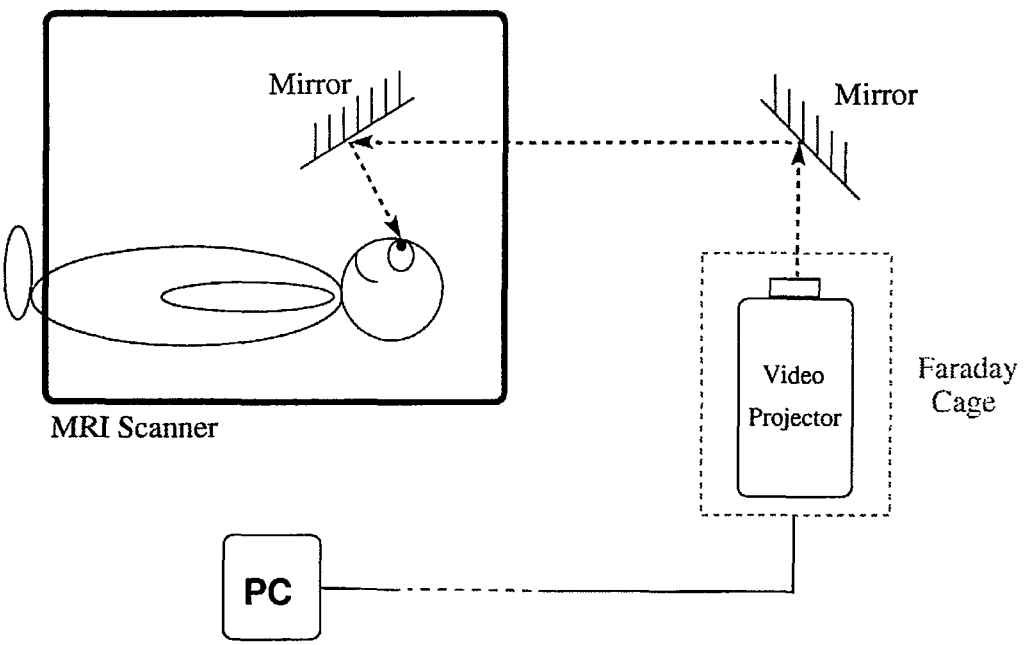


Figure 38. Visual Stimuli Presentation in the Scanner

Source: (Wotawa, 2006: 69)

6.13 Data Storage

All the scanning data were transported to and stored in the UNITE[22] database at Hammersmith Hospital, Imperial College London, in the form of both DICOM (Digital Imaging and Communications in Medicine) files (with participant and scanning information), as well as ANALYZE files that consist of header (.hdr) files accompanied by image (.img) files, and both .hdr and .img files were anonymized. They were password-protected and only accessible to authorized researchers. To protect against data loss, all the data were backed up on an external disc for exclusive storage in the Hospital.

For data storage outside the premises of Hammersmith Hospital, the investigator was allowed to back up only .hdr and .img files (anonymized), as opposed to DICOM files (allowed for use only within the Imaging Sciences Department), on an external drive to facilitate data analysis outside the Hospital.

6.14 Data Analysis

6.14.1 Behavior Results

The behavioral performances (i.e. the outputs) of the subjects during reading and sight interpreting were monitored and recorded through the intercom to ensure that they were conducting the task as instructed. In addition, two Mandarin language/interpreting teachers with English as their second language graded the overall interpreting and reading outputs, based on their standard scoring scheme (100=full mark or 100% accuracy; 60=pass or 60% accuracy). Table 36 shows the scores they awarded.

The behavioral results show that the subjects did conduct the tasks as instructed by the investigator, thus validating the neurological data for further analyses. Interpreting into

[22] The UNITE database is a password-protected database storing all MRI-related data. It is accessible only to faculty members and researchers at the Imaging Sciences Department, Hammersmith Hospital, Imperial College London, UK. The word UNITE is not an abbreviation.

English scored lower than into Mandarin Chinese. This suggests that A→B interpreting was more cognitively demanding than B→A interpreting. The A→B interpreting output was clearly of lower accuracy, as is evident from the grades; the question is whether this is a necessary result of it being a more difficult task, which is quite likely. However, it has to be acknowledged that another arguable interpretation is presumably that an interpreting performance could be of lower accuracy, independently of the cognitive loading.

	N	Minimum	Maximum	Mean	Std. Deviation
Mandarin Story 1 (Reading) %	5	80	92	89.0	5.099
English Story 2 (Reading) %	5	77	90	86.8	5.440
Mandarin Story 1 (A→B Interpreting) %	5	65	85	77.8	9.576
English Story 2 (B→A Interpreting) %	5	69	87	79.8	9.418
Valid Number of Subjects	5				

Table 36. Descriptive Statistics on Scorings

6.14.2 Neurological Results

The investigator used the anonymous .hdr and .img files for data analysis outside the Hospital. Statistical Parametric Mapping 5 or SPM5 (Department of Cognitive Neurology, University College London, UK; http://www.fil.ion.ucl.ac.uk), implemented on MatLab v.R2007a.Student (Mathworks Inc. Sherborn, Mass., USA; http://www.mathworks.com/) was used for data analysis. Figure 39 shows the MatLab interface for SPM5 implementation. Brain activations were explored, following the steps described below. A table for recording all the steps in the analysis was created to ensure that all the procedures were followed without fail.

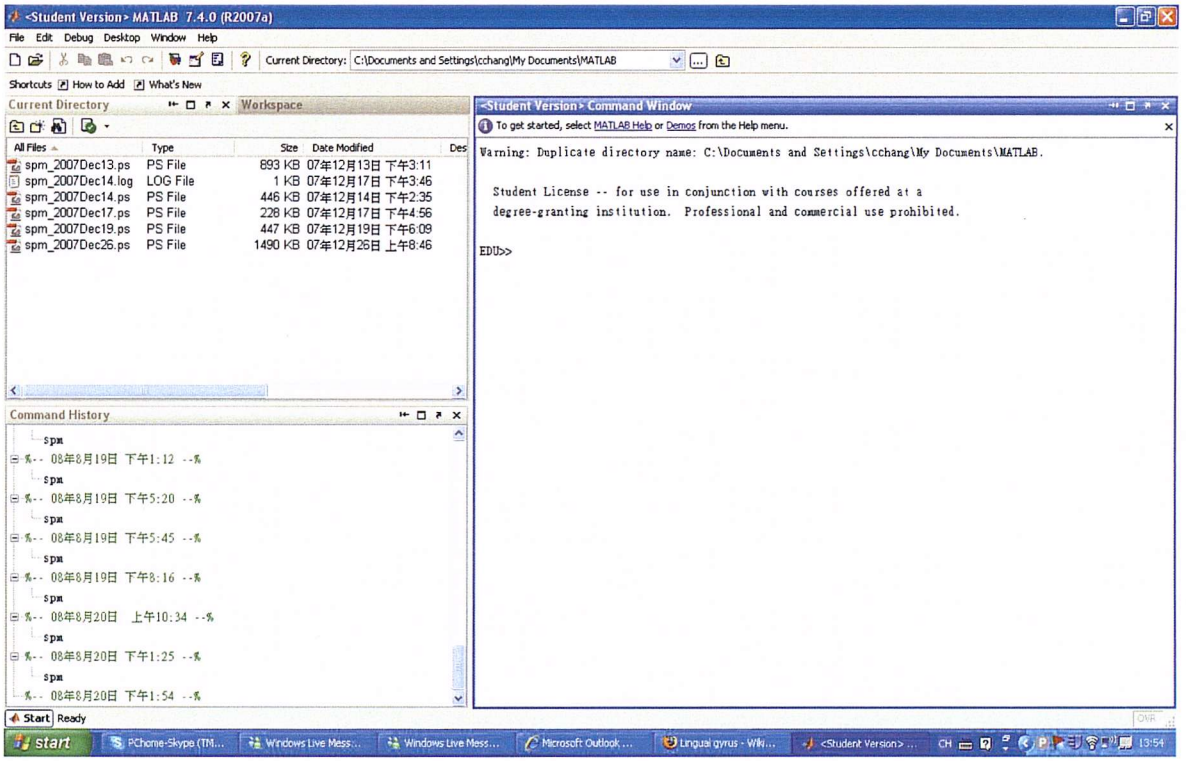


Figure 39. MatLab Environment

As explained earlier, there are two kinds of images acquired from an fMRI experiment: high-resolution, static structural brain images and functional, dynamic planar slice images that make up a brain volume. The analysis of these images using SPM5 was

conducted partially with expert assistance from the Laboratory of Integrated Brain Research, Department of Research and Education, Taipei Veteran General Hospital, Taipei, Taiwan[23].

6.15 Steps Involved in Analyzing the Neurological Data

Analyzing the neurological data involved seven major steps in which many parameters relating to the experimental design were involved. The SPM5 manual (Ashburner *et al.*, 2006; for a detailed discussion of SPM5 data analysis in an audiovisual DVD format, see SPM, 2006; for a detailed introduction on fMRI experimental designs and data analysis, see Aguirre, Online) was used to ensure the parameters and the steps were followed thoroughly.

6.15.1 Slice Timing

As explained earlier, a complete functional brain volume is made up of planar brain image slices which are composed of voxels (three-dimensional image units). These slices are acquired 'sequentially', rather than 'simultaneously'. The 'slice timing' functionality on SPM5 is designed to take into consideration the sequential relationship between slices. SPM5 computes the data regarding the slices forming one brain volume, in such a way that all slices will be treated as a single, simultaneous 'snapshot' of the brain volume.

6.15.2 Realignment

Movement on the x axis is called 'pitch'; movement on the y axis is called 'roll'; and movement on the z axis is called 'yaw'.

Potential head motion on the x, y or z axis, however slight, can cause problems in an fMRI study, possibly causing erroneous localization of the brain with consequent

[23] Contact information: Laboratory of Integrated Brain Research, Department of Medical Research and Education, Taipei Veterans General Hospital, No. 201, Sec. 2, Shih-Pai Road, Taipei, 112, Taiwan. Tel: + 886 (0) 2 2875 7480, Fax: + 886 (0) 2 2874 5182, Email: ibri@vghtpe.gov.tw.

misalignment of the sequentially arranged slice images. No matter how hard a subject tries not to move, a certain degree of head motion, whether conscious or unconscious, is likely to occur. To solve this problem, SPM5 has a functionality called 'realign' which compensates for any head motion on the part of a participant, ensuring that all image slices are motion-corrected and eliminating potential artefacts generated by head motion. During this step in SPM5, the experimenter has to select a slice image to which all other slices will be realigned, usually the first slice.

The smaller the amount of movement on any of the axes, the better. It is recommended that the movement should be within the range between 2 to -2 degrees on any axis (Huettel *et al.*, 2003: 264). Otherwise, data may have to be discarded from analysis. In the fMRI study presented here, the realignment results were within the recommended range.

6.15.3 Coregistration (Coregistering Functional Images with Structural Images)

In this step, all the planar slice images that have been time- and motion-corrected are aligned with the structural image of the participant. This step of 'coregistration' involves the identification of image boundaries on the planar image slices that can be matched with features in the structural images, and the two images thus mapped onto one another. Figure 40 shows how a slice of planar image is mapped onto the structural image during the process of coregistration.

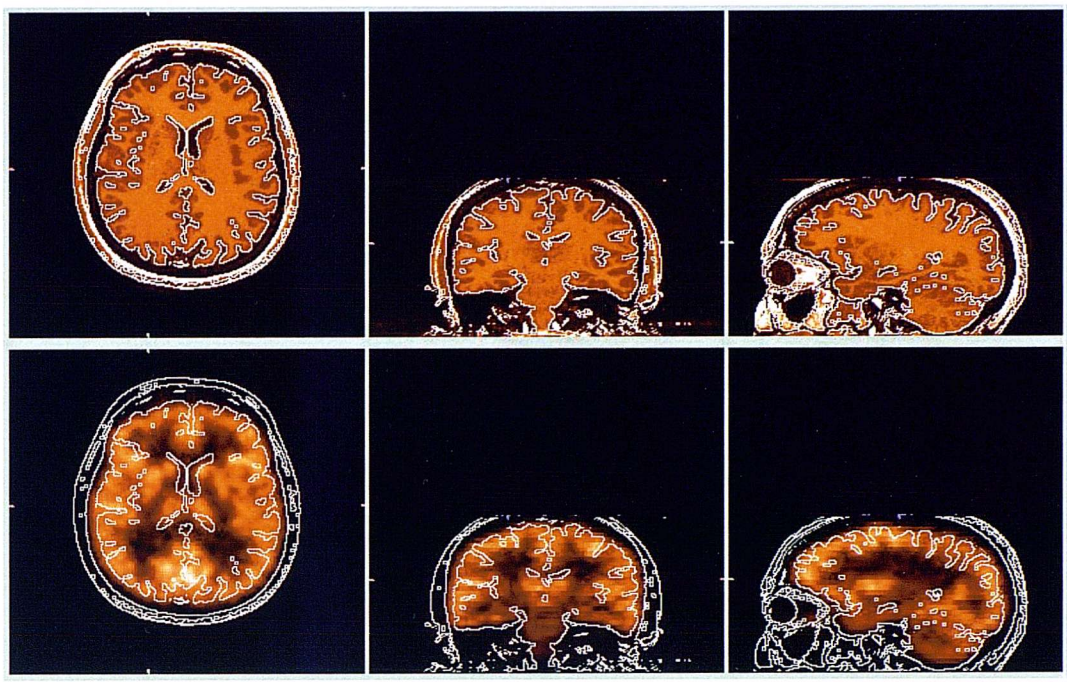


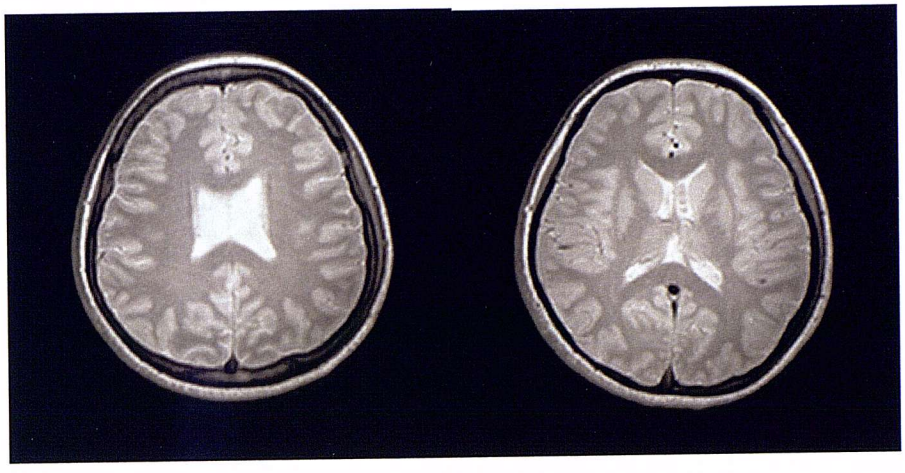
Figure 40. Coregistration

Source: (WAISMAN, online)

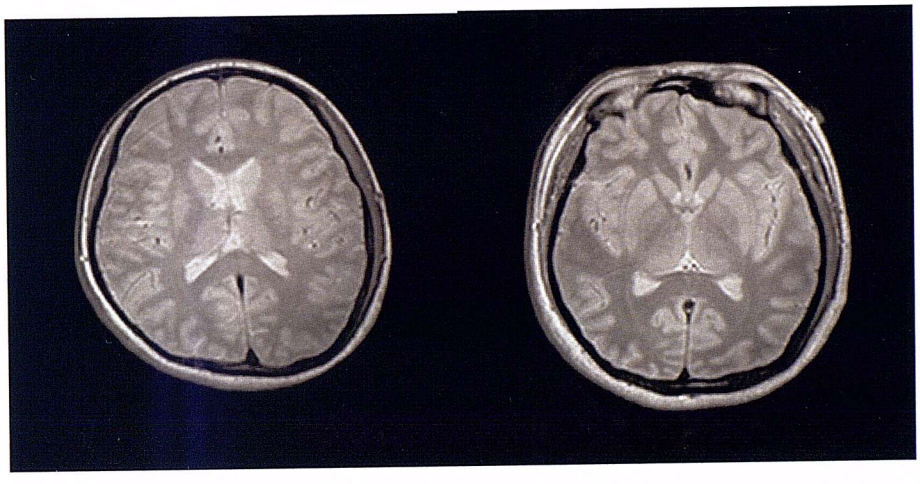
In the figure, red images are the structural images and the images with white-colored boundaries are planar images. On the bottom, the two types of images are not coregistered yet, whereas, on the top, the two types of images have been matched and coregistered.

6.15.4 Normalization

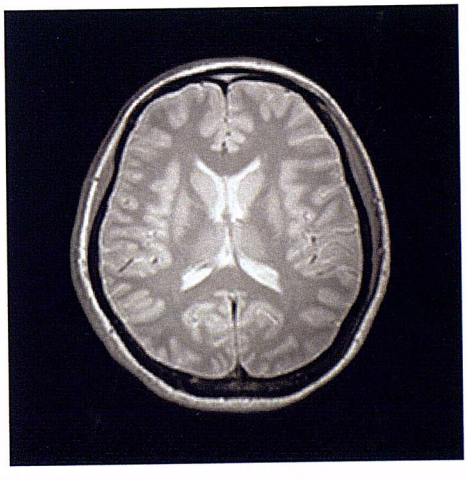
Healthy human brains have different sizes. Figure 41 shows how the brain sizes of the five subjects in the fMRI study differed from one another.



Subject 1 Subject2



Subject 3 Subject 4



Subject 5

Figure 41. Variations of Brain Sizes of the Five Subjects in the fMRI Study

For accurate brain localization, brains of different sizes have to be matched and scaled to a template available on SPM5, which is a process called 'normalization'. Huettel *et al.* (2003: 196) defines normalization as '... the transformation of MRI data from an individual subject to match the spatial properties of a standardized image'. In this study, a template called 'avgis2T2.nii' was used as a standardized image. This template is known as a Talairach space (Talairach and Tournoux, 1988). Huettel *et al.* (2003: 273) explains, 'Talairach space was created by the French physician Jean Talairach and colleagues and is based upon a simple stereotaxic framework derived from measurements on a single brain'. Further, a stereotaxic space is '[a] precise mapping system (e.g. of the brain) using three-dimensional coordinates' (ibid.: 273). During this step, all participants' brain images (both structural and functional) were scaled, based on the Talairach template size (avgis2T2.nii), so that the brains' sizes were comparable across all subjects. This is also to help with precise localization of brain activities in later analysis. Figure 42 shows a stereotaxic space on the left and a Talairach space on the right.

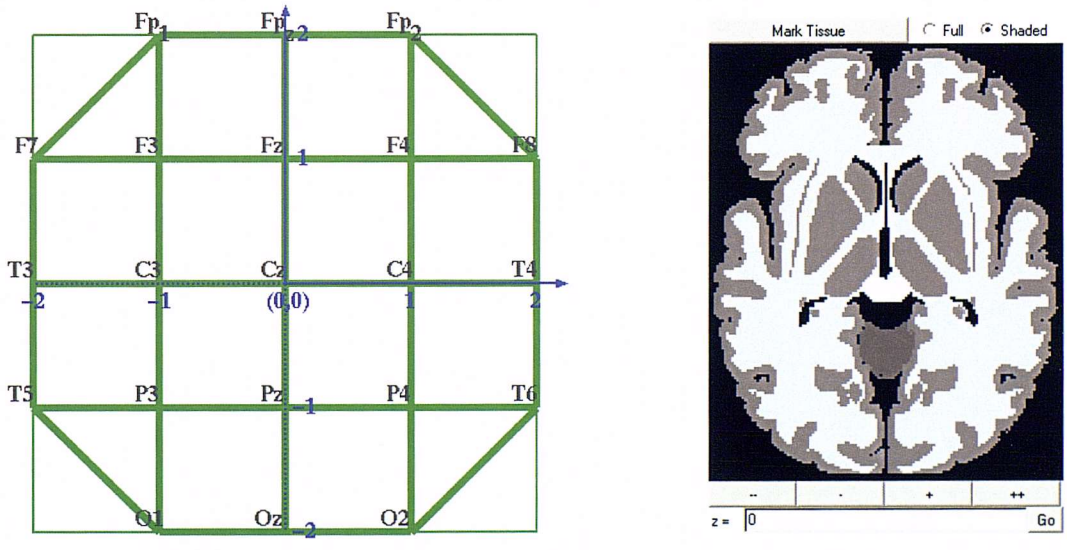


Figure 42. Stereotaxic Space on the Left and a Talairach Space on the Right

Source: (Steinsträter, online)

6.15.5 Smoothing

A voxel is a unit of space, and planar images consist of activation *within* a voxel. 'Smoothing' is simply a way of eliminating the artificial sharp edges that are generated by the punctuation of the voxel grid.

6.15.6 First-Level Analysis

First-level analysis is an 'intra-subject' analysis, based on the post-subtraction contrast between two fMRI images regarding the two tasks. Within 'each' participant, an image 'contrast' was generated. In this study, brain activities accompanying 'interpreting' were in question, and the cognitive subtraction paradigm used was [sight interpreting - reading]. This means that images regarding both 'sight interpreting' as well as 'reading' were yielded. Subtracting the two types of image created the 'contrasts' *within* each participant required for the next step of analysis that would then generate results of brain activities accompanying 'interpreting' *across* the participants.

6.15.7 Second-Level Analysis

Second-level analysis is an inter-subject analysis (called conjunction analysis) (for a detailed introduction on conjunction analysis, see Price and Friston, 1997); conjunction analysis 'combines' the features of interest identified for each individual subject in the first-level analysis across the whole sample. In this way, generalized brain activation patterns associated with the cognitive task can be generated. In essence, this is the final step to show the end results of brain activation patterns of a cognitive task across subjects, based on the multiple intra-subject files for each individual generated during the first-level analysis. After the second-level analysis, a file is produced to help generate the ultimate results.

6.16 Results

The first five steps listed above (namely, slice timing, realignment, coregistration, normalization and smoothing) are together called 'reprocessing'; they were performed to make the brain volumes obtained in the study comparable in size and shape. Then, the first-level analysis and second-level analysis yielded results, based on image contrasts. As previously indicated in **Section 6.6.1**, the images of *interpreting* MINUS those of *reading* generated image contrasts across all the participants, and the contrasts represented the cortical results of the fMRI study.

6.16.1 Interpreting from English into Mandarin Chinese vs. Interpreting from Mandarin Chinese into English

Table 37 shows the coordinates of gray matter areas activated across subjects during sight interpreting from English into Mandarin Chinese. Table 38 shows the coordinates of gray matter areas activated across subjects during sight interpreting from Mandarin Chinese into English.

Statistics: ***p-values adjusted for search volume***

set-level		cluster-level			voxel-level					mm mm mm		
p	c	$p_{corrected}$	k_E	$p_{uncorrected}$	$p_{FWE-corr}$	$p_{FDR-corr}$	T	$(Z_{\equiv})$	$p_{uncorrected}$			
0.952	6	0.111	63	0.000	0.344	0.344	34.90	4.61	0.000	24	-14	-14
					1.000	0.993	14.42	3.82	0.000	24	-24	-14
		0.997	25	0.011	1.000	0.506	25.21	4.33	0.000	-46	-70	30
					1.000	0.993	6.53	2.99	0.001	-42	-68	38
		0.993	26	0.009	1.000	0.993	11.32	3.58	0.000	14	10	-8
		0.579	42	0.002	1.000	0.993	9.31	3.37	0.000	-46	20	20
					1.000	0.993	8.90	3.33	0.000	-40	14	22
		0.836	35	0.003	1.000	0.993	9.22	3.36	0.000	-26	-14	-24
					1.000	0.993	6.26	2.94	0.002	-14	-8	-28
		0.971	29	0.007	1.000	0.993	8.07	3.22	0.001	-2	22	2

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 3.75, p = 0.010 (1.000) {p<0.01 (unc.)}
Extent threshold: k = 20 voxels, p = 0.020 (1.000)
Expected voxels per cluster, <k> = 3.434
Expected number of clusters, <c> = 10.60
Expected false discovery rate, <= 0.99

Degrees of freedom = [1.0, 4.0]
FWHM = 7.4 7.2 6.0 mm mm mm; 3.7 3.6 3.0 {voxels};
Volume: 1368792; 171099 voxels; 3977.0 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 39.95 voxels)

Table 37. Coordinates of Gray Matter Areas Activated during Interpreting from English into Mandarin Chinese

Statistics: ***p-values adjusted for search volume***

set-level		cluster-level			voxel-level					mm mm mm		
p	c	$p_{corrected}$	k_E	$p_{uncorrected}$	$p_{FWE-corr}$	$p_{FDR-corr}$	T	$(Z_{\equiv})$	$p_{uncorrected}$			
0.491	11	0.005	103	0.000	1.000	0.828	19.76	4.12	0.000	38	6	-6
					1.000	0.828	15.26	3.87	0.000	40	-10	0
					1.000	0.828	10.50	3.50	0.000	34	-14	8
		0.012	90	0.000	1.000	0.828	19.13	4.09	0.000	2	-24	32
					1.000	0.828	6.45	2.97	0.001	-20	-30	32
					1.000	0.828	4.30	2.49	0.006	-12	-32	30
		0.940	31	0.005	1.000	0.828	13.91	3.78	0.000	-20	8	40
					1.000	0.828	10.50	3.50	0.000	-16	2	46
		0.073	68	0.000	1.000	0.828	13.80	3.77	0.000	26	-56	0
					1.000	0.828	8.83	3.32	0.000	24	-44	-2
		0.000	173	0.000	1.000	0.828	13.56	3.76	0.000	-40	-44	18
					1.000	0.828	12.34	3.66	0.000	-54	-40	8
					1.000	0.828	8.94	3.33	0.000	-36	-34	12
		0.997	25	0.011	1.000	0.828	13.46	3.75	0.000	-30	2	-18
					1.000	0.828	4.77	2.62	0.004	-34	12	-20
		1.000	21	0.018	1.000	0.828	11.43	3.59	0.000	24	-32	30
		0.217	55	0.000	1.000	0.828	11.13	3.56	0.000	-38	-16	56
					1.000	0.828	7.00	3.06	0.001	-34	-18	48
		0.507	44	0.001	1.000	0.828	10.61	3.51	0.000	-36	-44	-24
					1.000	0.828	6.01	2.89	0.002	-30	-48	-18
		0.169	58	0.000	1.000	0.828	9.97	3.45	0.000	-14	-46	-6
					1.000	0.828	9.22	3.36	0.000	-18	-52	-12
					1.000	0.828	4.20	2.47	0.007	-22	-48	-4
		0.730	38	0.002	1.000	0.828	9.88	3.44	0.000	56	-16	24

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 3.75, p = 0.010 (1.000)(p<0.01 (unc.))
Extent threshold: k = 20 voxels, p = 0.020 (1.000)
Expected voxels per cluster, <k> = 3.434
Expected number of clusters, <c> = 10.60
Expected false discovery rate, <= 0.83

Degrees of freedom = [1.0, 4.0]
FWHM = 7.4 7.2 6.0 mm mm mm; 3.7 3.6 3.0 (voxels)
Volume: 1368792; 171099 voxels; 3977.0 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 39.95 voxels)

Table 38. Coordinates of Gray Matter Areas Activated during Interpreting from Mandarin Chinese into English

In these tables, the column entitled 'cluster level' indicates p values of statistically significant activations. The column entitled 'mm mm mm' refers to the coordinates of the significant activations. For Mandarin → English interpreting, there were a total of twenty-four coordinates, whereas there were only ten for English → Mandarin interpreting. However, the total number of the coordinates did not equal to the number of brain areas activated, as some coordinates referred to the same area.

Once the coordinates were revealed, a free on-line resource called 'Münster T2T-Converter' (Steinsträter, Online) that offers brain localization using coordinates was used to demonstrate the activation of various gray matter areas for the two tasks. Some of the coordinates pointed to the same cortical area. Results presented later will show that eight areas of the brain were activated during English → Mandarin interpreting, and twelve for

Mandarin → English interpreting.

Based on the tables presented above, more cortical areas were recruited to carry out interpreting from Mandarin Chinese into English (A→B), compared with interpreting from English into Mandarin Chinese (B→A). This suggested that the predictions made by the Revised Hierarchical Model were valid, even though the number of the participants involved was limited.

6.16.2 Brain Activation Patterns during Interpreting from English into Mandarin Chinese

A gross three-dimensional image of brain activation patterns for this interpreting direction was produced by SPM5 (Figure 43). The image only shows the gray matter 'on the brain surface'; later figures will display the localization of gray matter in the 'limbic system'.

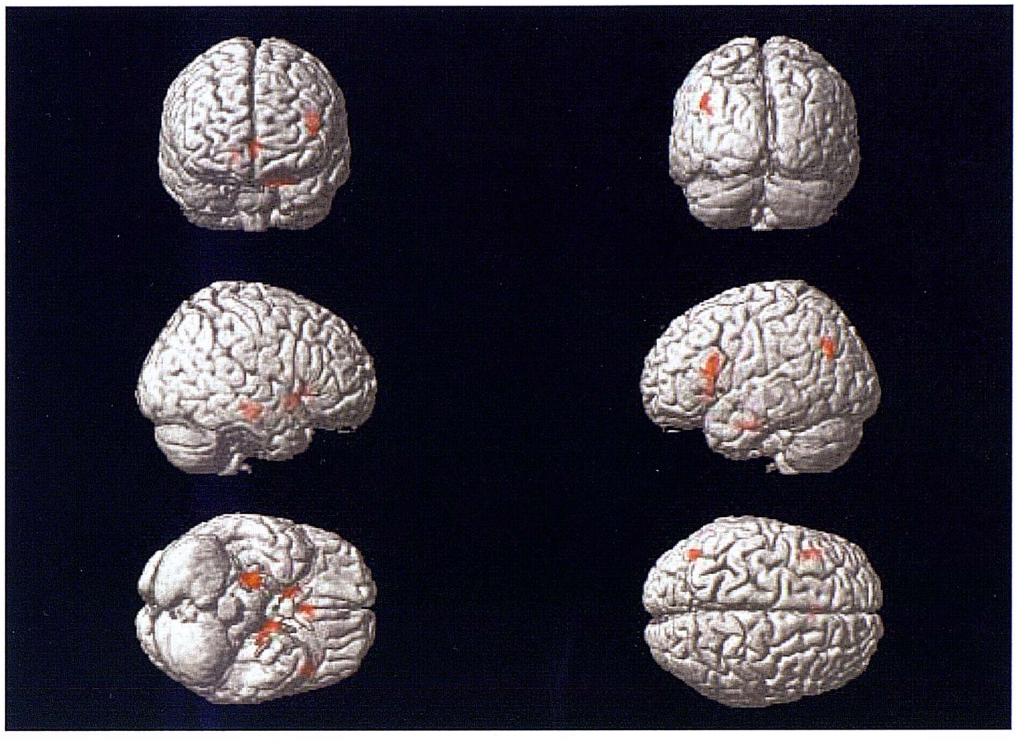


Figure 43. Brain Activation Patterns during Interpreting from English into Mandarin Chinese

In the figure, cortical areas colored in red represent brain activations during the task in question. On the left side of the figure, the uppermost image shows frontal activation; the middle one shows right-hemispheric activation; and the lowermost one reveals activation on the bottom of the brain. On the right side of the figure, the uppermost image shows activation in the back of the brain; the middle one indicates left-hemispheric activation; and the lowermost one shows cortical activation on the very top of the brain.

For each activated brain area for the task, three slice images (coronal, sagittal, and horizontal) of the activation will be displayed. A pictorial example is shown in Figure 44. All the images regarding specific cortical activations presented below are the averaged image results across the five subjects involved in the study. Further, the results reported in the images below are based on the selection only of areas with strong activation indicated by a color spectrum. A summary of brain activation patterns for each direction of interpreting is provided at the beginning of **Section 6.17**.

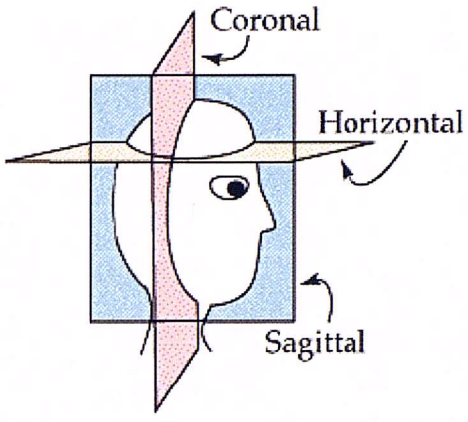


Figure 44. Coronal, Sagittal, and Horizontal Brain Images

Source: (Wellesley, Online)

A coronal brain image (colored pink) is taken between the front and the back of a brain. A horizontal image (colored beige) is collected between the top and the bottom of the brain. A sagittal image (colored blue) is obtained between the two sides of the brain.

(1) & (2) Bilateral (right & left) parahippocampal gyrus activation:

The parahippocampal gyrus is the gray matter that surrounds the hippocampus. Research has shown (Alkire *et al.*, 1998; Whatmough and Chertkow, 2007) that bilateral parahippocampal activation correlates strongly with word retrieval, word recall and semantic processing. Figures 45 and 46 show right and left parahippocampal gyrus activations respectively.

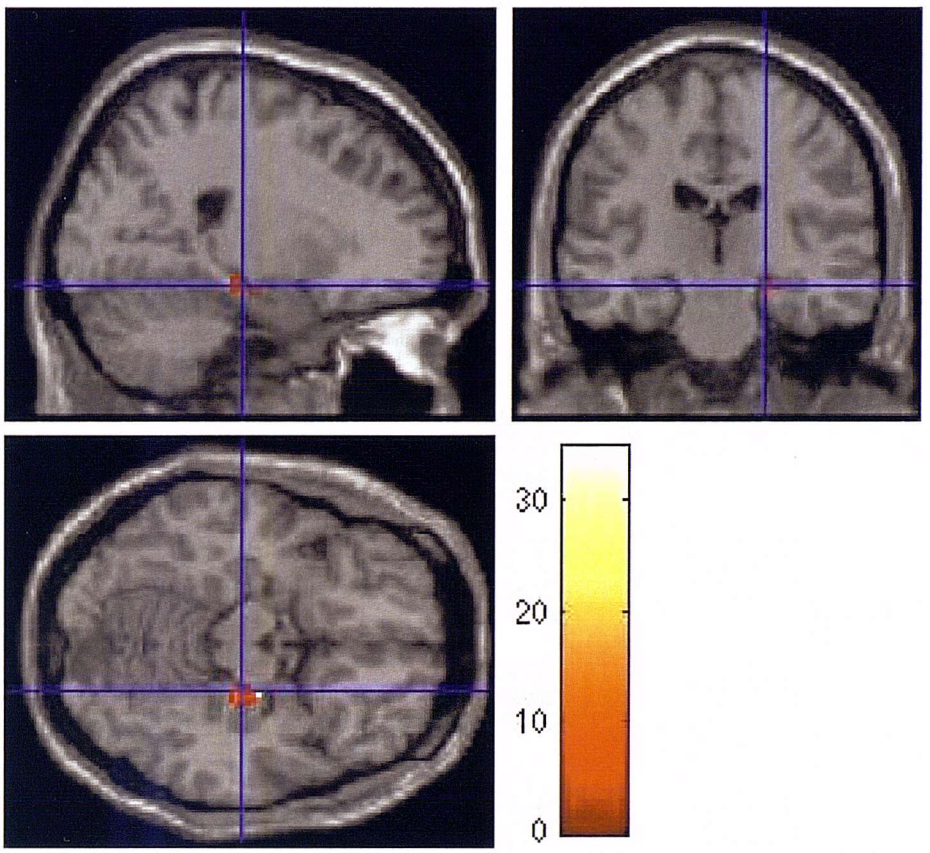


Figure 45. Activation of Right Parahippocampal Gyrus

The upper-left image is sagittal view; the upper-right one is a coronal image; and the lower-left image is a horizontal one. The colored bar spectrum represents the strength of activation of a particular area. Red means very strong activation, whereas yellow means relatively less strong activation. The numbers next to the spectrum represent the probabilities (*p* value). For example, 10 means that ten out of 100 (90%) times, an area will not be activated for a particular task; similarly, 5 means a probability of 95%. It has to be noted that the *p*-value criterion of the analysis in the study presented here was set at 0.05. The activated brain areas reported here passed this *p*-value criterion (see Tables 39 in **Section 6.18**).

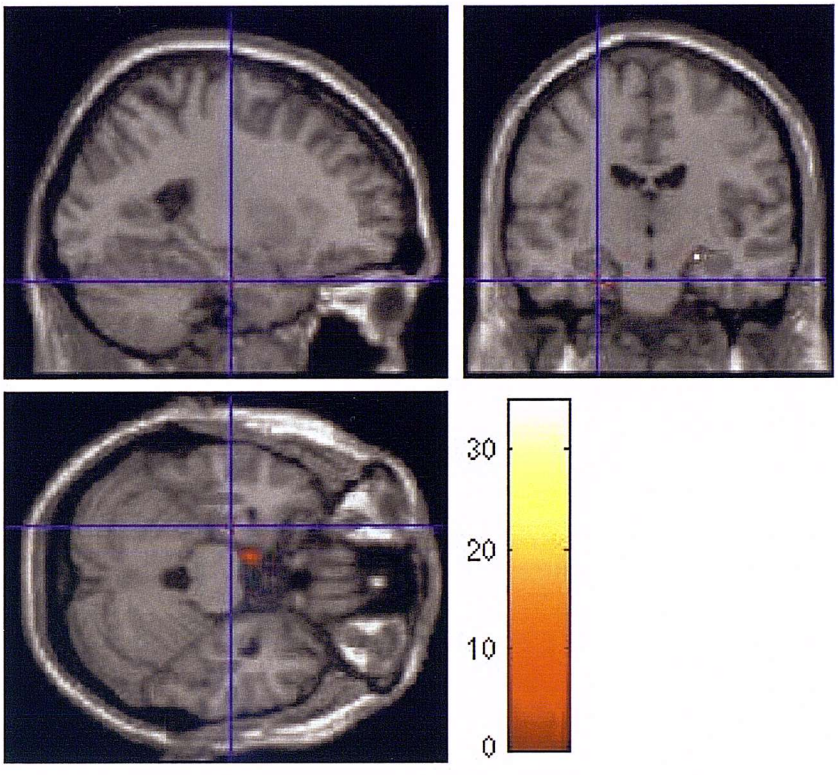


Figure 46. Activation of Left Parahippocampal Gyrus

(3) Activation of the left angular gyrus in the temporal lobe:

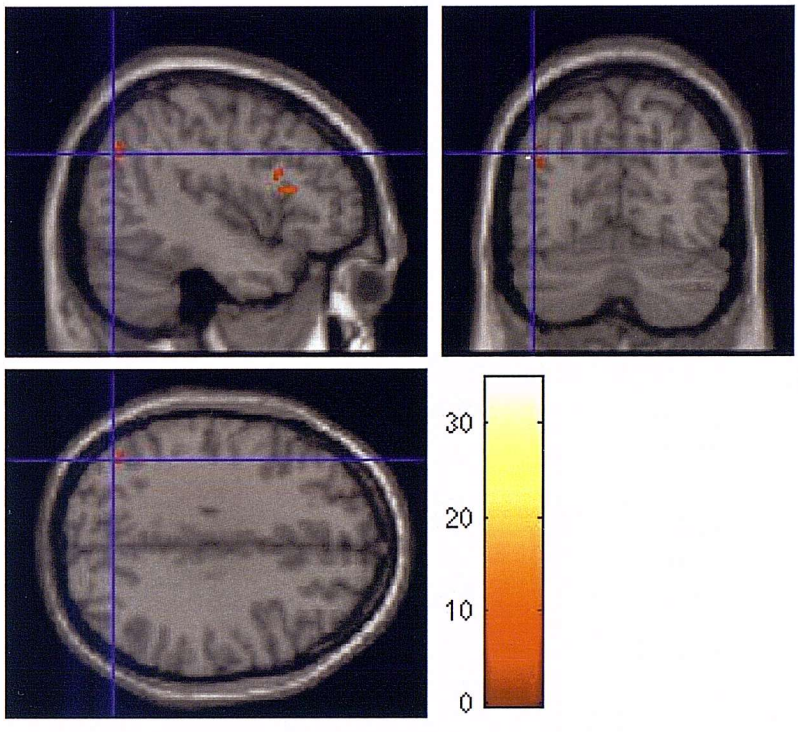


Figure 47. Activation of the Left Angular Gyrus

Prior research has shown that the left angular gyrus is associated with semantic processing (Mechelli *et al.*, 2007). It has also been reported (Law, 2007) that the activation response of the left angular gyrus is very important in the neural circuitry of reading and comprehending Japanese, an ideogram-based language. In the study presented in this thesis, the presence of Mandarin (also an ideogram-based language) in the language combination may have contributed to the activation of the left angular gyrus.

(4) Activation of the left inferior parietal lobule:

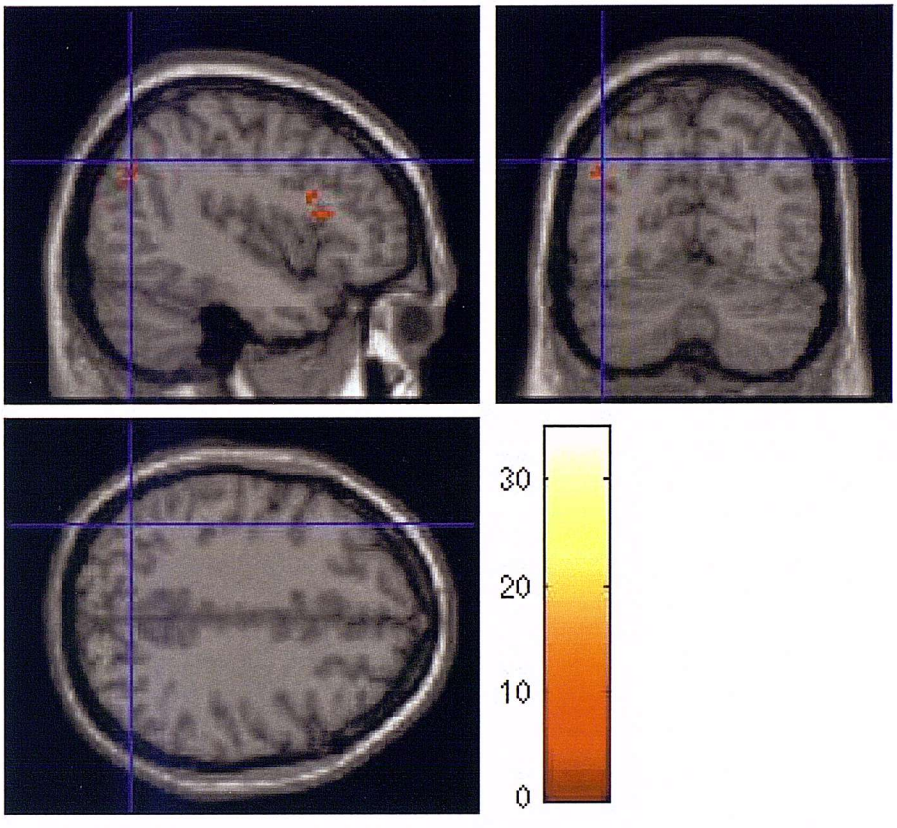


Figure 48. Activation of the left inferior parietal lobule

The left inferior parietal lobule is related to processing word imageability (Bedny and Thompson-Schill, 2006), and is also associated with verbal working memory (Kirschen *et al.*, 2005).

(5) Activation of the right putamen:

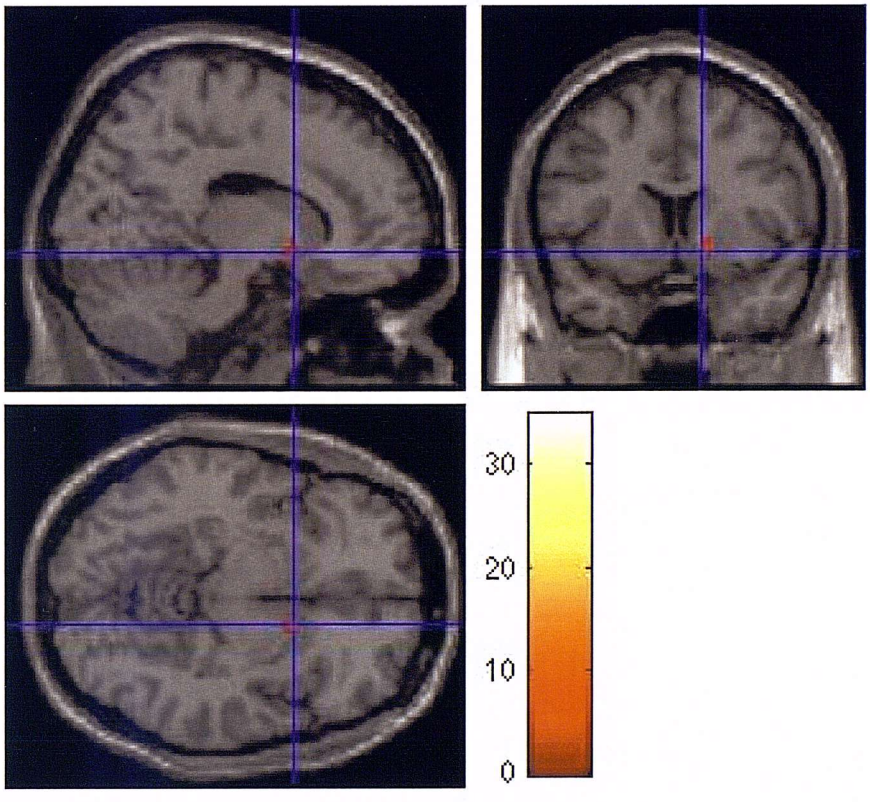


Figure 49. Activation of the Right Putamen

As previously demonstrated in a PET study by Price *et al.* (1999), the putamen is involved in single word translation. Further, the putamen is related to motor control (Ketcham *et al.*, 2003) and distinguishing language ambiguity (Ketteler *et al.*, 2008).

(6) Activation of the left inferior frontal gyrus:

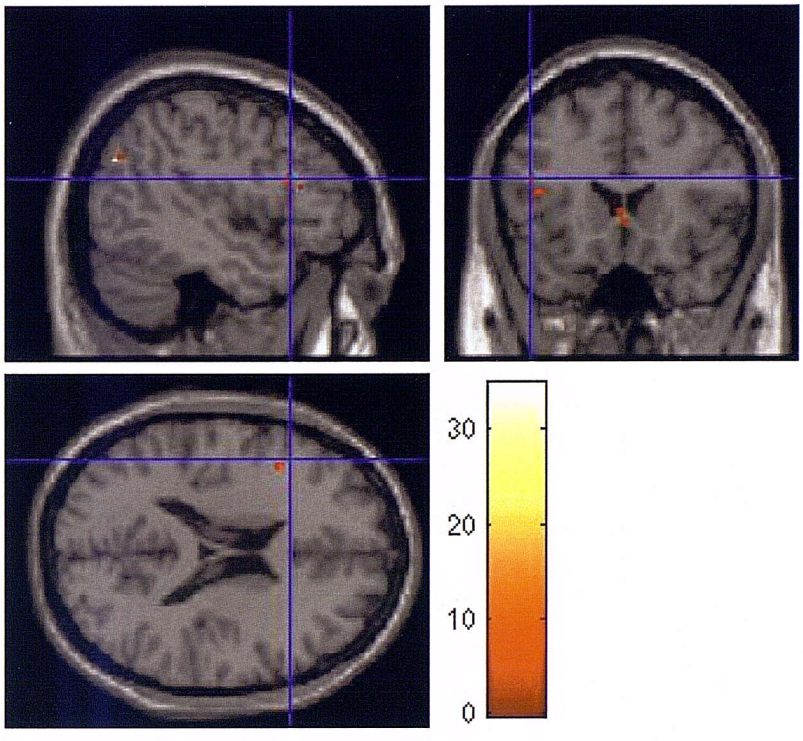


Figure 50. Activation of the Left Inferior Frontal Gyrus

This part of the brain is located in the Broca's area, and is therefore associated with language production and syntactic processing (Wang *et al.*, 2008). This area is also associated with language processing when language-related stimuli are visually presented (Sabri *et al.*, 2008).

(7) Activation of the uncus:

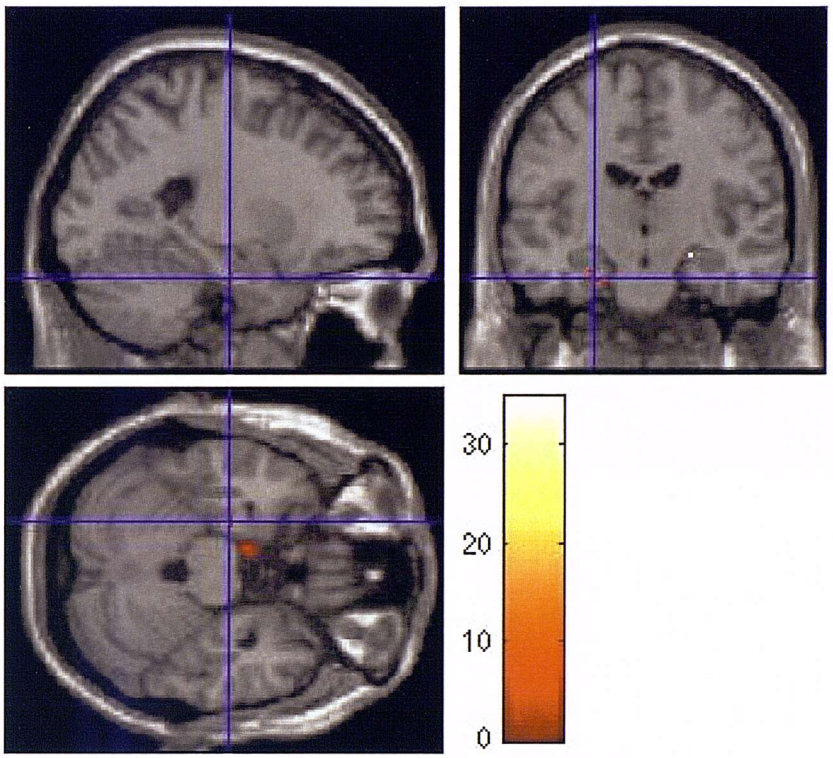


Figure 51. Activation of the Uncus

Research has shown that the uncus is associated with anxiety in producing difficult words of varying difficulty levels (Lorberbaum *et al.*, 2004).

(8) Activation of the left caudate head:

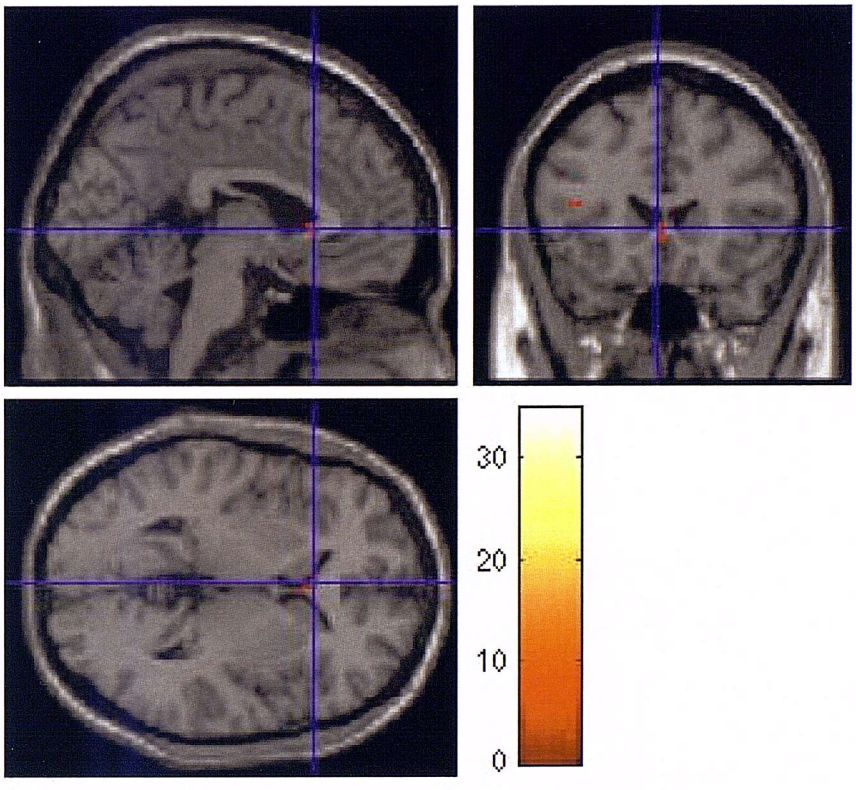


Figure 52. Activation of the left caudate head

Again, as reported in the PET study by Price *et al.* (1999), the head of the caudate is recruited for single word translation. Further, like the putamen, the caudate is associated with motor control (Ketcham *et al.*, 2003) and distinguishing language ambiguity (Ketteler *et al.*, 2008).

6.16.3 Brain Activation Patterns during Interpreting from Mandarin Chinese into English

A three-dimensional image of brain activation patterns for this interpreting direction was likewise produced by SPM5 (Figure 53). Images of gray matter activations in the limbic system will once again be shown in subsequent figures.

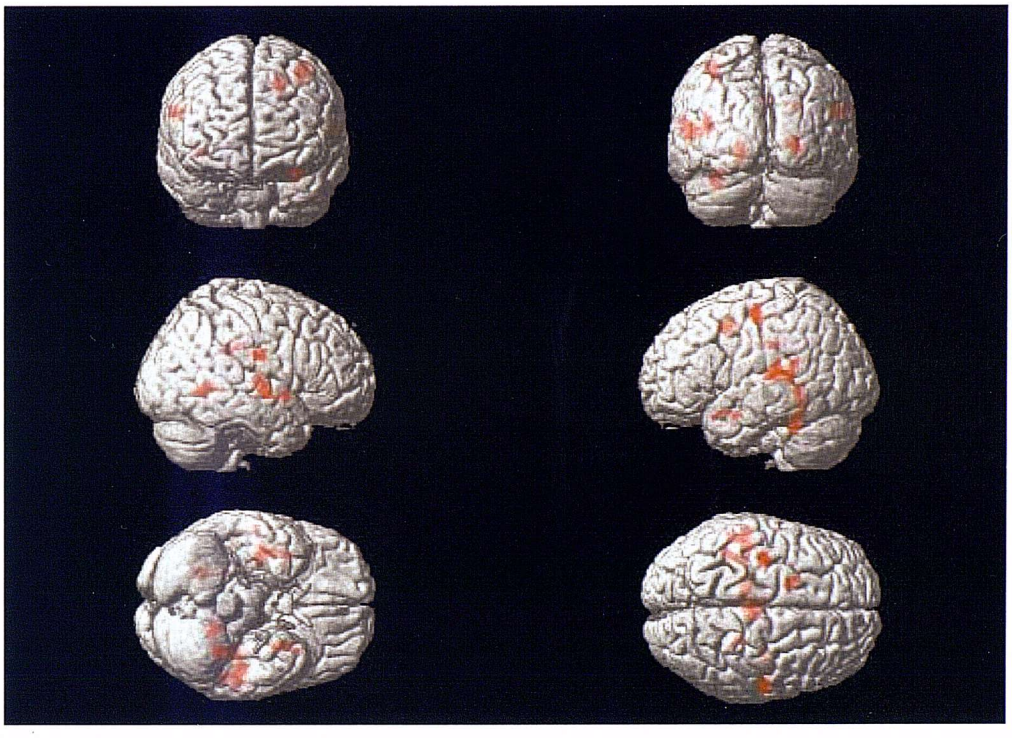


Figure 53. Brain Activation Patterns during Interpreting from Mandarin Chinese into English

(1) Activation of the right insula:

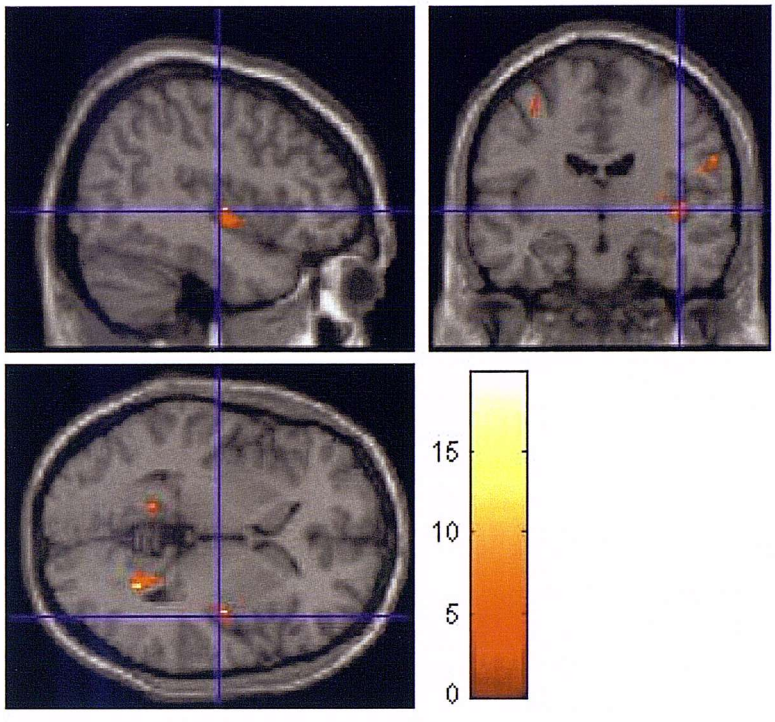


Figure 54. Activation of the Right Insula

It has been shown that the insula is involved in language-switching (Krick *et al.*, 2006). Riecker *et al.* (2000) have also demonstrated that the right insular cortex is involved in mediating speech and singing motor abilities, and in particular that 'the right insula might mediate temporo-spatial control vocal tract musculature' (ibid.: 1997) during sound production.

(2) Activation of the right putamen:

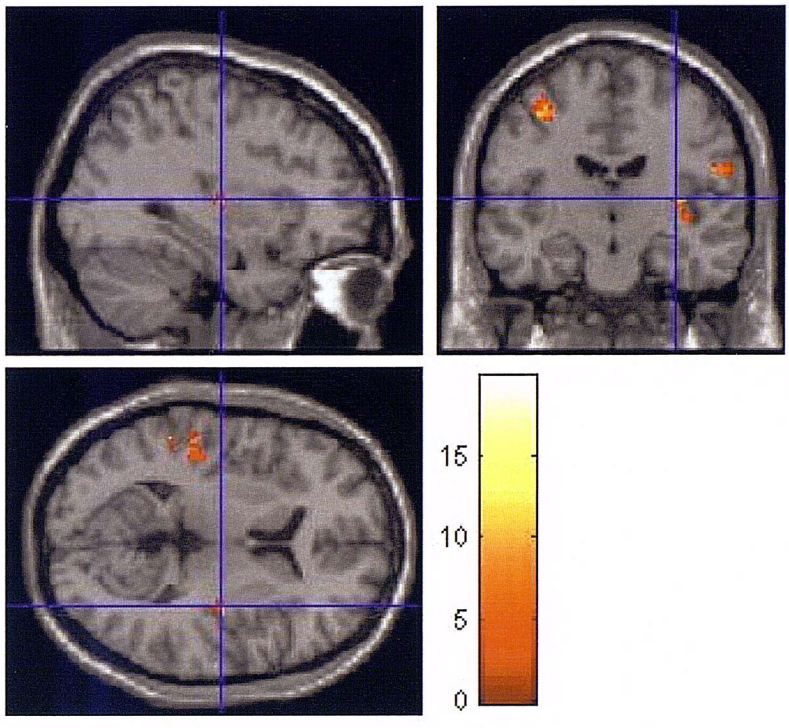


Figure 55. Activation of the Right Putamen

It has been previously mentioned that Price *et al.* (1999) showed that the putamen is involved in single word translation. Further, the putamen is related to motor control (Ketcham *et al.*, 2003) and distinguishing language ambiguity (Ketteler *et al.*, 2008).

(3) Activation of the cingulate gyrus:

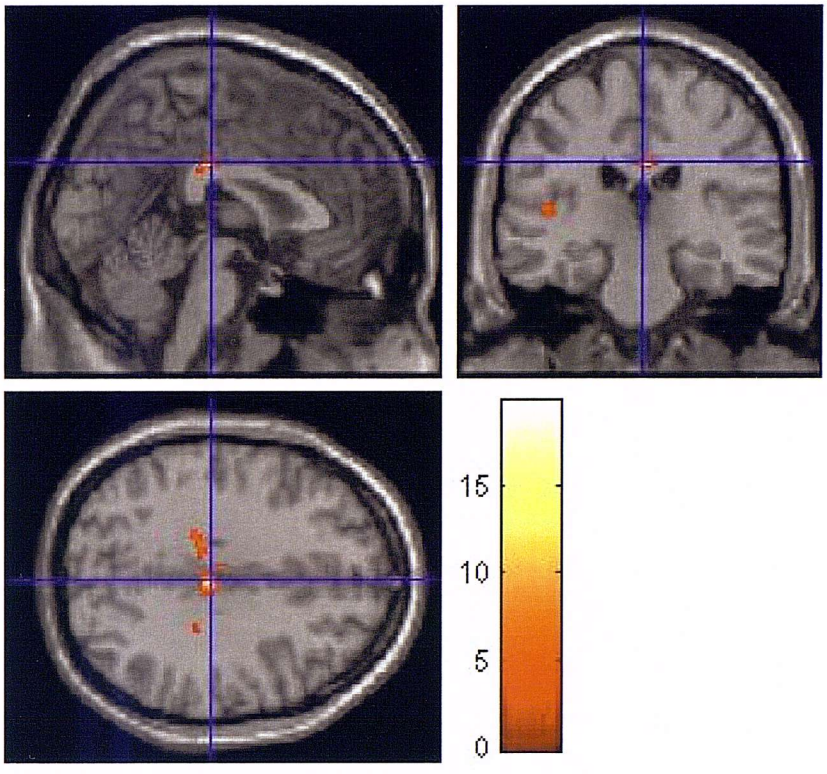


Figure 56. Activation of the Cingulate Gyrus

The cingulate gyrus is associated with alertness (Rota *et al.*, 2008), and has also been shown to be related to the retrieval of semantic memory (Assaf *et al.*, 2006) and strategic semantic search when a meaning is ambiguous (Ketteler *et al.*, 2008).

(4) Activation of the left dorsolateral prefrontal cortex:

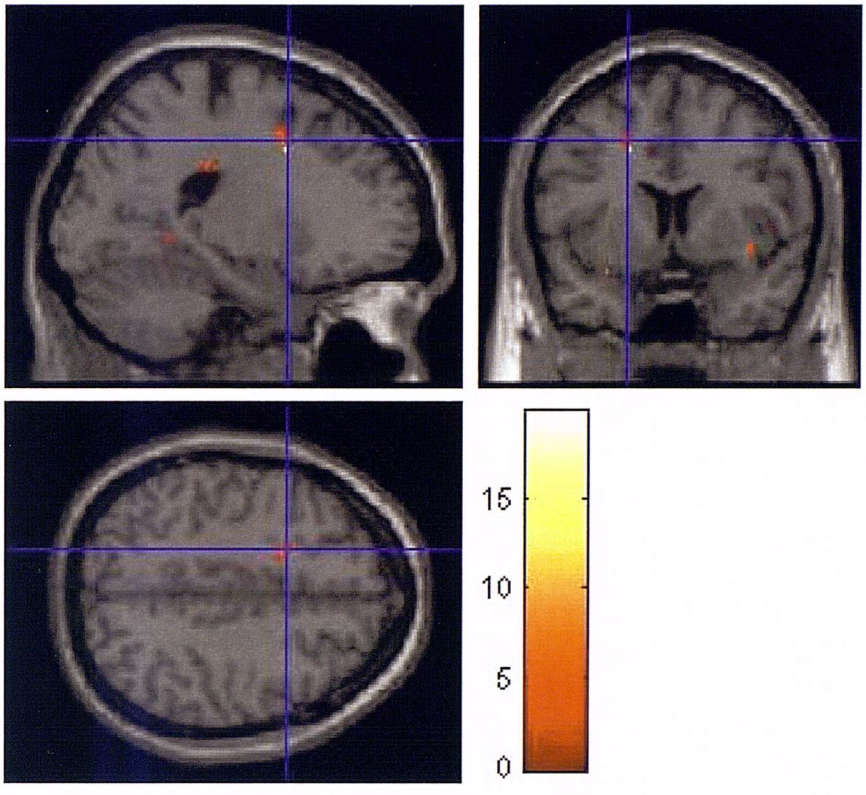


Figure 57. Activation of the Left Dorsolateral Prefrontal Cortex

The left dorsolateral prefrontal cortex is associated with executive function (Ward, 2006: 285). Further, this cortical area has been demonstrated to be crucial in supporting the cognitive task of single word translation (Klein *et al.*, 1995; Price *et al.*, 1999) and language switching (Krick *et al.*, 2006).

(5) Activation of the right lingual gyrus in the occipital lobe:

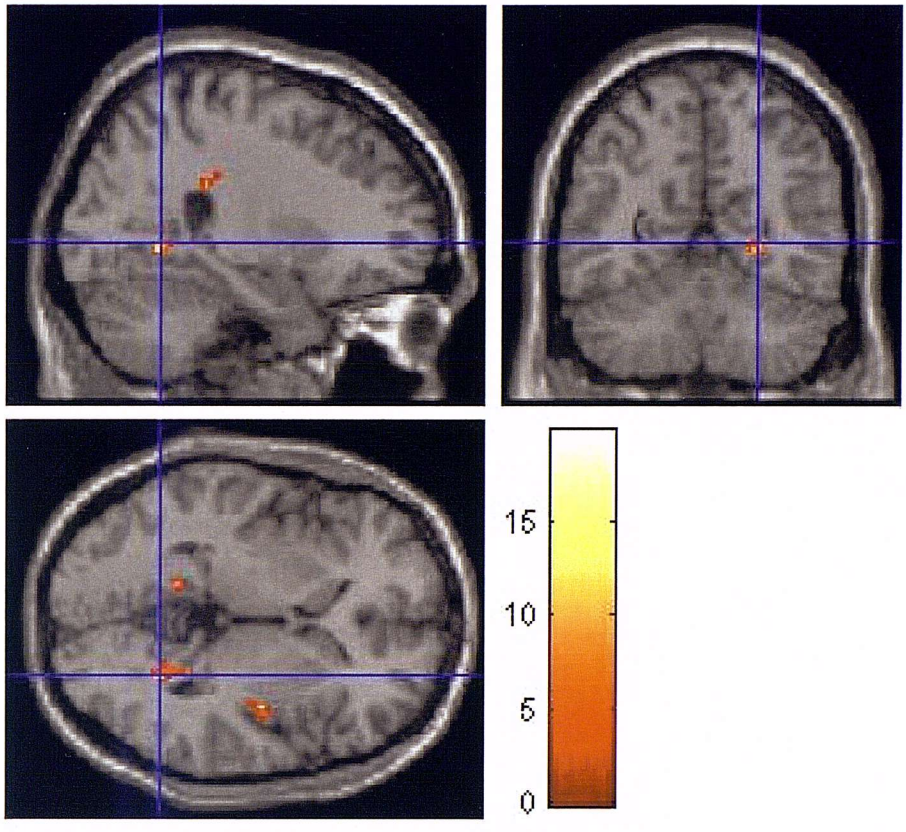


Figure 58. Activation of the Right Lingual Gyrus

The lingual gyrus in the occipital lobe is related to processing visually presented words and objects (Borowsky *et al.*, 2007). As it is situated in the occipital lobe, it mainly deals with visual inputs (Gazzaniga *et al.*, 2002).

(6) Activation of the left superior temporal gyrus:

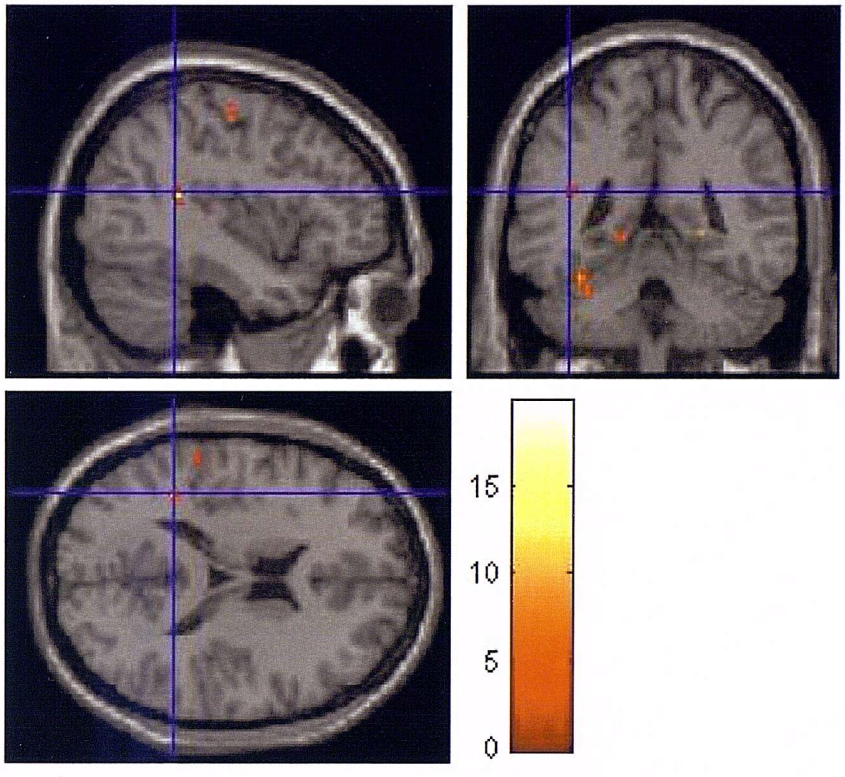


Figure 59. Activation of the Left Superior Temporal Gyrus

The left superior temporal gyrus is part of Wernicke's area (Gazzaniga *et al.*, 2002), which is responsible for speech or language comprehension. Other research has also demonstrated that the left superior temporal gyrus strongly correlates with speech and language comprehension (Virtue *et al.*, 2008) and perception (Sabri *et al.*, 2008).

(7) Activation of the transverse temporal gyrus (or named the Heschl's gyrus):

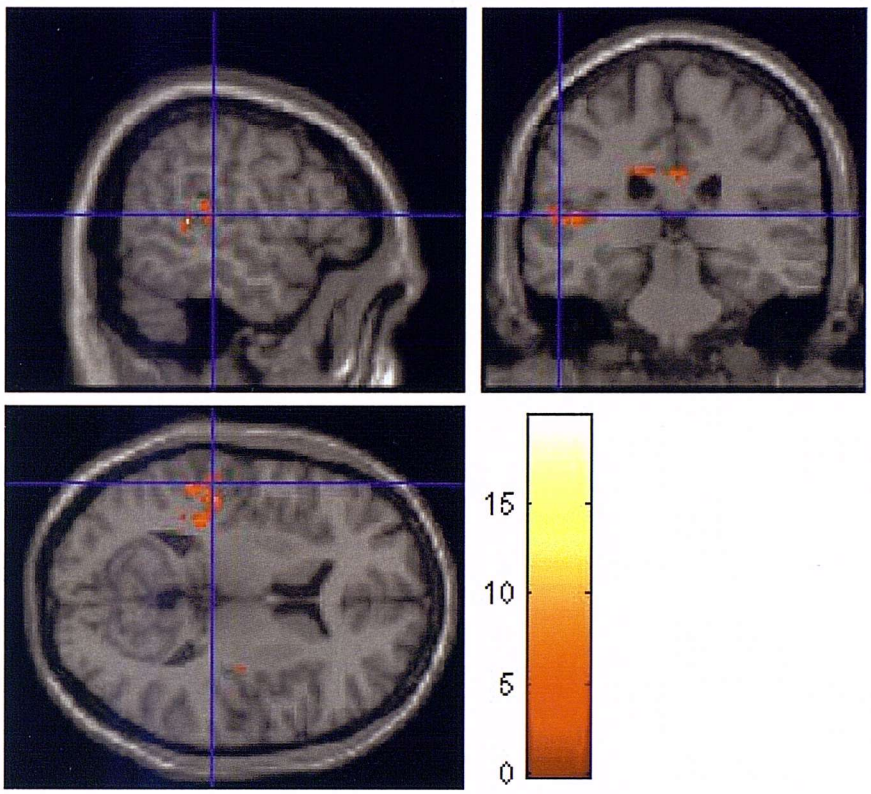


Figure 60. Activation of the Left Superior Temporal Gyrus

The transverse temporal gyrus is also called Heschl's gyrus, and is associated with processing auditory information (Ward, 2006). However, research has also shown that Heschl's gyrus is involved in language comprehension, as it is adjacent to the left superior temporal gyrus (Brauer *et al.*, 2008).

(8) Activation of the left parahippocampal gyrus:

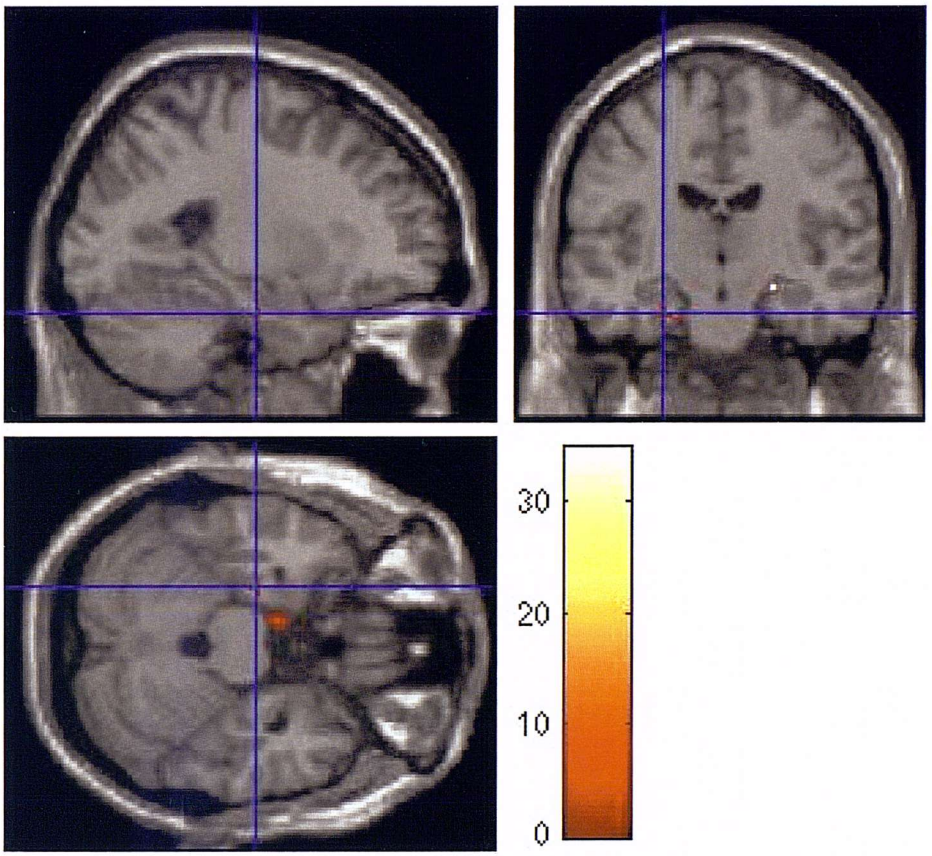


Figure 61. Activation of the Left Parahippocampal Gyrus

As previously indicated, research has shown (Alkire *et al.*, 1998; Whatmough and Chertkow, 2007) that parahippocampal activation correlates strongly with word retrieval, word recall and semantic processing.

(9) Activation of the right superior temporal gyrus:

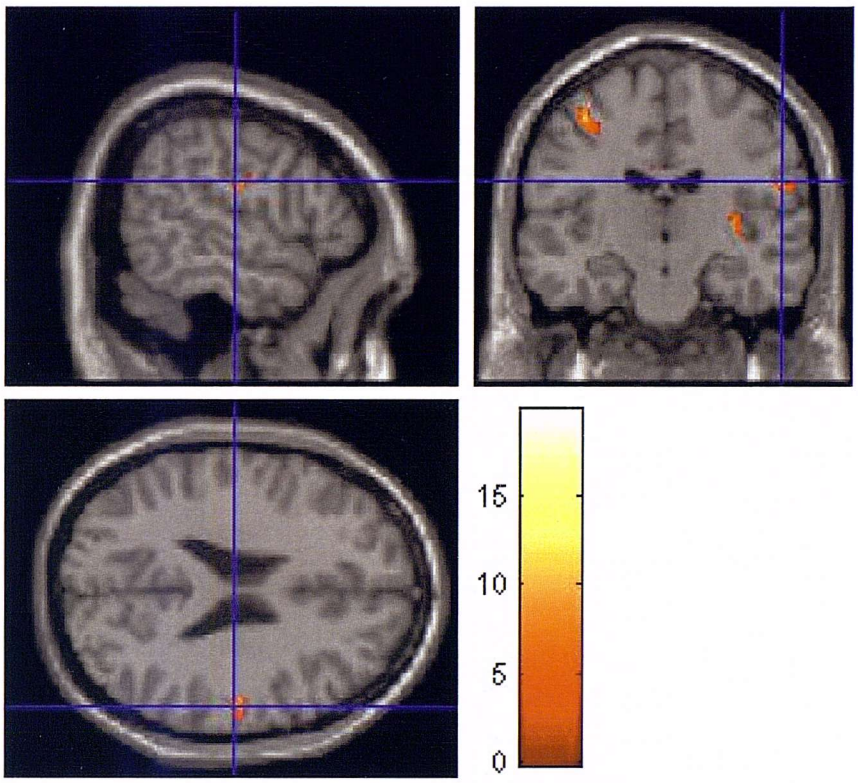


Figure 62. Activation of the Right Superior Temporal Gyrus

Like the left superior temporal gyrus, the right superior temporal gyrus is associated with speech and language comprehension. However, the left hemisphere usually deals with logic and meanings at face value, whereas the right hemisphere handles intended emotions and intonations. The two areas work together to help an individual to understand not only a language (the left superior temporal gyrus) but also the feelings associated with it (the right superior temporal gyrus) (for a more detailed discussion, see Gazzaniga *et al.*, 2002).

(10) Activation of the precentral gyrus in the frontal lobe:

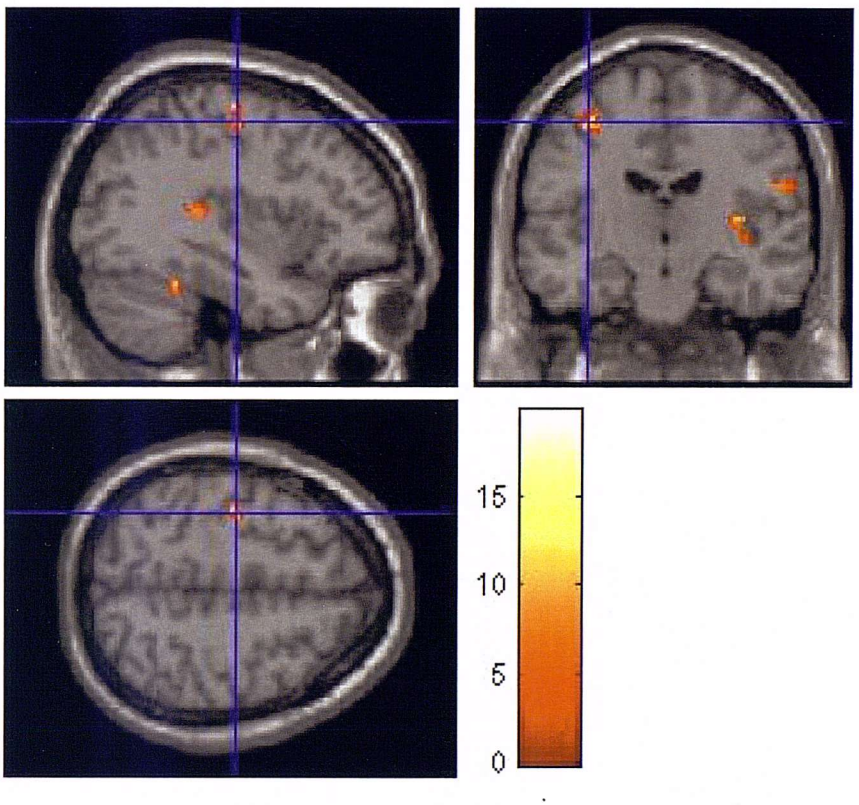


Figure 63. Activation of the Precentral Gyrus

The precentral gyrus is also called the primary motor cortex, and it controls mechanical movement. Thus, it is associated with controlled orofacial movement for speech articulation (Song *et al.*, 2007).

(11) Activation of the cerebellum:

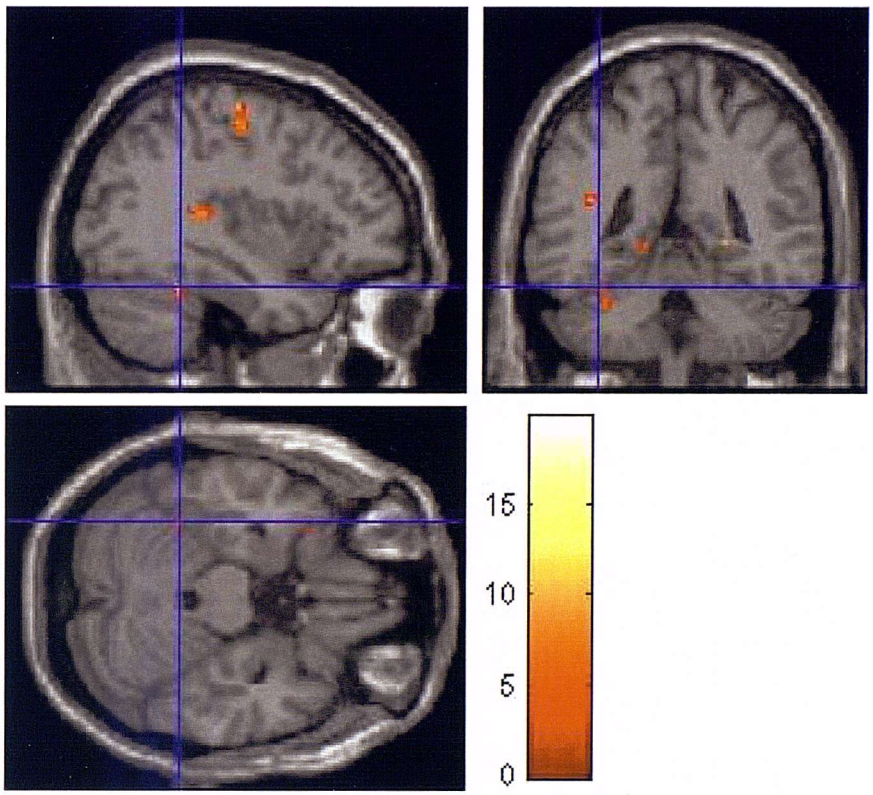


Figure 64. Activation of the Cerebellum

Cerebellum activation is known to be associated with speech production (Borbely *et al.*, 2005; Wu and Lin, 2008). It has also been shown that the cerebellum helps to sustain verbal working memory (Chiricozzi *et al.*, 2008).

(12) Activation of the right fusiform gyrus in the temporal lobe:

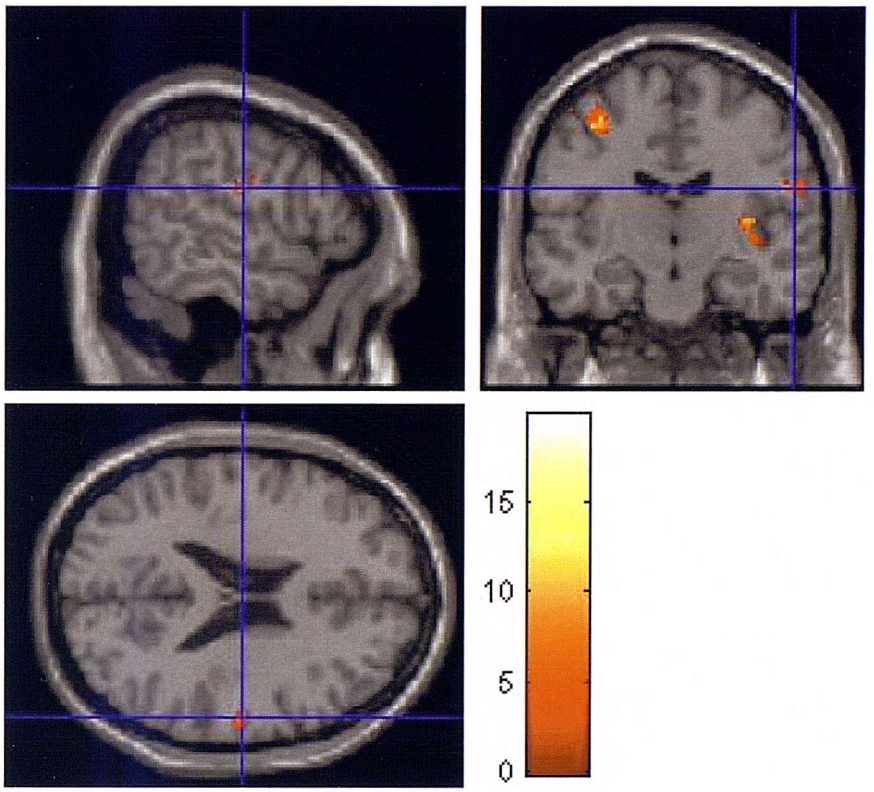


Figure 65. Activation of the Right Fusiform Gyrus

The right fusiform gyrus is associated with word and number recognition during reading (Gazzaniga *et al.*, 2002; Ward, 2006; Devlin *et al.*, 2006; Liu *et al.*, 2008).

6.17 Findings

To summarize the neural correlates regarding A→B and B→A interpreting, the brain activation patterns of the two tasks are listed as follows:

- **A→B Interpreting:**

(1) Right insula; (2) the right putamen; (3) the cingulate gyrus; (4) the dorsolateral

prefrontal cortex; (5) the right lingual gyrus; (6) the left superior temporal gyrus; (7) the Heschl's gyrus; (8) the left parahippocampal gyrus; (9) the right superior temporal gyrus; (10) the precentral gyrus; (11) the cerebellum; and (12) the right fusiform gyrus

- **B→A interpreting:**

  (1)The right parahippocampal gyrus; (2) left parahippocampal gyrus; (3) the left angular gyrus in the temporal lobe; (4) the left inferior parietal lobe; (5) the right putamen; (6) the left inferior frontal gyrus; (7) the uncus; (8) and the head of the left caudate.

As noted previously, sight interpreting at a textual level is a cognitive task that requires orofacial movement, which can be a source of problems in an fMRI experiment designed to explore its brain activation patterns. This issue has not been resolved in previous research. However, the above findings suggest that 'sparse temporal sampling' (Hall *et al.*, 1999) can eliminate potential artefacts caused by orofacial movement. With this paradigm, fMRI can be satisfactorily applied to the exploration of the cortical activity accompanying interpreting in both directions. This suggests that fMRI can be adequately applied to T&I Studies.

Further, the results based on a limited number of five participants suggest that sight interpreting into a second language recruits more gray matter activations than sight interpreting into a first language. This, in turn, suggests that A→B sight interpreting is cognitively more demanding than B→A sight interpreting. Even though sight interpreting is a verbal language conversion task which is different from written translation, the fMRI results have indicated similar results (regarding asymmetry) to those yielded by the two eye-tracking experiments that involved written text translation presented in this thesis. Despite of the limited number of the subjects involved in the fMRI study and the preliminary, tentative results yielded, the fMRI results to a certain extent lent support to that notion that the predictions suggested by the Revised Hierarchical Model were valid at a textual level.

6.18 Discussion

The above results reveal that some cortical activities are common to A→B and B→A sight interpreting. These are activations in the left parahippocampal gyrus, frontal lobe and temporal lobe. It has been noted that the parahippocampal gyrus is associated with facilitating word retrieval, word recall and semantic processing (Alkire *et al.*, 1998; Whatmough and Chertkow, 2007). This suggests that these functions are crucial to the cognitive task of sight interpreting, regardless of directionality.

When a person is sight interpreting, s/he is listening/reading and comprehending a speech or text while producing a translation in another language. These events all happen at the same time, and require the recall and retrieval of words for both language comprehension and production. This may be why there was activation of the parahippocampal gyrus in both directions in the study presented in this thesis. Even though this result was not reflected in the studies by Klein *et al.* (1995), Price *et al.* (1999) and Rinne *et al.* (2000) (see Table 11), this is the first study that draws attention to the potentially significant role of the parahippocampal gyrus during sight interpreting. A possible explanation for this is because the fMRI study presented in this thesis employed 'textual stimuli' which perhaps required a higher level of word retrieval, word recall and semantic processing as shown by Alkire *et al.* (1998) and Whatmough and Chertkow (2007).

In addition, both A→B and B→A sight interpreting involved frontal and temporal lobe activations. As previously introduced in **Section 6.3**, the frontal lobe is mostly associated with language production (the Broca's area), planning and execution of body and eye movements, cognition and emotions. The primary motor area (precentral gyrus) in the frontal lobe controls the mechanical actions of movements, whilst the premotor cortex (anterior to the primary motor cortex) is involved in decision-making and planning of motor activities. The prefrontal cortex is related to executive function, decision-making, cognition, emotions and memories (Xiao *et al.*, 2008). The temporal lobe also plays a crucial part for language processing. The Wernicke's area is involved in language perception and

comprehension (Andoh *et al.*, 2008). Adjacent to this area are the cortices often associated with long-term lexical and semantic memory (Venneri *et al.*, 2008). Given some cognitive similarities between A→B and B→A sight interpreting, the common frontal and temporal activations indicated by the study here seemed to make sense.

It is interesting to see that sight interpreting in both directions activated some parts of cortical areas in the limbic system (A→B sight interpreting: putamen; B→A sight interpreting: caudate head). Further, A→B sight interpreting recruited cerebellar activation. All these areas are related to articulation (see Dronkers, 1996; Price *et al.*, 1996; Wise *et al.*, 1999). Also, activation of the precentral gyrus (primary motor cortex) during A→B sight interpreting shows that this task requires more motor control regarding orofacial movement for articulation. The results of extra cerebella activation during A→B sight interpreting suggest that the cognitive task requires more articulatory effort. This is perhaps because articulating a second language is less spontaneous than articulating a first language. Interestingly, Price *et al.* (1999) also found extra activation of the insula and cerebella during single word translation, compared to mere reading. Klein *et al.* (1995) also found that translating a single word into a second language recruited the putamen, and this is consistent with the results of A→B sight interpreting in the present study. However, Rinne *et al.*'s (2000) results did not show any activation in the limbic system or cerebellum in either direction. This may be because the subjects they used in their study were professional interpreters who required less conscious effort in articulation due to their abundant experience, as opposed to the novice interpreters used in this research, who were inevitably less expert in articulation.

The results yielded by the PET study conducted by Rinne *et al.* (2000) showed that simultaneously sight interpreting into a second language recruited more extensive left-sided fronto-temporal activation. In contrast, the results of the study presented here show that there is more right-sided fronto-temporal activation during A→B sight interpreting. It has been shown that handling a tonal language, e.g. Mandarin Chinese, whether aurally or verbally, requires more right hemispheric activation (Gandour *et al.*, 2003). This may explain

why right hemispheric activations were observed in this study for the language combination of Mandarin Chinese and English. Even so, frontal activation is strongly associated with language production and decision-making and temporal activation relates to language comprehension and memory retrieval. This explains why there is activation in fronto-temporal cortices during both tasks, regardless of directionality.

Klein *et al.* (1995) reported that during single word translation into a second language, activation was mainly shown at the dorsolateral prefrontal cortex, associated with executive function or decision making (Ward, 2006). Rinne *et al.* (2000) also reported intensive activation in the dorsolateral prefrontal cortex during simultaneous sight interpreting into a second language. This is confirmed by the results of A→B sight interpreting in the study presented here, which have also shown strong activation in the dorsolateral prefrontal cortex. This may be because sight interpreting into a second language requires *more* executive control than sight interpreting into the mother tongue. Klein *et al.* (1995) emphasize the importance of the dorsolateral prefrontal cortex during A→B single word translation, because it is an area strongly associated with working memory and executive function which is considered very important to the task of interpreting (Gile, 1995; Gile, 1997; Liu, 2001; Christoffels *et al.*, 2003).

Rinne *et al.* (2000) observed that the professional interpreters they used as subjects in their PET study performed well in both interpreting directions (2000: 88), and their results did not reveal a clear difference in the number of activated areas in A→B vs. B→A interpreting. However, in this fMRI study (using novice interpreters), the number of activated cortical areas is larger during A→B sight interpreting than that during B→A sight interpreting. This suggests that the interpreter's proficiency or experience may have an effect on the neural representation of cognitive loading during bilateral sight interpreting. In other words, this suggests that the more novice an interpreter is, the more likely it is for him/her to show a clear difference in the number of activated gray matter areas during A→B and B→A sight interpreting.

The left inferior frontal gyrus has been shown to be associated with semantic

processing and lexical search (Petersen *et al.*, 1988; Wise *et al.*, 1991; Demonet *et al.*, 1992). In this study, such an activation was only observed during B→A sight interpreting. A possible explanation of the result is that 'semantic processing' during B→A sight interpreting (where reading and *semantically processing* a B language before a verbal output is required) was more prominent.

Like Price *et al.* (1999), the fMRI study presented here found activation of the insula during A→B interpreting. Research has shown that the insula is associated with such cognitive functions as risk prediction and rapid updating of reward anticipation (Preuschoff *et al.*, 2008). This suggests that A→B sight interpreting may require more risk prediction, as producing a second language may involve more uncertainty.

Activation of the cingulate gyrus was observed during A→B sight interpreting in this study. The cingulate gyrus has been demonstrated to be associated with alertness (Mander *et al.*, 2008), attention, language control (Christensen *et al.*, 2008; Abutalebi *et al.*, 2008) and optimizing behavior (Quilodran *et al.*, 2008). This suggests that A→B sight interpreting may require more alertness, attention control, and optimization of output behavior, compared with B→A sight interpreting. The activation of the cingulated gyrus was also reflected in Price *et al.*'s (1999) study.

It is worth noting that, during A→B sight interpreting, the lingual gyrus in the occipital lobe was recruited, which is not observed during B→A sight interpreting. This may be because the experimental task was 'sight interpreting' which involved visual stimuli. Further, Mandarin Chinese is a logographic, ideogram-based language. Even though it is the subjects' native language, processing a language represented by square-shaped, picture-like symbols may still require extra occipital activation. Further, occipital activation was not shown in studies conducted by Klein *et al.* (1995), Price *et al.* (1999) and Rinne *et al.* (2000) perhaps because their stimuli were auditorily presented and because their language combinations did not include any ideogram-based, tonal language.

Unlike Klein *et al.*, the present study found activation in the left inferior parietal cortex during *B→A sight interpreting*. This area is associated with the voluntary reorientation

of attention (Hopfinger *et al.*, 2000). Sight interpreting is a cognitive task that requires an interpreter to voluntarily reorient his/her attention moment-by-moment (Gile, 1997), which may explain why the left inferior parietal cortex was recruited. Even though the nature of the task was different from that of the present study, Krick *et al.* (2006) also found activation of this area during language switching, which also requires voluntary reorientation of attention. However, this area was recruited only during B→A sight interpreting in the study presented in this thesis, whereas the same area was demonstrated to be active during A→B translation by Klein *et al.* (1995). This suggests that experiments involving tasks that require similar mental processes are not guaranteed to yield the same cerebral results.

In the present study, during A→B sight interpreting, areas adjacent to Wernicke's area found to be activated were Heschl's gyrus, the bilateral superior temporal gyrus and the fusiform gyrus. Even though they are mainly responsible for processing auditory stimuli, in combination with Wernicke's area, they help the human brain to comprehend language (Gazzaniga *et al.*, 2002). They not only assist in the understanding of an incoming stimulus, but also help to monitor what is being verbally articulated in another language during sight interpreting. This may explain why these areas were strongly activated during A→B sight interpreting, as perhaps more effort is needed to auditorily monitor/comprehend the oral output in a second language during the ongoing task.

The results of three earlier studies relevant to this fMRI study, presented alongside the results of the present study, are tabulated to facilitate discussions (Table 39); however, it has to be noted that, due to the different experimental designs of the various studies in the table, no attempt is made to yield direct comparisons. They used PET to examine translation, language switching or sight interpreting, and yielded results with good spatial resolution. Further, their neurological data were validated by behavioral results.

Studies	Klein *et al.* (1995: 2901)	Price *et al.* (1999: 2227)	Rinne *et al.* (2000: 86)	fMRI Study Presented in this Thesis
A→B or B→A	A→B	A→B & B→A	A→B	A→B
Results	1) Left ventrolateral prefrontal cortex 2) Left dorsolateral prefrontal cortex 3) Left media prefrontal cortex 4) Left inferior temporal cortex 5) Left posterior parietal cortex 6) Left inferior parietal cortex 7) Parietal occipital cortex 8) Cerebellum 9) Putaman 10) Thalamus; Pulvinar	1) Anterior cingulated 2) Left putamen; head of caudate 3) Supplementary motor area (SMA) 4) Right putamen; head of caudate 5) Left anterior inferior insula 6) Medial cerebellum 7) Left medial fusiform gyrus 8) Left cerebellum 9) Right cerebellum	1) Ventrolateral frontal cortex 2) Inferior temporal cortex 3) Premotor cortex 4) Cerebellum	1) Right insula 2) Right putamen 3) Cingulate gyrus 4) Left dorsolateral prefrontal cortex 5) Right lingual gyrus 6) Left superior temporal gyrus 7) Heschl's gyrus 8) Left parahippocampal gyrus 9) Right superior temporal gyrus 10) Precentral gyrus 11) Cerebellum 12) Right fusiform gyrus
				B→A
				1) Right parahippocampal gyrus 2) Left parahippocampal gyrus 3) Left angular gyrus 4) Left inferior parietal cortex 5) Right putamen 6) Left inferior frontal gyrus 7) Uncus 8) Head of the left caudate

Table 39. Comparison of the Present Results and Those from Prior Studies

6.19 Conclusion

The exploratory fMRI study presented here has demonstrated that the method of fMRI can be adequately applied to the field of T&I Studies. As previously indicated, prior neuroimaging research specifically investigating T&I suffers from a number of weaknesses: (1) orofacial movement was not allowed during scanning; (2) the subjects recruited were bilinguals or professional interpreters; and (3) textual stimuli were not used, which decreases the level of ecological validity. In the fMRI study presented in this thesis, even though there were only five participants, all of them were trained novice interpreters and the study did indeed generate interesting preliminary results. Further, textual stimuli were selected for the study, and the 'sparse paradigm' was adequately employed to allow for orofacial movement. This helped elevate the level of ecological validity of the study. Therefore, the results have shown that fMRI is a viable method for the discipline of T&I Studies.

To stress yet again, the fMRI results have indicated that Mandarin → English interpreting recruited more cortical activations than English → Mandarin interpreting. This suggests that A→B interpreting was more cognitively demanding than B→A interpreting. Even though the number of participants involved was limited and the results were hence preliminary, they still lent support to the idea that the predictions suggested by the Revised Hierarchical Model were valid at a *textual* level.

Chapter Seven Discussion, Limitations and Recommended Future Work

7.1 Summary of the Research Program

This doctoral study was conducted with two primary aims: The first aim was to test the applicability of two methods of research or techniques, i.e. eye-tracking and fMRI, to the field of T&I Studies, in particular by investigating cognitive loading of different translation/sight interpreting directions. The second aim was to explore whether or not assumptions and predictions suggested by the Revised Hierarchical Model regarding the asymmetry are valid at a *textual* level. Very little has been done previously to achieve these two aims. The entire research project described in this thesis has presented empirical evidence to successfully achieve *the two aims*.

The eye-tracking laboratory was set up from scratch by the researcher. The fMRI laboratory, on the other hand, was already set up for use, and the study would not have been possible without substantial technical support from physicists, engineers and doctors associated with the Imaging Sciences Department of Imperial College London. The experimental design of the fMRI study stemmed from intensive discussions and collaborations with these expert colleagues.

The methodological rationale and practical design of all aspects of this study have been reported in detail, so future researchers with a Humanities rather than a Science background will be able to pursue a similar line of investigation.

Based on the results and findings, it can be concluded that both eye-tracking and fMRI technologies have a useful contribution to make to the field of T&I Studies. Their applicability has been demonstrated by their capacity to explore, and at least in the case of eye-tracking, to answer the research questions regarding cognitive loading of different translation/sight interpreting directions. Furthermore, in the course of the study a number of

interesting issues that merit further investigation were identified as a result of careful statistical analysis. Both the main findings and the additional lines of future enquiry that these suggested are summarized in the sections that follow.

In addition, the three experiments presented in the thesis have lent strong, affirmative evidence to confirm that the predictions suggested by the Revised Hierarchical Model were valid at a 'textual' level. This will also be summarized in this chapter.

7.2 Relevance of the Research Findings to Theoretical Considerations

The eye-tracking experiments consistently confirmed that there was a significant difference in the cognitive loading between translating A→B and translating B→A. Additionally, the fMRI study also provided preliminary evidence for this same phenomenon. This can be explained, at least up to a point, by the Revised Hierarchical Model. However, this theory was developed using a 'single word' experimental paradigm, whereas translation mainly deals with multi-word texts. Given the nature of the experiments presented in this thesis, it has been demonstrated that the predictions suggested by the Revised Hierarchical Model were valid at a 'textual' level.

It will be remembered that the predictions of the Revised Hierarchical Model have been demonstrated to be valid by this research program for textual translation/sight interpreting which requires dealing with either concrete and abstract concepts. It has to be acknowledged though that the texts used in the research program presented in this thesis were simple, non-technical short texts, because factors, such as text difficulty and genre, may have an effect on 'translation asymmetry' suggested by the Revised Hierarchical Model.

7.3 Summary of Research Conclusions

It is appropriate to summarize the answers provided by this research program to the research questions posed in **Section 1.2**.

- Apart from pupillometric data, other eye-tracking data are potentially useful, in

particular 'fixation frequency'. This affirmatively shows the broader applicability of eye-tracking technology in the T&I field.

- The eye-tracking experiments have given rise to additional interesting findings that can be utilized as new research hypotheses to be tested by future investigations. For instance, novice translators paid more visual attention to the target text independently of the direction of translation; typing Spanish or Mandarin was cognitively more demanding than typing English, even though the subjects were native Spanish or Mandarin speakers; reading an individual's A language was less cognitively demanding than reading her/his B language; and the 'familiarity effect' was not a universal phenomenon. This clearly demonstrates the investigative potential and broader applicability of eye-tracking technology in the T&I field.

- The technique of 'sparse temporal sampling' permits an adequate fMRI experimental paradigm that overcomes the potential difficulties associated with orofacial movement during imaging. In this way, fMRI can be successfully applied to explore neural correlates of A→B and B→A sight interpreting, confirming that, like eye-tracking, this technology also has application to T&I Studies.

- A detailed analysis of eye-tracking data, supported by preliminary fMRI results, confirms that there is a significant difference in the cognitive loading experienced by novice translators and interpreters during A→B and B→A translation and interpreting of textual material.

- It is confirmed that the predictions suggested by the Revised Hierarchical Model indicating directionality has an effect on cognitive loading are valid at a *textual* level. Thus all of the research questions that motivated the study have been given satisfactory answers, based on carefully designed experiments within an adequate methodological framework, and including rigorous statistical analysis.

To summarize, in the research program presented in this thesis, both the eye-

tracking and fMRI methods have been applied in improved conditions, and therefore the research program presented here contributes to advancing methodology in T&I Studies. Further, new findings were identified, e.g. more attention was paid to TT and the lack of confirmation for the 'familiarity effect'. This contributes to forming a basis for a number of more concrete hypotheses to be tested by future research in T&I Studies. Also, the research program has confirmed that the predictions suggested by the Revised Hierarchical Model were valid at a *textual* level, based on the results from both the eye-tracking and fMRI experiments, and this contributes to theoretical advances in T&I Studies.

7.4 Recommended Future Work

7.4.1 Extending the Languages Investigated

This research has studied only two language combinations: Mandarin - English and Spanish - English. The fact that very similar results regarding cognitive loading were reached in two such diverse cases strongly suggests that these results are potentially universal for all language combinations. However, some further work to confirm this is desirable. In particular, it would be interesting to know whether the difference in cognitive loading is reduced in the case of cognate language pairs which share similar vocabulary and grammar, such as two Bantu African languages, or in the case of Chinese dialects that share the same written form, e.g. Cantonese and Mandarin. Replications of the above studies on different language combinations would be potentially very useful.

In addition, a longitudinal study recording the different cognitive loading between A→B and B→A translation or interpreting can be interesting, as it could potentially test if training over time has an effect on cognitive loading. Such results can also help to inform translator and interpreter pedagogy.

7.4.2 Extending the Eye-tracking Work

The eye-tracking experiments have revealed that (1) A→B translation was more cognitively demanding than B→A translation; (2) the predictions suggested by the Revised Hierarchical Model were valid at a *textual* level; and (3) the additional interesting results and findings can potentially form a basis for formulating new research hypotheses to be tested by future investigations, as previously stated in **Section 7.3**.

In terms of additional interesting findings that can be used as new research hypotheses to be tested by future investigations, one example is that it has been suggested that translators paid more attention to the target text independently of translation directions. This has also been demonstrated by Pavlović and Jensen (2009). Future research with different language combinations, varying levels of text difficulties and lengths and different research populations investigating this suggestion is warranted.

In addition, the pupillometric results from the two eye-tracking experiments suggested that there was a lack of confirmation for the 'familiarity effect' during *written translation*. Whether or not this is a universal phenomenon for written translation with different language combinations, text types, text lengths and text difficulties, as well as within a different research population, has yet to be tested.

Further, given the non-head-mounted model of the monocular eye-tracker used in the study, blink frequency, a potentially important indicator of cognitive loading, was not accurately recorded. However, based on the result of the principal component analysis, blink frequency could be very powerful in terms of describing the variability of the dataset. Further work which employs a model in which blink frequency can be measured is thus desirable.

Given that the broader applicability of eye-tracking to T&I Studies has been demonstrated, the use of eye-tracking can be extended to investigate other research topics within the realm of T&I Studies. For example: (1) How do people visually process subtitles? (2) What is the general length of a 'meaning unit' in a source text for a translator to have to process before s/he starts to produce the target text? (3) During the very act of interpreting,

what do interpreters look at a conference venue to help to obtain non-text-related information to facilitate interpreting? (4) Where best to place a translated text in a video game, so as to attract the highest level of visual attention? The study presented in this thesis has laid a foundation that encourages future researchers to use eye-tracking to study topics associated with T&I.

Pupillometry and other ocular measures are only some of the physiological indicators for cognitive loading. Other physiological measures are heart rate, skin conductance, blood pressure, and neuroendocrine secretion. How to adequately apply these methods to the field of T&I Studies to *converge* different types of data in order to *further strengthen* results and findings is a question that has remained hugely under-researched. It is hoped that this doctoral study can open the eyes of T&I researchers to the physiological approach to studying the phenomena of translation and interpreting with a variety of physiological measures.

7.4.2.1 Potential Hypotheses for Future Testing

Based on the additional results and findings generated by the two eye-tracking experiments, a number of potential hypotheses or research questions warrant further investigation by future T&I researchers. For example: (1) Do novice translators tend to pay more visual attention to the target text, regardless of directionality? (2) Given that the 'familiarity effect' did not exist in the two eye-tracking experiments, is this phenomenon applicable to other language combinations, texts of a longer length, or texts of a higher level of difficulty? (3) What is the potential reason(s) behind the disappearance of the 'familiarity effect' toward the end of a written translation task? (4) Given the right Spanish keyboard, would typing Spanish be less cognitively demanding than typing English? (5) Can a head-mounted eye-tracker accurately capture the measurement of 'blink frequency' in order to employ it as a potentially useful dependent variable? The thesis presented here has helped to form a basis for formulating a number of more concrete hypotheses and questions to be studied by future researchers.

7.4.3 Extending the fMRI Work

The fMRI experiment has tentatively shown that (1) A→B was less cognitively demanding than B→A interpreting, due to a lower number of activated brain areas recruited; (2) even though the number of the participants was limited and hence the results were tentative, it has been indicated that the predictions suggested by the Revised Hierarchical Model were valid at a *textual* level; (3) A→B and B→A did share some similar cerebral activations, e.g. frontal and temporal activation, as well as parahippocampal gyrus activation (see **Section 6.18**); and (4) this is the 'first' study that shows activation of parahippocampal gyrus during textual interpreting. In addition, the study has further confirmed the importance of the left dorsolateral prefrontal cortex, which subserves verbal working memory (Klein *et al.*, 1995; Price *et al.*, 1999). Interpreting requires concurrent processing of multiple sources of information, and working memory has been shown to be crucial to interpreting performance (Gile, 1997; Liu, 2001). It has been demonstrated that working memory or memory span can be trained by a carefully selected task, for example, an ordinary individual is comfortable with remembering seven digits within a short period of time, which is dependent on working memory associated with the left dorsolateral prefrontal cortical connections; however, thanks to training, an individual can be trained to remember more than seven digits (Olesen *et al.*, 2004). If a task, for instance, a word span task, can be carefully designed and employed to help an interpreter establish stronger cortical connections in the left dorsolateral prefrontal cortex and retain more word-level information in the working memory during interpreting, this may help improve the interpreting skill or performance.

Again, the fMRI study involved only a very limited number of subjects. As a result, the conclusions reached must be regarded as preliminary and tentative. Enough has been done, however, to demonstrate the potential applicability of fMRI to T&I Studies. Further work using larger numbers of subjects may well produce fascinating insights into the neural pathways involved in translation or interpreting. This in turn may well shed light on the

intra-cerebral mechanisms responsible for the asymmetry in cognitive loading which this study has demonstrated. Given the limited number of subjects for the fMRI experiment in the study, it is advisable to carry out further research with more subjects in order to increase the statistical robustness and validity of the study. Even though it can be difficult to recruit participants that meet extremely specific criteria, e.g. translator or interpreters (novice or professional alike), future research should still strive to recruit them. A quick survey of published fMRI papers in a journal with a high impact factor usually included at least twelve subjects. Therefore, it would be desirable to increase the number of subjects to at least twelve for a more thorough fMRI study.

fMRI is only one of a number of techniques which are currently being developed for the study of brain processes. For instance, DTI (diffusion tensor imaging) can help to determine the tractography of the white matter tracts that connect activated gray matter areas.

Moreover, once gray matter activations and white matter connectivity are established by fMRI and DTI respectively, EEG can be applied to explore the time course of the gray matter activations. Thus, the combination of these three methods (fMRI, DTI and EEG) could provide a most powerful tool for investigating the structure of a translator's or interpreter's brain patterns.

By incorporating spectroscopy into MRI, a potentially higher concentration of a certain chemical compound, e.g. serotonin, required by a particular gray matter area for processing a certain cognitive task can be explored. Therefore, for example, based on *a priori* knowledge (Price *et al.*, 1999; Rinne *et al.*, 2000), it has been established that the left dorsolateral prefrontal cortex, subserving executive function (Ward, 2006: 285), is crucial to translating or interpreting. Spectroscopy is useful in exploring the level of chemical concentration at the left dorsolateral prefrontal cortex, in comparison with that at other cerebral sites during translating or interpreting.

Transcranial magnetic stimulation (TMS) is another powerful imaging technique. It involves the temporary disturbance of the neuronal activity of a particular cortical area

(using electromagnetic pulses) with an aim of establishing a causal link between the dysfunction of this area and the performing of a cognitive task. Since the dorsolateral prefrontal cortex has been shown to be vital in both A→B and B→A translation, TMS can be applied to see if temporarily disturbing this area causes a reduction in translation competence or efficiency.

7.5 Conclusion

This doctoral study set out to answer two main research questions: (1) whether or not eye-tracking has a broader applicability (and in the case of fMRI, a potential applicability) to the field of translation and interpreting studies; and (2) whether or not the predictions suggested by the Revised Hierarchical Model are valid at a *textual* level, based on the physiological and neurological evidence regarding the cognitive loading caused by different translation/sight interpreting directions at a textual level. These two questions have been answered firmly in the affirmative. It is hoped that this success will encourage further application of rigorous scientific methodology to T&I Studies.

References

ABUTALEBI, J., ANNONI, J. M., ZIMINE, I., PEGNA, A. J., SEGHIER, M. L., LEE-JAHNKE, H., LAZEYRAS, F., CAPPA, S. F. & KHATEB, A. (2008). Language control and lexical competition in bilinguals: an event-related FMRI study. *Cereb Cortex,* 18, 1496-505.

AGUIRRE, G. K. (Online). Retrieved at http://www.cfn.upenn.edu/fundamentals_fmri.htm, University of Pennsylvania, Accessed 15 August 2006.

AHRENS, B., KALDERON, E., REITH, W. & KRICK, C. (in press). fMRI for exploring simultaneous interpreting.

ALKIRE, M. T., HAIER, R. J., FALLON, J. H. & CAHILL, L. (1998). Hippocampal, but not amygdala, activity at encoding correlates with long-term, free recall of nonemotional information. *Proc Natl Acad Sci U S A,* 95, 14506-10.

ALTARRIBA, J. & MATHIS, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language,* 36, 550-568.

ALVES, F. (Ed.) (2003). *Triangulating Translation: Perspectives in process oriented research,* John Benjamins.

ALVES, F. & VALE, D. (forthcoming). Probing the unit of translation in time: factors on the design and development of a web application for storing, annotating, and querying translation process data. IN JÄÄSKELÄINEN, R. & GÖPFERICH, S. (Eds.) *Across Language and Cultures, vol 2/2009.*

ANDOH, J., ARTIGES, E., PALLIER, C., RIVIERE, D., MANGIN, J. F., PAILLERE-

MARTINOT, M. L. & MARTINOT, J. L. (2008). Priming frequencies of transcranial magnetic stimulation over Wernicke's area modulate word detection. *Cereb Cortex,* 18, 210-6.

ARNÁIZ UZQUIZA, V. (2008). Eye-tracking: Adopting scientific tools to improve subtitling practices, paper presented at *Multidisciplinary Approaches,* University of Montpellier 3, 21 June 2008 (personal communication).

ARTHURS, O. J. & BONIFACE, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends Neurosci,* 25, 27-31.

ASHBURNER, J., FLANDIN, G., HENSON, R., KIEBEL, S., KILNER, J., MATTOUT, J., PENNY, W. & STEPHAN, K. (2006). SPM5 Manual. http://www.fil.ion.ucl.ac.uk/spm/, London, UK, The Functional Imaging Laboratory Methods Group, University College London.

ASL (2006). *Eye Tracking System Instructions ASL Eye-Trac 6000,* Bedford, MA, USA, Applied Science Group, Inc.

ASSAF, M., CALHOUN, V. D., KUZU, C. H., KRAUT, M. A., RIVKIN, P. R., HART, J., JR. & PEARLSON, G. D. (2006). Neural correlates of the object-recall process in semantic memory. *Psychiatry Res,* 147, 115-26.

ATARI, O. (2005). Saudi students' translation strategies in an undergraduate translator training program. *Meta,* 50, 180-193.

BAKER, M. (1998). *Routledge encyclopedia of translation studies,* London and New York, Routledge.

BALLING, L. W. (2008). A brief introduction to regression designs and mixed-effects modelling by a recent convert. IN GÖPFERICH, S., JAKCOBSEN, A. L. &

MEES, I. M. (Eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing.* Copenhagen, Samfundslitteratur.

BARIK, H. C. (1969). A study of simultaneous interpretation. University of North Carolina.

BARTHOLOMEW, D. J., STEELE, F., MOUSTAKI, I. & GALBRAITH, J. I. (2002). *The Analysis and Interpretation of Multivariate Data for Social Scientists,* New York & Washington, DC, Chapman & Hall/CRC.

BEAR, M. F., CONNORS, B. W. & PARADISO, M. A. (2007). *Neuroscience: exploring the brain,* Philadelphia, Lippincott Williams & Wilkins.

BEATTY, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin,* 91, 276-292.

BEDNY, M. & THOMPSON-SCHILL, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain Lang,* 98, 127-39.

BIRDSONG, D. (1999). *Second Language Acquisition and the Critical Period Hypothesis,* Mahwah, NJ, Lawrence Erlbaum Asscociates.

BORBELY, K., NAGY, D., TOTH, M., DONAUER, N., VAJDA, J. & VARADY, P. (2005). Speech activation SPECT and plasticity of language networks. *J Neuroradiol,* 32, 345-7.

BOROWSKY, R., ESOPENKO, C., CUMMINE, J. & SARTY, G. E. (2007). Neural representations of visual words and objects: a functional MRI study on the modularity of reading and object processing. *Brain Topogr,* 20, 89-96.

BRADBURN, N. M., SUDMAN, S. & WANSINK, B. (2004). *Asking Questions: The Definitive Guide to Questionnaire Design -- For Market Research, Political Polls, and Social and Health Questionnaires,* Jossey-Bass.

BRAUER, J., NEUMANN, J. & FRIEDERICI, A. D. (2008). Temporal dynamics of perisylvian activation during language processing in children and adults. *Neuroimage,* 41, 1484-92.

BRITANNICA, E. (Online). Retrieved at http://cache.eb.com/eb/image?id=99532&rendTypeId=4, Accessed 1 March 2008.

BROWN, J. D. (1988). *Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research Design,* Cambridge University Press.

BRYSBAERT, M. & DIJKSTRA, T. (2006). Changing views on word recognition in bilinguals. IN MORAIS, J. & D'YDEWALLE, G. (Eds.) *Bilingualism and Second Language Acquisition.* Brussels, KVAB.

BUONOMANO, D. V. & MERZENICH, M. M. (1998). Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.,* 21, 149-186.

BUTTERWORTH, B. (1999). A head for figures. *Science,* 284, 928-9.

BUXTON, R. B. (2002). *Functional Magnetic Resonance Imaging: Principles and Techniques,* Cambridge, University Press.

CAFFREY, C. (2008). Using pupillometric, fixation-based and subjective measures to measure the processing effort experienced when viewing subtitled TV anime with pop-up gloss. IN GÖPFERICH, S., JAKOBSEN, A. L. & MEES, I. M. (Eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing.*

Copenhagen, Samfundslitteratur.

CAMPBELL, S. J. (1991). Towards a model of translation competence. *Meta,* 36, 329-343.

CAMPBELL, S. J. (2000). Choice Network Analysis in Translation Research. IN OLOHAN, M. (Ed.) *Intercultural Faultlines: Research Models in Translation Studies: Textual and Cognitive Aspects.* Manchester, St. Jeromes, pp. 29-42.

CAREW, J. (1978). *The Lost Love and Other Stories,* London, Longman Group Limited.

CARL, M., JAKOBSEN, A. L. & JENSEN, K. T. H. (2008). Studying Human Translation Behavior with User-Activity Data. In: Proceedings of NLPCS workshop at ICEIS. Barcelona 2008, 114-123.

CHANG, C. (2005). Directionality in Chinese/English simultaneous interpreting: impact on performance and strategy use. *Unpublished doctoral thesis.* University of Texas at Austin, Austin USA.

CHERNOV, G. (1979). Semantic aspects of psycholinguistics research in simultaneous interpretation. *Language and Speech,* 22, 277-295.

CHIRICOZZI, F. R., CLAUSI, S., MOLINARI, M. & LEGGIO, M. G. (2008). Phonological short-term store impairment after cerebellar lesion: a single case study. *Neuropsychologia,* 46, 1940-53.

CHRISTENSEN, T. A., ANTONUCCI, S. M., LOCKWOOD, J. L., KITTLESON, M. & PLANTE, E. (2008). Cortical and subcortical contributions to the attentive processing of speech. *Neuroreport,* 19, 1101-5.

CHRISTOFFELS, I. K. & DE GROOT, A. M. B. (2004). Components of simultaneous

interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition,* 7, 227-240.

CHRISTOFFELS, I. K., DE GROOT, A. M. B. & KROLL, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language,* 54, 324-345.

CHRISTOFFELS, I. K., DE GROOT, A. M. B. & WALDORP, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition,* 6, 201-211.

COGAIN (2008). http://www.cogain.org/glossary/. *Accessed 14 January 2008.*

COGAIN (Online). Retrieved at http://www.cogain.org/glossary/, Accessed 14 January 2008.

COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20, 37-46.

COHEN, J. (1992). A power primer. *Psychological Bulletin,* 112, 155-159.

COOK, V. (1996). *Second Language Learning and Teaching,* New York, Arnold.

COPENHAVER, B. R., RABIN, L. A., SAYKIN, A. J., ROTH, R. M., WISHART, H. A., FLASHMAN, L. A., SANTULLI, R. B., MCHUGH, T. L. & MAMOURIAN, A. C. (2006). The fornix and mammillary bodies in older adults with Alzheimer's disease, mild cognitive impairment, and cognitive complaints: a volumetric MRI study. *Psychiatry Res,* 147, 93-103.

CRAIG, G. (2007). *Repeated Measures ANOVA: ANOVA with Within-subject Factors (lecturer notes and slides),* London, University College London.

CSE (Online). Edinburgh Handedness Inventory, Retrieved at http://www.cse.yorku.ca/course_archive/2006-07/W/4441/EdinburghInventory.html. York University, Canada, Accessed 20 July 2007.

CSIC (Online). Retrieved at http://www.csic.cornell.edu/Elrod/t-test/two-normals.gif, Accessed 10 July 2008.

CULHAM, J. (Online). Basics of fMRI and the BOLD Response (Online lecture slides). Retrieved at http://psychology.uwo.ca/fmri4newbies/Tutorials.htm, London, Ontario Canada, University of Western Ontario, Accessed 11 August 2008.

CUMMINGS, E. (under review). Gender, emotion and pupillary response. *Cognition and Emotion*.

DANKS, J., SHREVE, G., FOUNTAIN, S. & MCBEATH, M. (Eds.) (1997). *Cognitive Processes in Translation and Interpreting*, Sage Publications.

DE BOT, K. (2000). Simultaneous interpreting as language production. IN E., D. B. & K., H. (Eds.) *Language processing and simultaneous interpreting.* Amsterdam, John Benjamins.

DE GROOT, A. M. B. (1992a). Bilingual lexical representation: a closer look at conceptual representations. IN FORST, R. & KATZ, L. (Eds.) *Orthography, Phonology, Morphology, and Meaning.* 339-412. Amsterdam, Elsevier.

DE GROOT, A. M. B. (1992b). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 18, 1001-1018.

DE GROOT, A. M. B. & CHRISTOFFELS, I. K. (2006). Language control in bilinguals:

Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition,* 9, 189-201.

DE GROOT, A. M. B., DANNENBURG, L. & VEN HELL, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language,* 33, 600-629.

DEB, A. (2003). EyeTrack: An application to record and play-back data from experiments using the ASL Model 504 Eye-Tracking system. *Unpublished MSc dissertation.* Imperial College London, London, UK.

DEMONET, J. F., CHOLLET, F., RAMSAY, S., CARDEBAT, D., NESPOULOUS, J. L., WISE, R., RASCOL, A. & FRACKOWIAK, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain,* 115 (Pt 6), 1753-68.

DEVLIN, J. T., JAMISON, H. L., GONNERMAN, L. M. & MATTHEWS, P. M. (2006). The role of the posterior fusiform gyrus in reading. *J Cogn Neurosci,* 18, 911-22.

DIAMOND, M., SCHEIBEL, A. B. & ELSON, L. M. (1985). *The Human Brain Coloring Book,* New York, N.Y., USA, HarperCollins.

DRAGSTED, B. & HANSEN, I. G. (2008). Comprehension and production in translation: a pilot study on segmentation and the coordiation of reading and writing processes. IN GÖPFERICH, S., JAKOBSEN, A. L. & MEES, I. M. (Eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing.* Copenhagen, Samfundslitteratur.

DRESEL, C., CASTROP, F., HASLINGER, B., WOHLSCHLAEGER, A. M., HENNENLOTTER, A. & CEBALLOS-BAUMANN, A. O. (2005). The

functional neuroanatomy of coordinated orofacial movements: sparse sampling fMRI of whistling. *Neuroimage,* 28, 588-97.

DRONKERS, N. F. (1996). A new brain region for coordinating speech articulation. *Nature,* 384, 159-161.

DSC (Online). Retrieved at http://www.spect.com/images/flanged-ect_2a.gif, Accessed 10 March 2008.

DUCHOWSKI, A. (2003). *Eye tracking methodology: theory and practice,* London, Springer Verlag.

EGGERS, G., RIEKER, M., KRESS, B., FIEBACH, J., DICKHAUS, H. & HASSFELD, S. (2005). Artefacts in magnetic resonance imaging caused by dental material. *Magma,* 18, 103-11.

ENGLUND-DIMITROVA, B. (2005). *Expertise and explicitation in the translation process,* Amsterdam.

ENGLUND-DIMITROVA, B. & HYLTENSTAM, K. (Eds.) (2000). *Language Processing and Simultaneous Interpreting: Interdisciplinary Perspectives* Amsterdam, John Benjamins Publishing.

EYE-TO-IT (Online). *Retrieved at http://cogs.nbu.bg/eye-to-it/.* Accessed 14 Feb. 2009.

FABBRO, F., GRAN, B. & GRAN, L. (1991). Hemispheric specialization for semantic and syntactic components of language in simultaneous interpreters. *Brain and Language,* 41, 1-42.

FANNING, D. W. (Online). Retrieved at http://www.dfanning.com/graphics_tips/mesh_mri_slices.jpg, Accessed 1

March 2008.

FIELD, A. (2005). *Discovering Statistics Using SPSS,* London, Sage Publications.

FRIDRIKSSON, J. & MORROW, L. (2005). Cortical activation and language task difficulty in aphasia. *Aphasiology,* 19, 239-250.

GANDOUR, J., WONG, D., DZEMIDZIC, M., LOWE, M., TONG, Y. & LI, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Hum Brain Mapp,* 18, 149-57.

GASS, S. M. & SELINKER, L. (2001). *Second language acquisition: An introductory course,* Mahwah, NJ, Lawrence Erlbaum.

GAZZANIGA, M. S., IVRY, R. B. & MANGUN, G. R. (2002). *Cognitive Neuroscience: The Biology of the Mind,* New York, W.W.Norton & Company, Inc.

GERVER, D. (1975). A psychological approach to simultaneous interpretation. *Meta,* 20, 119-128.

GILE, D. (1995). *Basic Concepts and Models for Interpreter and Translator Training,* Amsterdam, John Benjamins Publishing Co.

GILE, D. (1997). Conference interpreting as a cognitive management problem. IN DANKS, H. J., SHREVE, G. M., FOUNTAIN, S. B. & MCBEATH, M. K. (Eds.) *Cognitive processes in translation and interpreting.* Thousand Oaks, CA, Sage.

GOLDMAN-EISLER, F. (1974). An experimental study of interference between receptive and productive processes relating to simultaneous translation. *Language and Speech,* 17, 1-10.

GOODWIN, C. J. (1995). *Research in Psychology: Methods and Design,* New York, John

Wiley & Sons, Inc.

GRABE, W. (1991). Current developments in second language reading research. *TESOL Quarterly,* 25, 375-405.

GRABNER, R. H., BRUNNER, C., LEEB, R., NEUPER, C. & PFURTSCHELLER, G. (2007). Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Res Bull,* 72, 57-65.

GRACCO, V. L., TREMBLAY, P. & PIKE, B. (2005). Imaging speech production using fMRI. *Neuroimage,* 26, 294-301.

GRAN, L. & FABBRO, F. (1988). The role of neuroscience in the teaching of interpretation. *The Interpreters' Newsletter,* 1, 23-41.

HALL, D. A., HAGGARD, M. P., AKEROYD, M. A., PALMER, A. R., SUMMERFIELD, A. Q., ELLIOTT, M. R., GURNEY, E. M. & BOWTELL, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp,* 7, 213-23.

HANSEN, G. (2002). *Empirical Translation Studies: Process and Product,* Samfundslitteratur.

HARLEY, T. (2001). *The Psychology of Language,* Hove & New York, Psychological Press.

HARMS, M. P. & MELCHER, J. R. (2002). Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol,* 88, 1433-50.

HASEGAWA, M., CARPENTER, P., A. & JUST, M. A. (2002). An fMRI study on

bilingual sentence comprehension and workload. *Neuroimage,* 15, 647-660.

HEIM, S., AMUNTS, K., MOHLBERG, H., WILMS, M. & FRIEDERICI, A. D. (2006). Head motion during overt language production in functional magnetic resonance imaging. *Neuroreport,* 17, 579-82.

HERBIE, D. (Online). Retrieved at http://drherbie.files.wordpress.com/2006/10/WindowsLiveWriter/NormalDistributionPartII_D10A/same%20mean%20different%20stddev%5B2%5D.gif, Accessed 10 July 2008.

HEREDIA, R. R. (1996). Bilingual memory: A re-revised version of the hierarchical model of bilingual memory. *The Newsletter of the Center For Research in Language,* 10, 3-6.

HOPFINGER, J. B., BUONOCORE, M. H. & MANGUN, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat Neurosci,* 3, 284-91.

HOSHINO, N. & KROLL, J. F. (2007). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition,* 106, 501-511.

HOWSTUFFWORKS (Online). Retrieved at http://static.howstuffworks.com/gif/adam/images/en/brain-picture.jpg, Accessed 1 March 2008.

HUETTEL, S. A., SONG, A. W. & MCCARTHY, G. (2003). *Functonal Magnetic Resonance Imaging,* Sunderland, MA, USA, Sinauer Associates, Inc.

HYÖNÄ, J., TOMMOLA, J. & ALAJA, A. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology,* 48A, 598-612.

ICREC (Online). Available at http://www3.imperial.ac.uk/research/researchethics, Accessed 1 February 2006.

JÄÄSKELÄINEN, R. (1989). Translation Assignment in Professional vs. Non-professional Translation: A Think-aloud Protocol Study. IN SÉGUINOT, C. (Ed.) *The Translation Process.* Toronto, York University School of Translation.

JÄÄSKELÄINEN, R. (2002). Think-aloud protocol studies into translation: An annotated bibliography. *Target,* 14, 107-136.

JÄÄSKELÄINEN, R. (2005). Triangulating translation: Perspectives in process oriented research. *Target,* 17, 161-167.

JÄÄSKELÄINEN, R. & TIRKKONEN-CONDIT, S. (1991). Automatised Processes in Professional vs. Non-professional Translation: A Think-Aloud Protocol Study. IN TIRKKONEN-CONDIT, S. (Ed.) *Empirical Research in Translation and Intercultural Studies.* Tübingen, Gunter Narr.

JAKCOBSEN, A. L. & JENSEN, K. T. H. (2008). Eye movement behaviour across four different types of reading task. IN GÖPFERICH, S., JAKOBSEN, A. L. & MEES, I. M. (Eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing.* Copenhagen, Samfundslitteratur.

JAKOBSEN, A. L. (2006). Research methods in translation - Translog. IN SULLIVAN, K. P. H. & LINDGREN, E. (Eds.) *Computer Keystroke Logging and Writing (pp. 95-105).* Amsterdam, Elsevier.

JANISSE, M. P. (1977). *Pupillometry: The Psychology of the Pupillary Response,* Washington, Hemisphere Publishing.

JANYAN, A., GERGANOV, A., STAMENOV, M., POPIVANOV, I. & OSENOVA, P.

(forthcoming). The role of word order and definite article form in relative clause processing in Bulgarian: an eye movement study.

JENSEN, C. (2008). Assessing eye-tracking accuracy in translation studies. IN GÖPFERICH, S., JAKCOBSEN, A. L. & MEES, I. M. (Eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing.* Copenhagen, Samfundslitteratur.

JEONG, H., SUGIURA, M., SASSA, Y., YOKOYAMA, S., HORIE, K., SATO, S., TAIRA, M. & KAWASHIMA, R. (2007). Cross-language influence on brain activation during second language processing: An fMRI study. *Bilingualism: Language and Cognition,* 10, 175-187.

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

JOHNSON, J. S. & NEWPORT, E. L. (1991). Critical period effects on universal properties of language: the status of subjacency in the acquisition of a second language. *Cognition,* 39, 215-58.

KAHNEMAN, D. (1973). *Attention and effort,* Eaglewood Cliffs, NJ, Prentice-Hall.

KETCHAM, C. J., HODGSON, T. L., KENNARD, C. & STELMACH, G. E. (2003). Memory-motor transformations are impaired in Parkinson's disease. *Exp Brain Res,* 149, 30-9.

KETTELER, D., KASTRAU, F., VOHN, R. & HUBER, W. (2008). The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain; an fMRI study. *Neuroimage,* 39, 2002-9.

KIRSCHEN, M. P., CHEN, S. H., SCHRAEDLEY-DESMOND, P. & DESMOND, J. E.

(2005). Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *Neuroimage,* 24, 462-72.

KLEIN, D., MILNER, B., ZATORRE, R. J., MEYER, E. & EVANS, A. C. (1995). The neural substrates underlying word generation: a bilingual functional-imaging study. *Proc Natl Acad Sci U S A,* 92, 2899-903.

KNECHT, S., DRÄGER, B., DEPPE, M., BOBE, L., LOHMANN, H., FLÖEL, A., RINGELSTEIN, E.-B. & HENNINGSEN, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain,* 123, 2512-2518.

KRICHMAR, J. L., THOMAS, M. L., ROWLAND, L., SING, H. C. & THORNE, D. R. (1997). Pupil dynamics and eye movements as indicators of fatigue and sleepiness. *Invest Ophthalmol Vis Sci (Suppl),* 38:S71.

KRICK, C., REITH, W., BEHRENT, S. & FRANCESCHINI, R. (2006). Interpreters' brain function: a study sheds new light on interpreters' grey matter. *Bulletin of the Institute of Translation and Interpreting.*

KRING, H. P. (2001). *Repairing texts: empirical investigations of machine translation post-editing processes,* Kent, Ohio, Kent State Univ.P.

KROLL, J. F. (2009). *Re: VERY URGENT: A quick question please,* Personal email to C.-Y. Chang (chc131@mail.harvard.edu).

KROLL, J. F., DUFOUR, R., MICHAEL, E. & TOKOWICZ, N. (1998). *Bilingual Word Naming and Translation: The Role of Experience and Cognitive Skill in Developing Proficiency in a Second Language,* University Park, PA, The Pennsylvania State University.

KROLL, J. F., MICHAEL, E., TOKOWICZ, N. & DUFOUR, R. (2002). The

development of lexical fluency in a second language. *Second Language Research,* 18, 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, 149-174.

KROLL, J. F. & TOKOWICZ, N. (2001). The development of conceptual representation for words in a second language. IN NICOL, J. (Ed.) *One Mind, Two Languages: Bilingual Language Processing.* Malden, MA, Blackwell Publishers.

KURZ, I. (1994). A look into the 'Black Box'- EEG probability mapping during mental simultaneous interpreting. IN SNELL-HORNBY, M., POCHHACKER, F. & KAINDL, K. (Eds.)
Translation Studies- An Interdiscipline. Amsterdam and Philadelphia, John Benjamins.

LAI, C. S. (2006). *Chinese Mandarin: Second Semester of the Fourth Year,* Taipei, Taiwan, Kang Hsuan Publishing (in Chinese Mandarin).

LANDIS, J. R. & KOCH, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.

LAUTERBUR, P. C. (1973). Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *NatureProceedings,* 242, 190-191.

LAW, I. (2007). Human brain mapping under increasing cognitive complexity using regional cerebral blood flow measurements and positron emission tomography. *Dan Med Bull,* 54, 289-305.

LEE, T.-H. (1999). Speech proportion and accuracy in simultaneous interpretation from English into Korean. *Meta,* 44.

LEHTONEN, M. H., LAINE, M., NIEMI, J., THOMSON, T., VOROBYEV, V. A. & HUGDAHL, K. (2005). Brain correlates of sentence translation in Finnish-Norwegian bilinguals. *NeuroReport,* 16, 607-610.

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

LEVÝ, J. (2000). Translation as a decision process. IN VENUTI, L. & BAKER, M. (Eds.) *The Translation Studies Reader.* London, Routledge (pp. 149-159).

LI, P. & GREEN, D. W. (2007). Neurocognitive approaches to bilingualism: Asian languages. *Bilingualism: Language and Cognition,* 10, 117-119.

LIN, Y. H. (Online). Words of My Mother (in Mandarin Chinese). Retrieved at http://write.idv.tw/old/doc/b/b018.doc, Accessed 2 June 2006.

LIU, C., ZHANG, W. T., TANG, Y. Y., MAI, X. Q., CHEN, H. C., TARDIF, T. & LUO, Y. J. (2008). The Visual Word Form Area: evidence from an fMRI study of implicit processing of Chinese characters. *Neuroimage,* 40, 1350-61.

LIU, M. (2001). Expertise in simultaneous interpreting: A working memory analysis. The University of Texas at Austin.

LONSDALE, A. B. (1998). Directionality of translation (directionality). IN BAKER, M. & MALMKAER, K. (Eds.) *Routledge Encyclopedia of Translation Studies.* London & New York, Routledge.

LORBERBAUM, J. P., KOSE, S., JOHNSON, M. R., ARANA, G. W., SULLIVAN, L. K., HAMNER, M. B., BALLENGER, J. C., LYDIARD, R. B., BRODRICK, P. S.,

BOHNING, D. E. & GEORGE, M. S. (2004). Neural correlates of speech anticipatory anxiety in generalized social phobia. *Neuroreport,* 15, 2701-5.

LÖRSCHER, W. (1989). Models of the translation process: Claims and reality. *Target,* 1, 43-68.

LÖRSCHER, W. (1992). Process-oriented research into translation and implications for translation teaching. *TTR,* 5, 145-161.

LUCK, S. J. (2005). *An Introduction to the Event-Related Potential Technique,* Cambridge, MA, The MIT Press.

MAJARANTA, P. (Online). Implementation of New Interaction Techniques: Eye Tracking (online lecture slides). Retrieved at http://www.cs.uta.fi/init/slides/INIT-2008-EyeTracking-01.pdf, Visual Interaction Research Group, Tampere Unit for Computer-Human Interaction, Accessed 13 August 2008.

MALKIEL, B. (2004). Directionality and translational difficulty. *Perspectives: Studies in Translatology,* 12, 208-221.

MANDER, B. A., REID, K. J., DAVULURI, V. K., SMALL, D. M., PARRISH, T. B., MESULAM, M. M., ZEE, P. C. & GITELMAN, D. R. (2008). Sleep deprivation alters functioning within the neural network underlying the covert orienting of attention. *Brain Res,* 1217, 148-56.

MARTIN, J. H. (2003). *Neuroanatomy Text and Atlas,* McGraw-Hill.

MCILWAIN, H. & BACHELARD, H. S. (1985). *Biochemistry and the Central Nervous System,* Edinburgh, Churchill Livingstone.

MCLAUGHLIN, B. (1978). *Second-language acquisition in childhood,* Hillsdale, NJ, Lawrence Erlbaum.

MECHELLI, A., CRINION, J. T., NOPPENEY, U., O'DOHERTY, J., ASHBURNER, J., FRACKOWIAK, R. S. & PRICE, C. J. (2004). Structural plasticity in the bilingual brain. *Nature,* 431, 757.

MECHELLI, A., JOSEPHS, O., LAMBON RALPH, M. A., MCCLELLAND, J. L. & PRICE, C. J. (2007). Dissociating stimulus-driven semantic and phonological effect during reading and naming. *Hum Brain Mapp,* 28, 205-17.

MEHLER, J., CHRISTOPHE, A. & RAMUS, F. (2000). What we know about the initial state for language. IN MARANTZ, A., MIYASHITA, Y. & O'NEIL, W. (Eds.) *Image, Language, 14 Pierre-Yves Oudeyer Brain: Papers from the first Mind-Brain Articulation Project symposium.* Cambridge, MA USA, MIT Press.

MERRILL, R. T. & MCELHINNY, M. W. (1983). *The Earth's magnetic field: Its history, origin and planetary perspective,* London and New York, Academic Press.

MIALDEA, A. (2007). Re: textos en español. *Sent from a.mialdea@imperial.ac.uk.* Personal email to V.C.-Y. Chang (cy.chang05@imperial.ac.uk).

MILNE, J. (1980). *The Long Tunnel,* London, Heinemann Educational Books, Ltd.

MORIMOTO, C. H., KOONS, D., AMIR, A. & FLICKNER, M. (2000). Pupil detection and tracking using multiple light sources. *Image and Vision Computing,* 4, 331-335.

MOSER-MERCER, B. (1997). Skill components in simultaneous interpreting. IN GAMBIER, Y., DANIEL, G. & TAYLOR, C. (Eds.) *Conference Interpreting: Current Trends in Research.* Amsterdam, John Benjamins, pp. 133-148.

NEI (Online). Retrieved at http://www.nei.nih.gov/health/eyediagram/images/NEA08.gif. *Accessed 22 April 2009.*

NIEHAUS, L., GULDIN, B. & MEYER, B. (2001). Influence of transcranial magnetic stimulation on pupil size. *Journal of the Neurological Sciences,* 182, 123-128.

O'BRIEN, S. (2005). Methodologies for measuring the correlations between post-editing effort and machine translatability. *Machine Translation,* 19, 37-58.

O'BRIEN, S. (2006a). Eye-tracking and translation memory matches. *Perspectives: Studies in Translatology,* 14, 185-205.

O'BRIEN, S. (2006b). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures,* 7, 1-21.

OLESEN, P. J., WESTERBERG, H. & KLINGBERG, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nat Neurosci,* 7, 75-79.

OLIVER, S. (2007). *Re. Please Could You Advise?,* Personal email to V.C.-Y. Chang (cy.chang05@imperial.ac.uk).

ORDEN, K. F. V., LIMBERT, W., MAKEIG, S. & JUNG, T.-P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human factors: The Journal of the Human Factors and Ergonomics Society,* 43, 111-121.

PACTE (2000). Acquiring Translation Competence: Hypotheses and Methodological Problems in a Research Project. IN BEEBY, A., ENSINGER, D. & PRESAS, M. (Eds.) *Investigating Translation.* Amsterdam, John Benjamins, pp. 99-106.

PADILLA, F., BAJO, M. T. & MACIZO, P. (2005). Articulatory suppression in

language interpretation: Working memory capacity, dual tasking and word knowledge. *Bilingualism: Language and Cognition,* 8, 207-219.

PARK, J. W., KIM, S. H., KIM, Y. W., KIM, J. Y., PARK, S. Y., SON, S. M., BAI, D. S. & JANG, S. H. (2008). Motor control via spared peri-infarct corticospinal tract in patients with pontine infarct. *J Comput Assist Tomogr,* 32, 159-62.

PAVLOVIĆ, N. (2007). Directionality in Collaborative Translation Process: A Study of Novice Translators. *Unpublished doctoral thesis.* Universitat Rovira i Virgili, Tarragona Spain.

PAVLOVIĆ, N. & JENSEN, K. T. H. (2009). Eye tracking translation directionality. IN PYM, A. & PEREKRESTENKO, A. (Eds.) *Translation Research Projects 2.* Tarragona, Intercultural Studies Group.

PETERSEN, S. E., FOX, P. T., POSNER, M. I., MINTUN, M. & RAICHLE, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature,* 331, 585-9.

PETSCHE, H. & ETLINGER, S. C. (1998). *EEG and thinking. Power and coherence analysis of cognitive processes,* Vienna, Verlag der Osterreichischen Akademie der Wissenschaften.

PETSCHE, H., ETLINGER, S. C. & FILZ, O. (1993). Brain electrical mechanisms of bilingual speech management: an initial investigation. *Electroencephalogr Clin Neurophysiol,* 86, 385-94.

PHAN, K. L., ANGSTADT, M., GOLDEN, J., ONYEWUENYI, I., POPOVSKA, A. & DE WIT, H. (2008). Cannabinoid modulation of amygdala reactivity to social signals of threat in humans. *J Neurosci,* 28, 2313-9.

PHILLIPS, N. A., KLEIN, D., MERCIER, J. & DE BOYSSON, C. (2006). ERP measures of auditory word repetition and translation priming in bilinguals. *Brain Res,* 1125, 116-31.

PÖCHHACKER, F. (2004). *Introducing Interpreting Studies,* London, Routledge.

PREUSCHOFF, K., QUARTZ, S. R. & BOSSAERTS, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *J Neurosci,* 28, 2745-52.

PRICE, C. J. & FRISTON, K. J. (1997). Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage,* 5, 261-70.

PRICE, C. J., GREEN, D. W. & VON STUDNITZ, R. (1999). A functional imaging study of translation and language switching. *Brain,* 122 (Pt 12), 2221-35.

PRICE, C. J., WISE, R. J., WARBURTON, E. A., MOORE, C. J., HOWARD, D. & PATTERSON, K. (1996). Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain,* 119.

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, 1029-1038(10).

PU, Y., LIU, H. L., SPINKS, J. A., MAHANKALI, S., XIONG, J., FENG, C. M., TAN, L. H., FOX, P. T. & GAO, J. H. (2001). Cerebral hemodynamic response in Chinese (first) and English (second) language processing revealed by event-related functional MRI. *Magn Reson Imaging,* 19, 643-7.

QUILODRAN, R., ROTHE, M. & PROCYK, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron,* 57, 314-25.

RADACH, R. & KENNEDY, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. *European Journal of Cognitive Psychology,* 16, 3-26.

RAYNER, K. (1998). Eye movements and information processing: 20 years of research. *Psychological Bulletin,* 124, 372-422.

RAYNER, K., K.H., C., SLATTERY, T. J. & ASHBY, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading,* 10, 241-255.

RIECKER, A., ACKERMANN, H., WILDGRUBER, D., DOGIL, G. & GRODD, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport,* 11, 1997-2000.

RINNE, J. O., TOMMOLA, J., LAINE, M., KRAUSE, B. J., SCHMIDT, D., KAASINEN, V., TERAS, M., SIPILA, H. & SUNNARI, M. (2000). The translating brain: cerebral activation patterns during simultaneous interpreting. *Neurosci Lett,* 294, 85-8.

RINNE, T., STECKER, G. C., KANG, X., YUND, E. W., HERRON, T. J. & WOODS, D. L. (2007). Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *Neuroreport,* 18, 1311-4.

ROTA, G., VEIT, R., NARDO, D., WEISKOPF, N., BIRBAUMER, N. & DOGIL, G. (2008). Processing of inconsistent emotional information: an fMRI study. *Exp Brain Res,* 186, 401-7.

SABRI, M., BINDER, J. R., DESAI, R., MEDLER, D. A., LEITL, M. D. & LIEBENTHAL, E. (2008). Attentional and linguistic interactions in speech

perception. *Neuroimage,* 39, 1444-56.

SAVOY, R. L. (2002). Functional magnetic resonance imaging fMRI. IN RAMACHANDRAN, V. (Ed.) *Encyclopedia of the brain.* San Diego, Academic Press.

SCHICK, T. (2000). *Readings in the Philosophy of Science: From Positivism to Postmodernism,* Mountain View, California, Mayfield Publishing Company.

SCHILPEROORD, J. & SANDERS, T. (1997). Pauses, cognitive rhythms and discourse structure: An empirical study of discourse production. *Amsterdam Studies in the Theory and History of Linguistics Sciences Series 4,* 247-267.

SCHLESINGER, M. (2000). Interpreting as a cognitive process: How can we know what really happens? IN TIRKKONEN-CONDIT, S. & JÄÄSKELÄINEN, R. (Eds.) *Tapping and mapping the processes of translation and interpreting: Outlooks on empirical research.* Amsterdam, The Netherlands, John Benjamins Publishing Company.

SCHWARTZ, A. I., KROLL, J. F. & DIAZ, M. (2007). Reading words in Spanish and English: mapping orthography to phonology in two languages. *Language and Cognitive Processes,* 22, 106-129.

SCHWEIGERT, W. A. (1994). *Research Methods and Statistics for Psychology,* Pacific Grove, California, Brooks/Cole Publishing Company.

SÉGUINOT, C. (Ed.) (1989). *The Translation Process,* Toronto, H. G. Publications.

SETTON, R. (1999). *Simultaneous Interpretation: A Cognitive-Pragmatic Analysis,* Amsterdam/Philadelphia, John Bejamins.

SETTON, R. (2002). Meaning assembly in simultaneous interpretation. IN PÖCHHACKER, F. & SHLESINGER, M. (Eds.) *The Interpreting Studies Reader.* London, Routledge.

SHARMIN, S., ŠPAKOV, O., RÄIHÄ, K.-J. & JAKOBSEN, A. L. (2008). Effects of time pressure and text complexity on translators' fixations. IN RÄIHÄ, K.-J. & DUCHOWSKI, A. (Eds.) *Proceedings of the Eye Tracking Research & Application Symposium, ETRA 2008.* Savannah, Georgia, USA, pp. 123-126, ACM.

SHAUGHNESSY, J. J., ZECHMEISTER, E. B. & ZECHMEISTER, J. S. (2000). *Research Methods in Psychology,* New York, McGraw-Hill Higher Education.

SHELLOCK, F. G. & SHELLOCK, V. J. (1998). Cardiovascular catheters and accessories: ex vivo testing of ferromagnetism, heating, and artifacts associated with MRI. *J Magn Reson Imaging,* 8, 1338-42.

SHOLL, A., SANKARANARAYANAN, A. & KROLL, J. F. (1995). Transfer between picture naming and translation: a test of asymmetries in bilingual memory. *Psychological Science,* 6, 45-49.

SJØRUP, A. C. (2008). Metaphor comprehension in translation: methodological issues in a pilot study. IN GÖPFERICH, S., JAKOBSEN, A. L. & MEES, I. M. (Eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing.* Copenhagen, Samfundslitteratur.

SLOWINSKI, J. (Online). Retrieved at http://bowlingknowledge.info/index.php?option=com_content&task=view&id=128&Itemid=1, Accessed 20 June 2007.

SMITH, C. J. (2007). Faster isn't necessarily better: the role of individual differences

in processing words with multiple translations. *Unpublished BPhil dissertation.* University of Pittsburgh, Pittsburgh USA.

SNELL-HORNBY, M. (1991). Translation Studies - Art, Science or Utopia? IN VAN LEUVEN-ZWART, K. M. & NAAIJKENS, T. (Eds.) *Translation Studies: The State of the Art (pp. 13-23).* Amsterdam & Atlanta, GA, Rodopi.

SONG, L. P., PENG, D. L., JIN, Z., YAO, L., NING, N., GUO, X. J. & ZHANG, T. (2007). [Gray matter abnormalities in developmental stuttering determined with voxel-based morphometry]. *Zhonghua Yi Xue Za Zhi,* 87, 2884-8.

SPM (2006). DVD. Filmed by Peter Aston, London, UK, University College London.

STAWARSKI, C. & PHILLIPS, P. P. (2008). *Data Collection: Planning for and Collecting All Types of Data,* Pfeiffer

STEINSTRÄTER, O. (Online). Retrieved at http://wwwneuro03.uni-muenster.de/ger/t2tconv/conv3d.html. *Accessed 17 August 2008.*

SUNDERMAN, G. & KROLL, J. F. (2006). First language activation during second language lexical processing. *Studies in Second Language Acquisition,* 28, 387-422.

TALAIRACH, J. & TOURNOUX, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain,* Suttgart, Thieme.

THULBORN, K. R., WATERTON, J. C., MATTHEWS, P. M. & RADDA, G. K. (1982). Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochim Biophys Acta,* 714, 265-270.

TIRKKONEN-CONDIT, S. & JÄÄSKELÄINEN, R. (Eds.) (2000). *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on Empirical Research,*

Amsterdam & Philadelphia, John Benjamins Publishing Co.

TOKOWICZ, N. & KROLL, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes,* 22, 727-779(53).

TOLMAN, C. W. (1992). *Positivism in Psychology: Historical and Contemporary Problems,* New York, Springer-Verlag.

TOMMOLA, J. (1999). New trends in interpreting research: Going psycho - or neuro (pp. 321-330). IN ÁLVAREZ, L. & OCAMPO, F. (Eds.) *Anovar-anosar, estudios de traducción e interpretación.*

TOMMOLA, J. & NIEMI, P. (1986). Mental load in simultaneous interpreting: An on-line pilot study. IN EVENSEN, L. (Ed.) *Nordic research in text linguistics and discourse analysis (pp. 171-184).* Trondheim, Tapir.

TORTORA, G. J. & ANAGNOSTAKOS, N. P. (1990). *Principles of Anatomy and Physiology,* New York, Harper Collins.

UNVERSE-REVIEW (Online). Retrieved at http://universe-review.ca/I10-41-limbic.jpg, Accessed 1 March 2008.

VENNERI, A., MCGEOWN, W. J., HIETANEN, H. M., GUERRINI, C., ELLIS, A. W. & SHANKS, M. F. (2008). The anatomical bases of semantic retrieval deficits in early Alzheimer's disease. *Neuropsychologia,* 46, 497-510.

VIRTUE, S., PARRISH, T. & JUNG-BEEMAN, M. (2008). Inferences during Story Comprehension: Cortical Recruitment Affected by Predictability of Events and Working Memory Capacity. *J Cogn Neurosci.*

WANG, J., ZHAI, S. & SU, H. (2001). Chinese input with keyboard and eye-tracking-An anatomical study. *Conference on Human Factors in Computing Systems: Proceedings of the SIGCHI conference on Human factors in computing systems.* Seattle, Washington, United States ACM Press, pp. 349-356.

WANG, S., ZHU, Z., ZHANG, J. X., WANG, Z., XIAO, Z., XIANG, H. & CHEN, H. C. (2008). Broca's area plays a role in syntactic processing during Chinese reading comprehension. *Neuropsychologia,* 46, 1371-8.

WARD, J. (2006). *The Student's Guide to Cognitive Neuroscience,* Hove and New York, Psychological Press, Taylor and Francis Group.

WEATHERBY, J. (1998). Teaching Translation into L2. A TT- Oriented Approach. IN MALMKJAER, K. (Ed.) *Translation and Language Teaching.* Manchester, St. Jerome.

WEINREICH, U. (1968). *Languages in Contact,* The Hague, Mouton.

WELLESLEY, C. (Online). Retrieved at http://cs.wellesley.edu/~cs112/assignments/assign5/slice.gif, Accessed 11 July 2009.

WHATMOUGH, C. & CHERTKOW, H. (2007). rCBF to the hippocampal complex covaries with superior semantic memory retrieval. *Behav Brain Res,* 181, 262-9.

WHELAN, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia *Educational Research Review* 2, 1-12.

WILLIAMS, J. & CHESTERMAN, A. (2002). *The map: A beginner's guide to doing research in translation studies,* Manchester, St. Jerome.

WILSS, W. (1996). *Knowledge and skills in translator behavior,* Amsterdam & Philadelphia, John Benjamins.

WISE, R., CHOLLET, F., HADAR, U., FRISTON, K., HOFFNER, E. & FRACKOWIAK, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain,* 114 (Pt 4), 1803-17.

WISE, R. J. S., GREEN, J., BUCHEL, C. & SCOTT, S. K. (1999). Brain regions involved in articulation. *Lancet,* 353, 1057-1061.

WOTAWA, O. (2006). Système visuel cortical de bas niveau et perception du mouvement: un étude par IRM. *Unpublished doctoral thesis.* Université Nice-Sophia Antipolis - UFR Sciences.

WU, C. M. & LIN, Y. S. (2008). Neural correlates on Chinese speech production: an fMRI study. *J Acoust Soc Am,* 123, 3075.

XIAO, L., BECHARA, A., CEN, S., GRENARD, J. L., STACY, A. W., GALLAHER, P., WEI, Y., JIA, Y. & ANDERSON JOHNSON, C. (2008). Affective decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in 10th-grade Chinese adolescent smokers. *Nicotine Tob Res,* 10, 1085-97.

XIAO, Y. W. (Online). Shelly's Adventure (in Mandarin Chinese). Retrieved at http://write.idv.tw/old/novel/n019.doc, Accessed 2 June 2006.

YABE, S., YAMASHITA, M., MOMOSE, S., TAHIRA, K., YOSHIDA, S., LI, R., YIN, S. & SATO, T. (2001). Synthesis and UV-shielding properties of metal oxide doped ceria via soft solution chemical processes. *International Journal of Inorganic Materials,* 3, 1003-1008.

Appendix A: English Text for Translation

At universities in England, May is the month of exams. A student is lying on the grass in front of the examination hall. His two friends are sitting near him. The three of them are first year students at university. All three of them are sitting their first year exams.

Appendix B: English Text for Typing

I live in a house in central city. My house has a garden. I like my garden very much, because my garden has a lot of flowers and the grass is very green. The flowers are very pretty. I like to sit in my garden when it is sunny outside.

Appendix C: Mandarin Text for Translation

<Traditional Mandarin>

有一個人，他很不滿意自己現在的樣子，於是到處去拜訪成功的人，請教他們改變自己的方法。他聽到了許多成功的故事，發現這些人成功的方法，都是一樣的。那就是：不斷的學習和創造，生命就會不斷的改變。

<Simplified Mandarin>

有一个人，他很不满意自己现在的样子，于是到处去拜访成功的人，请教他们改变自己的方法。他听到了许多成功的故事，发现这些人成功的方法，都是一样的。那就是：不断的学习和创造，生命就会不断的改变。

English Back Translation:

There was someone, who was not happy with where he is in life. Therefore, he started visiting successful people and asking them how to change himself. He heard many success stories and discovered that the way in which those people succeeded was the same. That is: If a person continuously learns and creates, then life will change continuously.

Appendix D: Mandarin Text for Typing

<Traditional Mandarin>

你如果想要試圖改變自己，在這一本書籍裡，你都可以找到答案。因爲這一本書，不只是告訴你書裡的知識，同時還告訴你：世界是怎麼改變的；如何運用創意；如何享受生活的樂趣；還有，如何設計有趣的活動。

<Simplified Mandarin>

你如果想要试图改变自己，在这一本书籍里，你都可以找到答案。因为这一本书，不只是告诉你书里的知识，同时还告诉你：世界是怎么改变的；如何运用创意；如何享受生活的乐趣；还有，如何设计有趣的活动。

English Back Translation:

If you want to try and change yourself, you can get all the answers from this book. It is because this book not only imparts knowledge, but also tells you how the world changes, how to use creativity, how to enjoy life and how to design interesting activities.

Appendix E: Spanish Text for Translation and Reading

La niña era mucho menor que sus hermanos, y sin duda por eso la cuidaban y mimaban tanto sus padres. Ella había nacido cuando sus padres no esperaban ya tener más hijos. Sus padres no la querían más que a los otros, pero habían sido más estrictos con sus hermanos.

English Back Translation:

The child was much younger than her brothers, and no doubt because of that her parents cared for and pampered her so much. She was born when her parents did not expect to have any more children. Her parents did not love her more than her brothers but they had been stricter with her brothers than with her.

Appendix F: Spanish Text for Typing

El niño tiene seis años. Es bueno y ya sabe leer y va a la escuela. Su padre es carpintero y su hijo ya empieza a serle útil llevándole las herramientas que le pide. Durante los fines de semana, también ayuda a su madre en las tareas de la casa.

English Back Translation:

The child is six years old. He's a clever child, already knows how to read and goes to school. His father is a carpenter and the son already helps him by bringing the tools his father asks for. During weekends, he also helps his mother with the housework.

Appendix G: Consent Form for the Mandarin – English Eye-tracking Experiment

Consent Form for Translators: Eye-tracking Study

You are invited to participate in a study of translation. My name is Chieh-ying Chang, a Ph.D. student at the Department of Humanities of Imperial College London. This study is part of the thesis research for the completion of a Ph.D. degree. I hope to investigate how translation directionality relates to cognitive effort during translating. You were selected as a potential participant because you are a full-time translation student with the language pair of Mandarin and English. You will be one of the students in translation- or interpretation-related departments in universities or colleges who have been invited to participate in this study.

If you decide to participate, you will be asked to fill out a questionnaire on your demographic information, your IELTS score, age of second language acquisition, handedness, your translation training, prior educational background and work experience. You will be asked to conduct translation of two 50-word non-technical texts, one into English and the other into Mandarin, as well as two typing tasks respectively in English and Mandarin. You will be asked to sit in front of a computer, with a non-invasive infrared eye-tracking machine recording your eye movement and pupil dilation. The whole process will take 90 minutes at most. When you have finished the translation, you will be given a cash amount of 10 pounds as honorarium, as well as an acknowledgement letter issued by the Department of Humanities of Imperial College London for having participated in this doctoral study. Ultimately, you will be offered the results of this study, which I hope will be a

good learning experience for you. Any personal information that is obtained in connection with this study will remain confidential.

I want you to understand that your participation in this study is completely voluntary and that you are free to choose not to participate. If you do choose to participate, you may withdraw at any time during the study simply by telling me that you no longer wish to participate. Your decision whether or not to participate will not affect your future relations with Department of Humanities or Imperial College London.

If, having read the above information, you decide to participate, please sign below to indicate that you have understood all of the above and that you give your consent to participate in this study as it is described.

Please feel free to ask me any questions before or after signing this consent form. If you have any additional questions later, you may contact me at cy.chang05@ic.ac.uk, Department of Humanities, South Kensington Campus, Imperial College London, London, SW7 2AZ, UK. You may also contact my supervisors, Mr. Mark Shuttleworth and Dr. Gabriela Saldanha, South Kensington Campus, Imperial College London, London, SW7 2AZ, UK, or Professor Basant Puri, Faculty of Medicine, Clinical Sciences Center, Robert Steiner MR Unit, Hammersmith Hospital, Imperial College London. We will be happy to answer all your questions. You will be offered a copy of this form.

---------- ----------

Signature of Participant Date

---------- ----------

Signature of Investigator Date

Appendix H: Consent Form for the Spanish – English Eye-tracking Experiment

Consent Form for Translators: Eye-tracking Study

You are invited to participate in a study of translation. My name is Chieh-ying Chang, a Ph.D. student at the Department of Humanities of Imperial College London. This study is part of the thesis research for the completion of a Ph.D. degree. I hope to investigate how translation directionality relates to cognitive effort during translating. You were selected as a potential participant because you are a full-time translation student with the language pair of Spanish and English. You will be one of the students in translation- or interpretation-related departments in universities or colleges who have been invited to participate in this study.

If you decide to participate, you will be asked to fill out a questionnaire on your demographic information, your IELTS score, age of second language acquisition, handedness, your translation training, prior educational background and work experience. You will be asked to conduct translation of two 50-word non-technical texts, one into English and the other into Spanish. Before this, you will be asked to read aloud the texts once. Also, two typing tasks will be administered respectively in English and Spanish. You will be asked to sit in front of a computer, with a non-invasive infrared eye-tracking machine recording your eye movement and pupil dilation. The whole process will take 90 minutes at most. When you have finished the translation, you will be given a cash amount of 10 pounds as honorarium, as well as an acknowledgement letter issued by the Department of Humanities of Imperial College London for having participated in this doctoral study. Ultimately, you will

be offered the results of this study, which I hope will be a good learning experience for you. Any personal information that is obtained in connection with this study will remain confidential.

I want you to understand that your participation in this study is completely voluntary and that you are free to choose not to participate. If you do choose to participate, you may withdraw at any time during the study simply by telling me that you no longer wish to participate. Your decision whether or not to participate will not affect your future relations with Department of Humanities or Imperial College London.

If, having read the above information, you decide to participate, please sign below to indicate that you have understood all of the above and that you give your consent to participate in this study as it is described.

Please feel free to ask me any questions before or after signing this consent form. If you have any additional questions later, you may contact me at cy.chang05@ic.ac.uk, Department of Humanities, South Kensington Campus, Imperial College London, London, SW7 2AZ, UK. You may also contact my supervisors, Mr. Mark Shuttleworth and Dr. Gabriela Saldanha, South Kensington Campus, Imperial College London, London, SW7 2AZ, UK, or Professor Basant Puri, Faculty of Medicine, Clinical Sciences Center, Robert Steiner MR Unit, Hammersmith Hospital, Imperial College London. We will be happy to answer all your questions. You will be offered a copy of this form.

---------- ----------

Signature of Participant Date

---------- ----------

Signature of Investigator Date

Appendix I: Random Permutation Table

4	1	3	2
3	4	1	2
4	3	1	2
3	1	2	4
3	2	1	4
4	1	2	3
4	3	2	1
3	4	1	2
4	2	1	3
3	2	4	1
2	3	1	4
3	2	1	4
1	2	4	3
3	2	1	4
1	3	4	2

Appendix J: Language and Translation Background Questionnaire (Mandarin – English Eye-tracking Experiment)

Language and Translation Background Questionnaire

This questionnaire concerns your language and translation experiences over the course of your lifetime. You can choose to answer the following questions either in Mandarin or in English. All responses are confidential and will be analyzed only for research purposes. If there are questions you prefer not to answer, you may choose to skip them. Thank you for your participation.

PARTICIPANT _____

1. Where were you born? _____

2. What is your nationality? _____

3. What is your gender? Male / Female

4. What is your age?

18-25 26-35 36-45 46-55 56-65 over 65

5. What is your handedness? Right-handed /Left-handed

6. What is your native language? _____

7. What is your second language? _____

8. What is your language of habitual use now? _____

9. What form of Mandarin are you more comfortable with during reading?

Simplified Mandarin / Traditional Mandarin

10. What form of Mandarin are you more comfortable with during writing?

Simplified Mandarin / Traditional Mandarin

11. What other languages or dialects do you know (including both foreign languages and dialects such as Cantonese, Taiwanese or Shanghainese, etc.)?

_____

12. What is the Chinese input method that you are used to using on a computer?

_____

13. Which language do you find more difficult to key in on a computer and why?

A. English / B. Mandarin

Please explain: _____

14. At what age did you start learning your second language? _____

15. Up to today, what is the total amount of time you have spent on formally studying your second language in a school setting? _____ year(s)_____month(s)

16. Up to today, what is the total amount of time you have spent using your second language in a practical way (e.g. verbal communication, paper writing, reading for obtainment of knowledge, etc.)? _____year(s)_____month(s)

17. Up to today, what is the total amount of time you have spent being immersed in an English-speaking environment? _____year(s)_____month(s)

18. In what major was your **undergraduate** degree? _____

19. What is your overall band on IELTS? _____

20. What are the bands for different tests in IELTS?

A. Listening _____ B. Reading _____

C. Writing _____ D. Speaking _____

21. Which university or college do or did you attend for your postgraduate translation training? _____

22. How long is the **postgraduate** translation course mentioned in Question 21? - _____

23. Of all the **postgraduate** translation assignments, approximately how much is translating from English into Mandarin and how much from Mandarin into English?

From English into Mandarin _____%

From Mandarin into English _____%

24. Which DIRECTION of translation do you find more comfortable working with?

A. From English into Mandarin

B. From Mandarin into English

25. Have you ever worked as a translator on a professional basis (e.g. paid translation work)?

Yes (Please proceed to answer the next question).

No (Please proceed to answer Question 28.)

26. How long have you worked as a professional translator? _____

27. Of all the professional translation assignments, approximately how much is translating from English into Mandarin and how much from Mandarin into English?

From English into Mandarin _____%

From Mandarin into English _____%

28. What (other) work experiences have you had so far?

_____

THANK YOU FOR FILLING IN THIS QUESTIONNAIRE

Appendix K: Language and Translation Background Questionnaire (Spanish – English Eye-tracking Experiment)

Language and Translation Background Questionnaire

This questionnaire concerns your language and translation experiences over the course of your lifetime. You can choose to answer the following questions either in Mandarin or in English. All responses are confidential and will be analyzed only for research purposes. If there are questions you prefer not to answer, you may choose to skip them. Thank you for your participation.

PARTICIPANT _____

1. Where were you born? _____

2. What is your nationality? _____

3. What is your gender? Male / Female

4. What is your age?

18-25 26-35 36-45 46-55 56-65 over 65

5. What is your handedness? Right-handed /Left-handed

6. What is your native language? _____

7. What is your second language? _____

8. What is your language of habitual use now? _____

9. What other languages or dialects do you know (including both foreign languages and dialects such as Catalan, French, German, Italian etc.)?

_____

10. Which language do you find more difficult to key in on a computer and why?

A. English / B. Spanish

Please explain: _____

11. At what age did you start learning your second language? _____

12. Up to today, what is the total amount of time you have spent on formally studying your second language in a school setting? _____ year(s)_____month(s)

13. Up to today, what is the total amount of time you have spent using your second language in a practical way (e.g. verbal communication, paper writing, reading for obtainment of knowledge, etc.)? _____year(s)_____month(s)

14. Up to today, what is the total amount of time you have spent being immersed in an English-speaking environment? _____year(s)_____month(s)

15. In what major was your **undergraduate** degree? _____

16. What is your overall band on IELTS? _____

17. What are the bands for different tests in IELTS?

A. Listening _____ B. Reading _____

C. Writing _____ D. Speaking _____

18. Which university or college do or did you attend for your postgraduate translation training? _____

19. How long is the **postgraduate** translation course mentioned in Question 21? -

20. Of all the **postgraduate** translation assignments, approximately how much is translating from English into Spanish and how much from Spanish into English?

From English into Spanish _____%

From Spanish into English _____%

21. Which DIRECTION of translation do you find more comfortable working with?

A. From English into Spanish

B. From Spanish into English

22. Have you ever worked as a translator on a professional basis (e.g. paid translation work)?

Yes (Please proceed to answer the next question).

No (Please proceed to answer Question 25.)

23. How long have you worked as a professional translator? _____

24. Of all the professional translation assignments, approximately how much is translating from English into Spanish and how much from Spanish into English?

From English into Spanish _____%

From Spanish into English _____%

25. What (other) work experiences have you had so far?

_____

THANK YOU FOR FILLING IN THIS QUESTIONNAIRE

Appendix L: Description of the Procedure of the Study

Description of the Procedure of the Study

In this study, you will be asked to fill out a questionnaire on your language and translation background first. Then you will be asked to carry out 4 tasks: 1) translate a non-technical Mandarin text into English; 2) translate a non-technical English text into Mandarin; 3) type a non-technical text in Mandarin; and 4) type a non-technical text in English. The four texts were carefully selected so that there will be NO technical terms involved. You will be instructed on the sequence of the tasks by the investigator.

Before each task, you will be given the text to read out loudly once to ensure that you understand the content of the text. Then your eye will be calibrated with the eye-tracker to ensure that your eye movement will be correctly recorded during the task. During calibration, please avoid body and head movement and only use your eyes to track the numbers on the screen. After calibration, the investigator will give you the instruction on whether you will be translating or typing the previously read text. Once the study starts, you can carry out the translation or typing task the way you normally do on your own computer. You can look at the keyboard when necessary. However, please avoid body movement when carrying out the tasks.

The investigator will approach your computer from time to time to create a file name, initiate the software, adjust the camera, measure the distance between your head and the camera, check the screen display on your monitor or save the data in a file. When the investigator approaches your computer, please restrain from moving your body or your head.

To familiarize you with this process, there will be a brief training session. Between tasks, you will be offered to take a 5-minute break. The whole process will take about 90 minutes. Thank you for your participation.

Appendix M: Practice Material for English → Mandarin Translation

I always ask my friends to visit me in London, because there are many historic spots to see.

Appendix N: Practice Material for Typing English

If you do not like the weather here, you can always move to somewhere else.

Appendix O: Practice Material for Mandarin → English Translation

<Traditional Mandarin>

雖然在國外唸書，我還是時常自己煮家鄉菜。

<Simplified Mandarin>

虽然在国外念书，我还是时常自己煮家乡菜。

English Back Translation:

Although I am studying abroad, I often cook food of my home country.

Appendix P: Practice Material for Typing Mandarin

<Traditional Mandarin>

世界上絕大多數國家都是聯合國的成員國。

<Simplified Mandarin>

世界上绝大多数国家都是联合国的成员国。

English Back Translation:

Most of the countries in the world are member states of the United Nations.

Appendix Q: Practice Material for Spanish → English Translation

Vivo en Inglaterra, y soy estudiante de inglés que es una lengua interesante.

English Back Translation:

I live in England, and I am studying English which is an interesting language.

Appendix R: Practice Material for Typing and Reading Spanish

Yo quiero aprender español en la escuela, pero no tengo tiempo o dinero.

English Back Translation:

I want to learn Spanish in school, but I have no time or money.

Appendix S: Hospital Anxiety and Depression Scale (HADS)

Hospital Anxiety and Depression Scale (HADS)

nferNelson
understanding potential

Name: _____ Date: _____

FOLD HERE

Clinicians are aware that emotions pay an important part in most illnesses. If your clinician knows about these feelings he or she will be able to help you more.

This questionnaire is designed to help your clinician to know how you feel. Read each item below and **underline the reply** which comes closest to how you have been feeling in the past week. Ignore the numbers printed at the edge of the questionnaire.

Don't take too long over your replies, your immediate reaction to each item will probably be more accurate than a long, thought-out response.

FOLD HERE

	A	D
I feel tense or 'wound up'		
Most of the time		3
A lot of the time		2
From time to time, occasionally		1
Not at all		0
I still enjoy the things I used to enjoy		
Definitely as much		
Not quite so much		
Only a little		
Hardly at all		
I get a sort of frightened feeling as if something awful is about to happen		
Very definitely and quite badly		3
Yes, but not too badly		2
A little, but it doesn't worry me		1
Not at all		0
I can laugh and see the funny side of things		
As much as I always could		
Not quite so much now		
Definitely not so much now		
Not at all		
Worrying thoughts go through my mind		
A great deal of the time		
A lot of the time		
Not too often		
Very little		
I feel cheerful		
Never		
Not often		
Sometimes		
Most of the time		
I can sit at ease and feel relaxed		
Definitely		
Usually		
Not often		
Not at all		
I feel as if I am slowed down		
Nearly all the time		3
Very often		2
Sometimes		1
Not at all		0
I get a sort of frightened feeling like 'butterflies' in the stomach		
Not at all	0	
Occasionally	1	
Quite often	2	
Very often	3	
I have lost interest in my appearance		
Definitely		3
I don't take as much care as I should		2
I may not take quite as much care		1
I take just as much care as ever		0
I feel restless as if I have to be on the move		
Very much indeed	3	
Quite a lot	2	
Not very much	1	
Not at all	0	
I look forward with enjoyment to things		
As much as I ever did		0
Rather less than I used to		1
Definitely less than I used to		2
Hardly at all		3
I get sudden feelings of panic		
Very often indeed	3	
Quite often	2	
Not very often	1	
Not at all	0	
I can enjoy a good book or radio or television programme		
Often		0
Sometimes		1
Not often		2
Very seldom		3

Now check that you have answered all the questions

	A	D
TOTAL		

This form is printed in green. Any other colour is an unauthorized photocopy.
HADS copyright © R.P. Snaith and A.S. Zigmond, 1983, 1992, 1994.
Record form items originally published in *Acta Psychiatrica Scandinavica* 67, 361–70, copyright © Munksgaard International Publishers Ltd, Copenhagen, 1983.
This edition first published in 1994 by nferNelson Publishing Company Ltd.
4-14 Chiswick High Road, London W4 5TF
nferNelson is a division of Granada Learning Limited
Code 0090002511 Printed in Great Britain 7/10.061

Appendix T: Consent Form for the fMRI Experiment[24]

Appendix 2[a]

Participant Consent Form

Title of project:

Optimization of a new 3T magnetic resonance imaging (MRI) scanner and comparison with a 1.5T scanner

The study has been explained to me by Prof/Dr/Mr/Mrs/Ms_____

The participant should complete the whole of this sheet him or herself.

(please tick each statement if it applies to you)

I have read the Information Sheet for Patients and Healthy Volunteers. ☐

I have been given the opportunity to ask questions and discuss this study. ☐

I have received satisfactory answers to all my questions. ☐

I have received enough information about the study. ☐

I do not wish that a qualified person, e.g radiologist, interprets the images. ☐

I understand that sections of any of my medical notes and images may be looked at by responsible individuals from our research collaborators or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to *my* records and images. ☐

I understand that I am free to withdraw from the study at any time, without having to give a reason for withdrawing and without affecting my future medical care. ☐

I agree to take part in this study. ☐

Signed_____Date_____

(NAME IN BLOCK CAPITALS)_____

[24]It has to be noted that the image qualify of the fMRI consent form used in the MRI laboratory was poor; therefore, the image of the scanned version was adjusted to the best of the researcher's ability, but its image quality is still relative low.