

**Investigating the Effectiveness of Bilingual Subtitles for Incidental  
Vocabulary Learning: A Mixed Methods Study**

by

Andi Wang

A thesis submitted to the UCL Institute of Education  
in fulfilment of the requirements for the degree of  
Doctor of Philosophy

UCL Institute of Education

2022

I, Andi Wang confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## **ABSTRACT**

Previous studies have shown that watching captioned and subtitled foreign language videos facilitates second language vocabulary learning (e.g., Koolstra & Beentjes, 1999; Montero Perez, Van Den Noortgate, & Desmet, 2013). The majority of studies have focused on the examination of captions (i.e., on-screen text in the same language as the soundtrack) and first language (L1) subtitles (i.e., on-screen text in viewers' L1). Despite the widespread use of bilingual subtitles (i.e., the simultaneous presentation of captions and L1 subtitles) in certain contexts, empirical evidence demonstrating their potential benefits for incidental vocabulary learning is scarce. In addition, available studies have yielded conflicting findings (e.g., Li, 2016; Lwo & Lin, 2012), questioning the benefits of bilingual subtitles for vocabulary development. These inconclusive findings could be due to learners' different uses of various sources of input in bilingual subtitles. However, little is known about how learners process subtitled areas or novel words in bilingual subtitles and how learners' engagement with novel words may relate to their vocabulary gains.

This thesis reports a mixed methods study undertaken with 112 Chinese learners of English to investigate: 1) the relative effectiveness of bilingual subtitles for incidental vocabulary learning, compared to captions and L1 subtitles, by using offline tests; 2) learners' attention allocation to subtitled areas and novel words by using eye-tracking; 3) learners' awareness and processing strategies for novel words by means of stimulated recall interviews; and 4) the relationship between learners' engagement with novel words (as measured by eye movements and stimulated recall interviews) and vocabulary learning gains.

Overall, the results indicate that bilingual subtitles are less beneficial than captions for recognising word forms, but more beneficial for facilitating meaning knowledge than other subtitling types. When using bilingual subtitles, learners spent more time processing L1 lines and L1 translations of unknown words than English equivalents. The bilingual subtitles group reported less awareness of novel words than the captions group but more than the L1 subtitles group. Moreover, more cases of using L1 translations to check the meanings of novel words were recorded than in the L1 subtitles group. The use of L1 translations seemed to prompt establishing initial form-meaning

connections. Additionally, the time spent on English word forms, not L1 translations, facilitated vocabulary learning.

## IMPACT STATEMENT

This research was motivated by the widespread use of bilingual subtitles by Chinese learners of English and by the lack of investigation of their benefits for language learning. The study presented in this thesis describes a comprehensive examination of the relative effectiveness of bilingual subtitles for incidental vocabulary learning, as well as of learners' engagement with on-screen text and novel words while viewing. This research provides important insights into second language learning from viewing and multimedia learning for researchers, classroom practitioners, language learners, material developers, and policymakers.

For researchers, this research represents a considerable methodological innovation, which combined both quantitative (offline performance tests and online eye-tracking data) and qualitative data (stimulated recall) to paint a more comprehensive picture of learning from viewing. To the best of my knowledge, this is the first time these methods have been used in combination in a multimedia learning setting. The use of different forms of data helps to offset the limitations of each research method. Eye-tracking unobtrusively records learners' real-time eye movements during viewing. Together with learners' self-reported data, which further reveals learners' cognitive engagement, a more accurate and thorough understanding of the relationship between learning processes and outcomes is revealed. The study presented in this thesis can inform future research in the area of vocabulary learning through multimedia and guide future methodological decisions.

While this study was situated within the context of learning from viewing outside the classroom, the results also have important implications for classroom practitioners, as subtitled viewing is now a frequently used activity in the classroom. The present findings further support using audio-visual materials that suit learners' L2 proficiency to facilitate vocabulary learning and advocate using on-screen text containing L2 to further enhance this benefit. Bilingual subtitles are effective for establishing initial form-meaning connections by providing both L1 and L2. However, the study shows that users of bilingual subtitles tended to over-rely on L1 translations, which might not be conducive to the development of form, and resulted in a relatively weak establishment of form-meaning connections. Moreover, more attention being paid to L2 word forms but not L1 translations was found to relate to greater learning gains. Therefore, I

recommend that teachers combine the use of bilingual subtitles with other techniques (e.g., pre-teaching, test announcement) to boost learners' attention to word forms. Bilingual subtitles have the potential to lead to the development of both form and meaning, but only if used appropriately. Additionally, viewing can be combined with other deliberate learning activities to reinforce newly acquired vocabulary knowledge.

Subtitled viewing was examined as an out-of-classroom activity in the present study, whose results can also be used to inform how to maximise the learning potential of such an activity. Language learners should regard audio-visual materials as language facilitators and actively use them to increase their exposure to authentic L2 input. To optimise language learning, learners should try to pay more attention to L2 input during viewing and use L1 translations to actively engage with unknown language items when the content is understandable. They can also watch repeatedly the same viewing material and make use of on-screen text based on their own need to assist their language learning.

For developers of language learning materials, this research further confirms that L1 translations can be a short-cut for learners to instantly understand the meaning of novel language items, but they may also compete with learners' attention to L2 input. This study shows that it was learners' attention to the L2 input, rather than the L1, that facilitated their learning gains. Thus, when designing bilingual vocabulary learning materials, L2 information should be designed in a more salient and attractive way to arouse learners' engagement with L2 input.

For policymakers, this research encourages the use of bilingual subtitles for imported English audio-visual materials in China, since they made a similar contribution to comprehension as L1 subtitles but had the advantage of facilitating word meaning knowledge. Bilingual subtitles should also be included in different online video platforms as an option to meet different needs.

## ACKNOWLEDGEMENTS

Towards the end of this journey, it has become clear that the biggest achievement is not what I pursued at the beginning – the completion of a PhD, but the journey itself. I am grateful for the people I met throughout this journey, it is them who shaped and influenced me to achieve what I have achieved today.

This thesis would never have been completed without the help and support from many people.

First and foremost, my principal supervisor, Dr. Ana Pellicer-Sánchez. I always find it difficult, especially in my second language, to put my gratitude and admiration to Ana into words. She is an amazing supervisor and a super nice person. It is a blessing that I did my master dissertation under Ana's supervision, from where this great journey has its prologue. Over the past four years, Ana has shown me how great a supervisor can be, and how important this role can become for a PhD student. Ana is an excellent researcher and expert in vocabulary learning and eye-tracking, which fundamentally makes her a great supervisor for me. But it is her generosity in sharing her experience and knowledge, as well as her endless passion to motivate and support her students that really makes her a marvellous supervisor. I truly appreciate all the time and the insightful guidance that she has devoted to me. And it is her constant and unwavering support that has motivated me to overcome every difficulty that I encountered during this journey. She guided me to make progress that I could only dream about, and unclouded my self-doubt to make me realise how good I am, and how great I could even be. Thank you, Ana, I am forever grateful for being your student.

I also want to thank my secondary supervisor, Prof. Andrea Révész. I met Andrea when attending her Second Language Acquisition course when doing my master, and I was fascinated by her expertise in SLA and her elegance in delivering the course in such a smooth and understandable way. She sparked my passion to do research in SLA and has been supportive throughout my master and PhD studies. I really appreciate her insightful feedback for my PhD research, which was invaluable in improving the quality of my thesis. I would also like to express my appreciation to Prof. Amos Paran and Dr. Kazuya Saito for their thorough review of my upgrade document and their constructive

feedback, which helped me to refine the research design and to complete the whole research smoothly.

I also want to thank all members and student members of the London Second Language Acquisition Research Forum and the BAAL Vocabulary Special Interest Group. Over the years, I have been very lucky to present my on-going research at the conferences and seminars of these organisations. I have received constructive and insightful feedback which in many ways has improved the quality of my thesis. I am thankful for their interest in my work and the members' generosity for sharing their expertise.

One of the biggest challenges in completing this thesis was using a novel software, R, to analyse my data. I could have never completed the analyses without attending the statistics courses delivered by the Centre for Applied Statistical Courses at UCL. I am grateful for Dr. Eirini Koutoumanou and Dr. Chibueze Ogbonnaya for being patient to answer my questions and reply to my emails. Besides, I would like to thank my PhD peers, Dr. Samuel Christopher Barclay, Dr. Xiaojun Lu, and Dr. Yi Wang for sharing their experience of learning R and providing support to help me out with some specific analysis problems.

Most importantly, I would like to thank all my dear participants, who have generously shared their precious time with me, and showed their genuine interest in this research. It is their encouraging feedback that made me even more motivated to complete this study. I also appreciate my PhD peer, Danni Shi, for her time to act as the second-rater and second-coder to code my data. A special thanks also go to the eye-tracking lab (although it is not a human being), it accompanied me for countless hours, supporting my data collection and analysis without any sudden breakdown to prevent my mental breakdown. This thesis would never have been completed without all their help and support.

I would also like to thank China Scholarship Council for the generous financial support in the past four years, covering my tuition fees and living expenses in London. And I appreciate the grants offered by the Institute of Education, University College London for my attendance to the 2021 annual conferences of American Association for Applied Linguistics and European Second Language Association.

Last but not least, I would like to thank my dearest friends and family. It is never easy for someone to study abroad alone over a number of years, especially when parts of the study overlapped with Covid pandemic and lockdown. But I am incredibly blessed and fortunate to have made many amazing friends during this PhD journey, which did not give me a moment to feel isolated. I would never forget the weekly jogging and chatting time with Xin, all the “dark” (but surprisingly tasty) meals that she made; all the happy, relaxing and wonderful times spent with Xin, Tian, Martha, Di, Rachel, with our footprints covering different places in the UK; the inspiring chatting and cooking times during the lockdown with Xumeng and Junping, and many more. I also truly appreciate the long-distance company from my mom, dad, grandmas, my boyfriend Ji Xiang, and my best friends, especially PP, Muiyang, Lisha, and Yanira. It is their steady love, unwavering support, and optimistic attitudes towards life that have motivated me from deep down. They are the fundamental momentum in supporting me to complete this thesis. Thank you all. 我爱你们，谢谢你们。

## DISSEMINATION

### Peer-Reviewed Journal

1. Parts of the offline and online data from quantitative part of this thesis have been accepted for publication as a journal article:  
Wang, A. & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning*.  
<https://doi.org/10.1111/lang.12495>

### Conference Presentations

1. Wang, A. (2020, December 4). *Exploring learners' engagement with unknown words when watching subtitled videos* [Paper presentation]. London Second Language Acquisition Research Forum (L-SLARF) PhD Autumn Conference, London, UK.
2. Wang, A. (2020, December 5). *Can watching subtitled videos help us learn English vocabulary?* [Paper presentation]. The Annual International Conference on Education, Waseda, Japan.
3. Wang, A. (2021, March 19-23). *Incidental vocabulary learning from dual subtitled viewing: An eye-tracking study*. American Association for Applied Linguistics (AAAL) 2021 Virtual Conference.
4. Wang, A. (2021, May 27). *Investigating the effectiveness of dual subtitles for incidental vocabulary learning: An eye-tracking study*. British Association for Applied Linguistics (BAAL) Vocabulary Studies SIG 2021, Sheffield, UK.
5. Wang, A. (2021, June 11). *Learning vocabulary through subtitled viewing: The role of subtitling types and attention*. Eye-tracking in SLA research: A global engagement seminar, London, UK.
6. Wang, A. (2021, June 30-July 3). *An investigation of the effectiveness of dual subtitles for Chinese EFL learners' incidental vocabulary learning: An eye-tracking study*. The European Second Language Association (EuroSLA) 30, Barcelona, Spain.

## Table of Contents

|                                                                                             |        |
|---------------------------------------------------------------------------------------------|--------|
| ABSTRACT .....                                                                              | 3      |
| IMPACT STATEMENT .....                                                                      | 5      |
| ACKNOWLEDGEMENTS .....                                                                      | 7      |
| DISSEMINATION .....                                                                         | 10     |
| Table of Contents .....                                                                     | 11     |
| List of Tables.....                                                                         | 14     |
| List of Figures .....                                                                       | 17     |
| List of Abbreviations.....                                                                  | 18     |
| <br>Chapter 1. Introduction .....                                                           | <br>19 |
| 1.1. Background to and Rationale of the Study .....                                         | 19     |
| 1.2. Aims of the Thesis .....                                                               | 24     |
| 1.3. Overview of the Thesis .....                                                           | 26     |
| <br>Chapter 2. Literature Review .....                                                      | <br>28 |
| 2.1. Vocabulary Knowledge.....                                                              | 29     |
| 2.1.1. Aspects of Vocabulary Knowledge.....                                                 | 30     |
| 2.1.2. Assessing the Form-Meaning Connection .....                                          | 34     |
| 2.2. Approaches to Vocabulary Learning .....                                                | 37     |
| 2.2.1. What is Incidental Vocabulary Learning?.....                                         | 38     |
| 2.2.2. Theoretical Support for Incidental Vocabulary Learning .....                         | 40     |
| 2.2.3. Empirical Studies on Incidental Vocabulary Learning and Influential<br>Factors ..... | 45     |
| 2.3. Incidental Vocabulary Learning from Viewing .....                                      | 49     |
| 2.3.1. Theoretical Support for Learning from Viewing .....                                  | 49     |
| 2.3.2. Use of Captions and L1 Subtitles while Viewing .....                                 | 62     |
| 2.3.3. Bilingual Subtitles.....                                                             | 78     |
| 2.4. Use of Eye-tracking in SLA.....                                                        | 91     |
| 2.4.1. Eye-Tracking Measures .....                                                          | 92     |
| 2.4.2. Eye Movements in Visual Scene Perception and Reading .....                           | 97     |
| 2.4.3. Empirical Eye-Tracking Studies on Captions/Subtitles Viewing .....                   | 99     |

|                                                                                                           |     |
|-----------------------------------------------------------------------------------------------------------|-----|
| 2.4.4. Empirical Eye-Tracking Studies on Incidental Vocabulary Learning from Reading and Viewing .....    | 110 |
| 2.5. Engagement.....                                                                                      | 122 |
| 2.5.1. Definitions of Engagement .....                                                                    | 123 |
| 2.5.2. Other Relevant Theories About Engagement in Language Learning .....                                | 125 |
| 2.5.3. Measuring Engagement.....                                                                          | 129 |
| 2.5.4. Empirical Studies Exploring L2 Learners' Engagement with L2 Vocabulary in Reading and Viewing..... | 131 |
| 2.5.5. Operationalisation of Engagement in the Current Research .....                                     | 148 |
| 2.6. Summary and Research Questions.....                                                                  | 150 |
| Chapter 3. Methodology .....                                                                              | 153 |
| 3.1. Initial Online Questionnaire Study .....                                                             | 153 |
| 3.2. Research Design.....                                                                                 | 155 |
| 3.3. Pilot Studies .....                                                                                  | 159 |
| 3.3.1. Pilot Study One.....                                                                               | 159 |
| 3.3.2. Pilot Study Two .....                                                                              | 160 |
| 3.4. Main Study .....                                                                                     | 162 |
| 3.4.1. Participants.....                                                                                  | 162 |
| 3.4.2. Research Ethics .....                                                                              | 164 |
| 3.4.3. Materials.....                                                                                     | 165 |
| 3.4.4. Instruments.....                                                                                   | 174 |
| 3.4.5. Experiment Design and Apparatus.....                                                               | 182 |
| 3.4.6. Stimulated Recall .....                                                                            | 183 |
| 3.4.7. Data Collection Procedure .....                                                                    | 184 |
| Chapter 4. QUAN Analyses and Results .....                                                                | 188 |
| 4.1. QUAN Analyses .....                                                                                  | 189 |
| 4.1.1. Scoring of Offline Tests .....                                                                     | 189 |
| 4.1.2. Eye-Movement Data Analysis .....                                                                   | 190 |
| 4.1.3. Quantitative Statistical Analyses.....                                                             | 195 |
| 4.1.4. Main Statistical Analyses.....                                                                     | 198 |
| 4.1.5. Preliminary Analyses .....                                                                         | 201 |
| 4.2. QUAN Results .....                                                                                   | 201 |
| 4.2.1. RQ1 Effects of Subtitles on Vocabulary Learning .....                                              | 201 |

|                                                                                   |     |
|-----------------------------------------------------------------------------------|-----|
| 4.2.2. RQ2 Effects of Subtitles on Comprehension .....                            | 213 |
| 4.2.3. RQ3 Eye Movements During Subtitled Viewing .....                           | 214 |
| 4.2.4. RQ4 Relationship Between Eye Movements and Vocabulary Tests .....          | 228 |
| 4.3. QUAN Interim Discussion .....                                                | 232 |
| 4.3.1. RQ1 – Vocabulary Tests .....                                               | 232 |
| 4.3.2. RQ2 – Comprehension .....                                                  | 236 |
| 4.3.3. RQ3 – Eye-Tracking data .....                                              | 237 |
| 4.3.4. RQ4 – Relationship Between Offline and Online Measures .....               | 245 |
| Chapter 5. Qual Analyses and Results .....                                        | 249 |
| 5.1. Stimulated Recall Analysis .....                                             | 249 |
| 5.1.1. Data Sampling .....                                                        | 249 |
| 5.1.2. Data Coding .....                                                          | 251 |
| 5.1.3. Inter-Coder Agreement in Coding .....                                      | 257 |
| 5.2. Qual Results .....                                                           | 257 |
| 5.2.1. Level 1: Awareness .....                                                   | 258 |
| 5.2.2. Level 2: Vocabulary Processing Strategies .....                            | 260 |
| 5.2.3. A Third Level of Engagement that Emerged: Intentionality of Learning ..... | 272 |
| 5.3. Qual Interim Discussion .....                                                | 274 |
| 5.3.1. Level 1: Awareness .....                                                   | 274 |
| 5.3.2. Level 2: Vocabulary Processing Strategies .....                            | 275 |
| 5.3.3. A Third Level of Engagement that Emerged: Intentionality of Learning ..... | 284 |
| 5.3.4. Triangulation of Qualitative and Quantitative Findings .....               | 286 |
| Chapter 6. General Discussion and Conclusion .....                                | 297 |
| 6.1. Summary of the Main Findings .....                                           | 297 |
| 6.2. Implications of the Study .....                                              | 301 |
| 6.2.1. Theoretical Implications .....                                             | 301 |
| 6.2.2. Methodological Implications .....                                          | 305 |
| 6.2.3. Pedagogical Implications .....                                             | 308 |
| 6.3. Limitations and Future Directions .....                                      | 311 |
| 6.4. Concluding Remarks .....                                                     | 316 |
| REFERENCES .....                                                                  | 319 |
| APPENDICES .....                                                                  | 349 |

## List of Tables

|                                                                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 1. Nation's (2001) Aspect of Word Knowledge Framework (p. 27) .....                                                                                                                    | 33  |
| Table 2. Laufer and Goldstein's (2004) Degrees of Vocabulary Knowledge (p. 407)...                                                                                                           | 35  |
| Table 3. Schmitt's (2010) Degrees of Vocabulary Knowledge (p. 86) .....                                                                                                                      | 36  |
| Table 4. Summary of Eye-Tracking Measures Reported in Video-Based Studies .....                                                                                                              | 95  |
| Table 5. Definitions of Commonly Reported Eye-Tracking Measures in Video-Based Studies (adapted from Conklin, Pellicer-Sánchez, and Carroll (2018, p. 68) and Godfroid (2020, p. 211)) ..... | 96  |
| Table 6. Leow's (2015) Operationalisation of Depth of Processing with Lexical Items (pp. 227-228).....                                                                                       | 127 |
| Table 7. Summary of Contextual Word Processing Strategies Reported in Previous Research (in chronological order) .....                                                                       | 142 |
| Table 8. Mixed Methods Design Matrix (adapted from Johnson and Christensen, 2012, p. 435) .....                                                                                              | 156 |
| Table 9. Descriptive Statistics for IELTS Scores and Vocabulary Size by Group .....                                                                                                          | 164 |
| Table 10. Characteristics of the 24 Target Words (in alphabetical order) .....                                                                                                               | 170 |
| Table 11. 33 Distractors Used in the Vocabulary Tests (in alphabetical order) .....                                                                                                          | 173 |
| Table 12. Example of the Form Recognition and Meaning Recall Pretest Answer Sheet .....                                                                                                      | 176 |
| Table 13. Example of the Form Recognition, Occurrence Decision and Meaning Recall Posttest Answer Sheet .....                                                                                | 179 |
| Table 14. Data Collection Procedures and Time Duration of the Present Main Study                                                                                                             | 185 |
| Table 15. The Definitions of Four Eye-Tracking Measures at Level 1 and Level 2 Analyses .....                                                                                                | 192 |
| Table 16. The Definitions of Seven Eye-Tracking Measures at Level 3 Analysis .....                                                                                                           | 195 |
| Table 17. Descriptive Statistics for the Vocabulary Pretests and Posttests by Group ..                                                                                                       | 203 |
| Table 18. Absolute Vocabulary Learning Gains of 24 Target Words on Four Offline Vocabulary Tests by Group .....                                                                              | 204 |
| Table 19. Results of the Logistic Mixed-Effects Models on Four Offline Vocabulary Tests Between Four Subtitling Groups .....                                                                 | 206 |
| Table 20. Results of Post Hoc Contrasts on Four Offline Vocabulary Tests Between Four Subtitling Groups.....                                                                                 | 209 |
| Table 21. Results of the Logistic Mixed-Effects Models Comparing Form Recognition Posttest Scores for 24 Target Words and 6 Distractors.....                                                 | 211 |
| Table 22. Results of the Logistic Mixed-Effects Models Comparing Meaning Recall Posttest Scores for 24 Target Words and 6 Distractors.....                                                   | 211 |

|                                                                                                                                                                                       |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 23. Results of the Logistic Mixed-Effects Models Comparing Meaning Recognition Posttest Scores for 24 Target Words and 6 Distractors .....                                      | 212 |
| Table 24. Descriptive Statistics for Comprehension Scores by Group .....                                                                                                              | 213 |
| Table 25. Results of the Linear Regression Model Examining Comprehension Performance .....                                                                                            | 213 |
| Table 26. Results of Post Hoc Contrasts on Comprehension Performance .....                                                                                                            | 214 |
| Table 27. Descriptive Statistics for Eye-Movement Data at Level 1 for Overall Subtitling and Image Area by Group .....                                                                | 216 |
| Table 28. Results of Post Hoc Contrasts for Total Reading Time % and Fixation % at Level 1 Overall Subtitling Area .....                                                              | 218 |
| Table 29. Results of Post Hoc Contrasts for Run Count and Skip Rate at Level 1 Overall Subtitling Area .....                                                                          | 218 |
| Table 30. Descriptive Statistics for Eye-Movement Data at Level 2 for L1 and L2 Line Area within the Bilingual Subtitles Group .....                                                  | 220 |
| Table 31. Descriptive Statistics for Eye-Movement Data at Level 2 for L2 Line Area in the Bilingual Subtitles and Captions Groups .....                                               | 221 |
| Table 32. Descriptive Statistics for Eye-Movement Data at Level 2 for L1 Line Area in the Bilingual Subtitles and L1 Subtitles Groups .....                                           | 222 |
| Table 33. Correlation Matrix of Seven Eye-Tracking Measures .....                                                                                                                     | 223 |
| Table 34. Descriptive Statistics for the Seven Eye-Movement Measures on Unknown Target Words and Corresponding L1 Translations by Group .....                                         | 225 |
| Table 35. Results of the Logistic Mixed-Effects Models Examining Relationship Between Processing of Unknown Target Words and Vocabulary Gains by Group .....                          | 230 |
| Table 36. Coding Categories and Examples of Participants' Familiarity with the Target Words in the Stimulated Recall .....                                                            | 252 |
| Table 37. Two Levels and Main Codes Generated from Stimulated Recall Interviews for Participants' Engagement .....                                                                    | 256 |
| Table 38. Results for Coding Categories of Participants' Reported Awareness (Level 1) of Unknown Target Words in Stimulated Recall Interviews .....                                   | 258 |
| Table 39. Results for the Frequency and Percentage of Participants' Reported Awareness (Level 1) of Unknown Target Words in Stimulated Recall Interviews by Group .....               | 259 |
| Table 40. Results for the Coding Categories of Participants' Vocabulary Processing Strategies (Level 2) of Unknown Target Words in Stimulated Recall Interviews .....                 | 261 |
| Table 41. Results for the Frequency and Percentage of Participants' Combination of Different Vocabulary Processing Strategies by Group .....                                          | 267 |
| Table 42. Results for the Frequency and Percentage of Participants' Vocabulary Processing Strategies (Level 2) of Unknown Target Words in Stimulated Recall Interviews by Group ..... | 268 |

|                                                                                                                                                                                             |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 43. Top Three Most Frequently Used Vocabulary Processing Strategy Categories and Subcategories in Three Subtitling Groups .....                                                       | 270 |
| Table 44. Results for the Frequency and Percentage of Subcategories and Further Subcategories Within the “3. Using L1 translations” Category in the L1 and Bilingual Subtitles Groups ..... | 272 |
| Table 45. Results for the Frequency and Percentage of Participants’ Reported Intentionality of Learning Unknown Target Words (Level 3) in Stimulated Recall Interviews by Group .....       | 273 |
| Table 46. Comparison of the Vocabulary Processing Strategies Generated by the Present Study and Previous Research .....                                                                     | 277 |

## List of Figures

|                                                                                                                                                                                                                                                            |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 1. Simple Illustration of the Mixed Methods Design and Data Analysis .....                                                                                                                                                                          | 26  |
| Figure 2. Example of a Vocabulary Knowledge Scale Test (Wesche & Paribakht, 1996, p. 30) .....                                                                                                                                                             | 31  |
| Figure 3. Example of a Word Associates Format test (Read, 1998, p. 46).....                                                                                                                                                                                | 32  |
| Figure 4. Relationship Between Awareness and the Three Components of Attention in Input Processing from Tomlin and Villa's (1994) Model of Attention (p. 197) .....                                                                                        | 44  |
| Figure 5. Dual Coding Theory (Paivio, 1986, p. 67) .....                                                                                                                                                                                                   | 50  |
| Figure 6. Bilingual Version of the Dual Coding Model (Paivio & Desrochers, 1980, p. 391) .....                                                                                                                                                             | 52  |
| Figure 7. Cognitive Theory of Multimedia Learning (Mayer, 2005a, p. 37).....                                                                                                                                                                               | 55  |
| Figure 8. Illustration and Definition of Commonly Reported Eye-Tracking Measures in Text-Based Studies (adapted from Conklin and Pellicer-Sánchez (2016, p. 456), Conklin, Pellicer-Sánchez, and Carroll (2018, p. 68), and Godfroid (2020, p. 211)) ..... | 94  |
| Figure 9. QUAN + qual Research Design (Creswell, 2003, p. 214).....                                                                                                                                                                                        | 158 |
| Figure 10. Visual Diagram of the Research Procedures in the Present Study .....                                                                                                                                                                            | 159 |
| Figure 11. Screenshots of the Four Subtitling Types in the Present Research .....                                                                                                                                                                          | 168 |
| Figure 12. Illustrations of Level 1 Areas of Interest for Eye-Movement Data Analysis in Four Groups .....                                                                                                                                                  | 191 |
| Figure 13. Illustrations of Level 2 Areas of Interest for Eye-Movement Data Analysis in Three Groups with On-Screen Text .....                                                                                                                             | 193 |
| Figure 14. Illustrations of Level 3 Areas of Interest for Eye-Movement Data Analysis in Three Groups with On-Screen Text .....                                                                                                                             | 194 |
| Figure 15. Illustration of Stimulated Recall Coding Procedures at Two Levels for Each Target Word .....                                                                                                                                                    | 254 |

## **List of Abbreviations**

|       |                                                      |
|-------|------------------------------------------------------|
| AOI   | Areas of interest                                    |
| BNC   | British National Corpus                              |
| CEFR  | Common European Framework of Reference for Languages |
| EFL   | English as a foreign language                        |
| ESL   | English as a second language                         |
| IELTS | International English Language Testing System        |
| IP    | Interest period                                      |
| L1    | First language                                       |
| L2    | Second language                                      |
| OR    | Odds ratio                                           |
| ROI   | Region of interest                                   |
| SLA   | Second language acquisition                          |
| TW    | Target word                                          |
| VKS   | Vocabulary Knowledge Scale                           |
| VLT   | Vocabulary Levels Test                               |

## Chapter 1. Introduction

### 1.1. Background to and Rationale of the Study

Vocabulary, regarded as “building blocks” in language use (Webb & Nation, 2017, p. 5), is an essential and fundamental component of second language acquisition (SLA) (Schmitt, 2010). There is a wealth of research evidence showing that second language (L2) learners’ vocabulary knowledge contributes greatly to their proficiency level and to the four language skills (i.e., reading, listening, speaking, and writing), making vocabulary knowledge a vital prerequisite for language learning success (Qian & Lin, 2020; Schmitt, 2010).

English as a Foreign Language (EFL) learners are expected to master 6,000–7,000 word families in order to accomplish informal daily conversation, and this number rises to around 8,000–9,000 when it comes to reading a range of authentic texts (Nation, 2006). Due to the limited amount of classroom time, it has been suggested that deliberate vocabulary learning needs to be supplemented by incidental learning (Krashen, 1989; Schmitt, 2010; Webb, 2020b; Webb & Nation, 2017). In general, *incidental vocabulary learning* is vocabulary learning that occurs as a by-product of meaning-focused activities or tasks, where learners’ primary objective is to focus on understanding the meaning without an effort focusing on learning language (Ellis, 1999; Hulstijn, 2003). Vocabulary researchers agree that being exposed to large amounts of comprehensible L2 input facilitates vocabulary learning, and that the advantages of incidental vocabulary learning should not be underestimated (Nation, 2013; Webb, 2020a; Webb & Nation, 2017).

Evidence for possible ways to learn vocabulary incidentally abounds. Research has shown that language learners can expand their vocabulary size and deepen their vocabulary knowledge as a by-product of communicative activities, where the main aim is comprehension rather than vocabulary learning (Hulstijn, 2001, 2003). Previous studies have demonstrated the effectiveness of reading (e.g., Elgort, Brysbaert, Stevens, & Van Assche, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006), listening (e.g., Brown, Waring, & Donkaewbua, 2008; Elley, 1989; Pavia, Webb, & Faez, 2019; Van Zeeland & Schmitt, 2013), and reading-while-listening (e.g., Chang, 2019; Vu & Peters, 2020; Webb, Newton, & Chang, 2013) for foreign language learners' incidental vocabulary growth. More recently, researchers have started to explore the effectiveness of viewing and have shown its benefits for incidental vocabulary learning (e.g., Montero Perez, Peters, & Desmet, 2018; Peters & Webb, 2018; Rodgers & Webb, 2019).

Apart from the advantages of wide availability and easy accessibility (Montero Perez, 2020b; Rodgers & Webb, 2011), watching authentic audio-visual materials has more potential to motivate L2 learners and further increase their language exposure compared to traditional L2 input (Peters, 2018; Webb & Rodgers, 2009). Moreover, the combination of both visual and aural input can be more conducive to incidental vocabulary learning than either visual or aural input alone (Duquette & Painchaud, 1996; Neuman & Koskinen, 1992). Especially, the use of *captions* (i.e., on-screen text in the same language as the soundtrack) and *first language (L1) subtitles* (i.e., on-screen text translated in the viewer's L1) has been found to support this process (e.g., Frumuselu, De Maeyer, Donche, & Colon Plana, 2015; Koolstra & Beentjes, 1999; Montero Perez et al., 2013; Peters, 2019; Pujadas & Muñoz, 2019; Winke, Gass, & Sydorenko, 2013). The majority of studies on viewing have examined the effectiveness of the use of

captions and/or L1 subtitles for learning, as they have been claimed to be the ones most frequently encountered by foreign language learners (Muñoz, 2017). However, in certain contexts, *bilingual subtitles* (i.e., the simultaneous presentation of L1 subtitles and captions; Bartolomé & Cabrera, 2005) are the preferred and most frequently used subtitling type and, despite their popularity, very little research has been conducted to examine their benefits for language learning.

In the Chinese context, where my study is situated, online viewing is a very popular form of entertainment. A 2021 report by the China Internet Network Information Center showed that there were 989 million online video users in mainland China, and 927 million of them accessed video sites online (CNNIC, 2021), which covers about 66 per cent of the overall population. Additionally, watching foreign language audio-visual material is a very common entertainment activity among Chinese EFL learners, and bilingual subtitles are a strong competitor for monolingual subtitles (Li, 2016). An initial online survey conducted as part of this thesis (see section 3.1) showed that Chinese learners had a clear preference for bilingual subtitles. However, despite their widespread use, empirical studies examining the effectiveness of bilingual subtitles for L2 vocabulary learning are still scarce.

Bilingual subtitles are believed to be conducive to vocabulary learning because L1 lines provide translations of unknown L2 words and facilitate comprehension, while L2 lines provide the form of unknown words and help learners link written and spoken forms (Li, 2016; Lwo & Lin, 2012). The possibility of connecting an L2 unknown form with its correct meaning might support vocabulary learning. However, according to the Depth of Processing Theory ( Craik & Lockhart, 1972), it can also be argued that having translations of L2 unknown words may reduce learners' cognitive analysis of their meanings and lead to shallower memory traces, which are then reflected in smaller

gains. Importantly, according to the Cognitive Load Theory (Sweller, 1988), there is also the potential for cognitive overload, as when identical information is presented in different forms, learners' working memory may be overloaded, resulting in a redundancy effect that hinders the learning process (Sweller, 2005b; Winke et al., 2013). Empirical evidence examining the effectiveness of bilingual subtitles for vocabulary learning is scarce and available research has yielded conflicting results, with studies reporting both an advantage for bilingual subtitles over captions and L1 subtitles (e.g., Lazareva & Loerts, 2017; Li, 2016) and a lack of any significant difference between bilingual subtitles and other subtitling conditions (e.g., Hao, Sheng, Ardasheva, & Wang, 2021; Lwo & Lin, 2012). These inconsistent findings could be due to learners' differential use of the sources of input available in bilingual subtitles.

Apart from the lack of research on bilingual subtitles, the majority of studies exploring the effectiveness of different subtitling types have mainly used offline, post-viewing tests, which, although highly informative, cannot tell us much about how learners make use of bilingual subtitles. Thus, it is still not clear how learners process different sources of input and how that relates to their vocabulary learning. One technique that can shed some light on learners' online processing of bilingual subtitles is eye-tracking. This technique has been used to detect learners' attention allocation during subtitled viewing (e.g., Gass, Winke, Isbell, & Ahn, 2019; Muñoz, 2017; Winke, Gass, & Sydorenko, 2013). However, only one eye-tracking study to date has investigated learners' attention allocation during bilingual subtitled viewing (Liao, Kruger, & Doherty, 2020). Due to the limited sample size and short length of stimuli, learners' processing of bilingual subtitles is still far from settled. Most importantly, vocabulary gains were found to closely relate to the amount of time spent processing novel words while viewing (Montero Perez, Peters, & Desmet, 2015). This is particularly relevant

for bilingual subtitles, where learners can choose how they want to allocate their attention (to a L2 novel word and/or its translation), but no studies so far have investigated this relationship in bilingual subtitled viewing.

Attention paid to a lexical item can partly reveal a learner's "engagement" with the word (Schmitt, 2008, p. 339). However, despite the essential role of attention in vocabulary learning, increased attention does not always lead to greater learning gains, since there are other factors influencing vocabulary learning that should also be taken into account (Schmitt, 2008, 2010). Eye-tracking data can only reveal where and for how long learners locate their attention while viewing, but cannot inform us about learners' underlying cognitive processes, i.e., what they are thinking about when processing a word (Godfroid & Winke, 2015; Montero Perez et al., 2015; Pellicer-Sánchez, 2020a). *Engagement* is a multifaceted construct that has many definitions in education and language learning research. In the present study, the definition of engagement with vocabulary is adapted from Svalberg's (2009) construct of *engagement with language*. Incorporating Svalberg's (2009) definition of cognitive engagement into the vocabulary learning field in particular, in this study, it is operationalised as attention, awareness, and vocabulary processing strategies. To have a more comprehensive view of learners' engagement with unknown vocabulary, it is necessary to not only investigate the level of learners' attention, as reflected in eye movements, but also to probe learners' awareness and different processing strategies used to engage with words by collecting learners' self-reported introspective data (as measured by stimulated recall interviews).

To date, little is known about how learners engage with novel words in bilingual subtitles and how learners' engagement may relate to their vocabulary gains. Having a better understanding of how learners process and make use of bilingual subtitles should

help to explain the current conflicting research findings on the effectiveness of bilingual subtitles. Furthermore, by comparing online and offline measures, we should be able to see if (and how) learners' allocation of attention is related to their vocabulary learning gains. Additionally, the triangulation of quantitative findings (i.e., online and offline measures) and qualitative findings (i.e., stimulated recall interviews) can paint a fuller picture of learners' engagement and help us better understand the relationship between learners' engagement and learning gains.

## **1.2. Aims of the Thesis**

As outlined above, despite the widespread use of bilingual subtitles among Chinese learners of English, there is a paucity of research investigating their effectiveness on incidental vocabulary learning. Furthermore, no research has explored how learners engage with unknown words during bilingual subtitled viewing. A thorough investigation of the use of bilingual subtitles can reveal the potential of bilingual subtitles for L2 learners' incidental vocabulary learning and maximise this learning potential. Since viewing is a meaning-focused activity and learners engage in this activity with the main aim of understanding its content, it is also important to explore potential differences in comprehension. Thus, while the primary focus of the thesis is to examine vocabulary learning, the relative effectiveness of bilingual subtitles on comprehension is also explored. The current research addresses the following aims:

1. To examine the relative effectiveness of bilingual subtitles for incidental vocabulary learning and comprehension, compared to captions, L1 subtitles, and no subtitles, by using offline tests.

2. To investigate learners' attention allocation to subtitled areas and novel words during bilingual subtitled viewing (in comparison to captions, L1 subtitles, and no

subtitles), through learners' recorded eye movements.

3. To explore learners' awareness of novel words and their use of vocabulary processing strategies during bilingual subtitled viewing (in comparison to captions, L1 subtitles, and no subtitles) by means of stimulated recall interviews.

4. To examine the relationship between learners' engagement with novel words (as measured by eye movements and stimulated recall interviews) and learning gains (as measured by offline vocabulary tests).

In order to achieve these aims, a mixed methods study was conducted. As shown in Figure 1, quantitative methods (as represented in two rectangles) include offline tests to measure the effects of bilingual subtitles on vocabulary learning and online eye-tracking data to capture learners' eye movements while viewing. The qualitative method (as represented in the oval) includes learners' verbal reports of their cognitive processes while viewing to examine their awareness and vocabulary processing strategies. Quantitative analyses of vocabulary tests and eye-movement data were first conducted, followed by qualitative analyses of stimulated recall interviews. The relationship between learners' vocabulary tests scores and eye-movement data was analysed quantitatively (as shown by the solid arrowed line). Results from the different analyses were finally triangulated (as shown by three arrowed lines) to examine the relationship between learners' engagement and vocabulary gains.

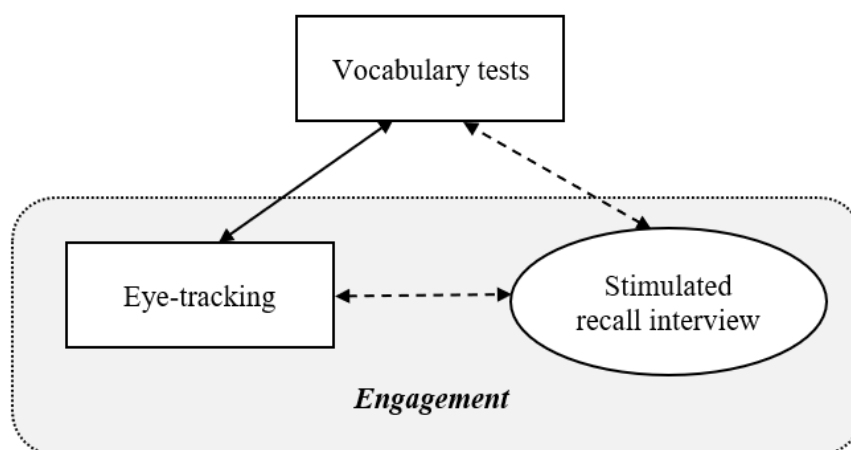


Figure 1. Simple Illustration of the Mixed Methods Design and Data Analysis

### 1.3. Overview of the Thesis

The remainder of the thesis consists of five chapters. Chapter 2 reviews relevant literature by first introducing the construct of vocabulary and the theoretical and empirical evidence supporting incidental vocabulary learning. This is followed by a review of the theoretical and empirical support for incidental vocabulary learning through viewing. Empirical studies exploring the use of captions and L1 subtitles in incidental vocabulary learning are reviewed afterwards. Bilingual subtitles, which are the focus of this study, are then introduced, followed by a review of empirical studies exploring their effectiveness for L2 vocabulary learning. Then, the application of eye-tracking methods in L2 vocabulary learning, with a specific focus on reading and viewing research, is discussed. This chapter ends with a review of the construct of engagement and empirical studies exploring L2 learners' engagement with vocabulary while reading and viewing.

Chapter 3 outlines the methodology of the present study. It starts by reporting the results of an initial online questionnaire conducted to support the rationale for the present study by demonstrating Chinese EFL learners' habits of viewing and subtitle use.

Then, the mixed methods design employed in the main study is described, followed by brief summaries of two pilot studies. The methodology employed in the main study is then presented. Data analysis procedures and results are presented separately for quantitative and qualitative data. Chapter 4 presents scoring and statistical analyses for quantitative data including offline tests and eye-movement data, followed by the results and interim discussion of the quantitative analysis. Chapter 5 explains the coding procedure for and analysis of stimulated recall data. The results of qualitative analyses are then summarised, followed by an interim discussion of the qualitative findings as well as the triangulation of three sets of data. Chapter 6 draws final conclusions by summarising the main findings of the study, followed by a discussion of its theoretical, methodological, and pedagogical implications. This chapter concludes by considering the limitations of the present study and suggesting possible directions for future research.

## **Chapter 2. Literature Review**

This chapter reviews literature relevant to the study presented in this thesis. The chapter begins by discussing the construct of vocabulary knowledge and assessment methods for form-meaning connection. Then, different approaches to vocabulary learning, with a particular focus on incidental vocabulary learning, are discussed. Several theoretical perspectives that can be used to support incidental vocabulary learning are summarised, along with supporting empirical evidence. Next, the theoretical basis for incidental vocabulary learning through viewing is presented, including Dual Coding Theory, the Cognitive Theory of Multimedia Learning, Cognitive Load Theory, and the redundancy principle grounded in the Cognitive Theory of Multimedia Learning, followed by a summary of empirical evidence. Then, the use of captions and L1 subtitles for vocabulary learning while viewing, together with a number of empirical studies examining the effectiveness of different subtitling types on vocabulary learning, is reviewed. It then summarises the use of bilingual subtitles and empirical studies that have explored their effectiveness for L2 vocabulary learning and comprehension. The use of eye-tracking as a method for measuring attention in reading and viewing studies is then reviewed and discussed. Definitions of engagement in learning are then summarised, together with other relevant theories concerning learners' engagement and language learning. Finally, empirical studies exploring learners' engagement with vocabulary while reading and viewing are reviewed.

## 2.1. Vocabulary Knowledge

The importance of vocabulary for language learning and meaning making is widely acknowledged. As Wilkins (1972) argues: “without grammar very little can be conveyed, without vocabulary nothing can be conveyed” (p. 111). As the most fundamental element of language use, via which sentences and paragraphs are formed, vocabulary has long been emphasised by language users and researchers (Read, 2000; Schmitt, 2010). Moreover, vocabulary knowledge contributes to overall language success, and it is an important predictor of overall proficiency (Coady & Huckin, 1997; Schmitt, 2010; Webb & Nation, 2017).

In the present thesis, I use the terms *vocabulary* and *word* interchangeably to indicate “a single unit of language that has meaning and can be spoken or written”, as defined by the online Cambridge Dictionary (<https://dictionary.cambridge.org/>). It should, however, be noted that while vocabulary knowledge has traditionally been conceptualised as knowledge of single words, there is now broad agreement that it also comprises knowledge of multiword items (Boers, 2020; Pellicer-Sánchez, 2020b; Webb & Nation, 2017). A multiword item is “a vocabulary item which consists of a sequence of two or more words (a word being simply an orthographic unit). This sequence of words semantically and/or syntactically forms a meaningful and inseparable unit” (Moon, 1997). It is usually used as an umbrella term to refer to different types of multiword combinations, for example, *lexical phrase*, *multiword unit*, *collocation*, *idiom*, and *formulaic sequence*, to name but a few (Boers, 2020). While the learning of multiword items is as important as that of single words, it is arguable that learning multiword items is different from single words due to their distinctive features and dissimilar learning burden (Pellicer-Sánchez, 2020b; Peters, 2014). Thus, the study

presented in this thesis only focuses on learning single words. In this thesis, precise terminology (e.g., *inflection*, *derivative*, *lemma*) is only used when needed to specify grammatical and morphological permutations of single words (Schmitt, 2000).

### **2.1.1. Aspects of Vocabulary Knowledge**

Vocabulary learning is challenging in two dimensions. The first dimension regards the *breadth* of vocabulary knowledge, which is the number of words that is known (Webb, 2020b). The estimated vocabulary size of an L1 English-speaking university graduate is about 20,000 word families (D'Anna, Zechmeister, & Hall, 1991; Goulden, Nation, & Read, 1990). Each *word family* includes a *base word* (the simplest form), all its *inflections* (with grammatical affixes), and its common *derivatives* (with word-class-change affixes) (Schmitt, 2000). Although some may argue that EFL learners do not need a native-like vocabulary size, according to Nation (2006), in order to understand a wide range of written and spoken texts, 8,000–9,000 and 6,000–7,000 word families are needed, respectively.

Apart from the large amount of vocabulary required for L2 learners, the second dimension concerns the *depth* of vocabulary knowledge, which is “typically defined as how well a word is known” (Yanagisawa & Webb, 2020, p. 371). It is widely acknowledged that learning a word is not an all-or-nothing process but is incremental in nature. In a recently published handbook on vocabulary studies, Yanagisawa and Webb (2020) discuss the construct of depth of vocabulary knowledge and summarise three different approaches that have been used to conceptualise and measure it.

First is the *developmental approach*. It considers the development of vocabulary knowledge, from no knowledge to full knowledge, by using scales to indicate developmental stages (Schmitt, 2010; Yanagisawa & Webb, 2020). The most widely

applied example of this approach is the Vocabulary Knowledge Scale (VKS; Paribakht & Wesche, 1993; Wesche & Paribakht, 1996). As shown in Figure 2, the VKS serves as a combination of self-reporting and performance tests by tapping into the recognition of word forms, knowledge of word meanings, and learners' ability to use words grammatically and semantically (Wesche & Paribakht, 1996). However, this approach has often been criticised for its linear developmental assumption and the lack of validity as regards measuring different types of knowledge (Yanagisawa & Webb, 2019; Read, 2000; Schmitt, 2010).

| VKS elicitation scale self-report categories |                                                                                               |
|----------------------------------------------|-----------------------------------------------------------------------------------------------|
| Self-report categories                       |                                                                                               |
| I.                                           | I don't remember having seen this word before.                                                |
| II.                                          | I have seen this word before, but I don't know what it means.                                 |
| III.                                         | I have seen this word before, and I <i>think</i> it means_____. (synonym or translation)      |
| IV.                                          | I <i>know</i> this word. It means_____. (synonym or translation)                              |
| V.                                           | I can use this word in a sentence:_____. (If you do this section, please also do Section IV.) |

Figure 2. Example of a Vocabulary Knowledge Scale Test (Wesche & Paribakht, 1996, p. 30)

The second approach, the *lexical network approach*, operationalises depth of vocabulary knowledge as learners' ability to connect different words in their mental lexicon (Yanagisawa & Webb, 2020). One of the most established tests is the Word Associates Format test designed by Read (1993, 1998). This test presents test takers with eight single words and asks them to select four words that have either a paradigmatic (i.e., being synonyms) or syntagmatic (i.e., being collocates) relationship with the target word (TW) as shown in Figure 3. Despite measuring semantic associations and collocations rather than merely focusing on single words, this test still fails to cover other aspects of vocabulary knowledge (e.g., spelling, pronunciation, and

grammatical functions) (Yanagisawa & Webb, 2020). Moreover, the scoring cannot represent learners' different degrees of knowledge well since it does not distinguish between a lack of response to a TW and incorrect responses to it (Read, 1993).

Additionally, research findings are neither easily interpreted nor comparable due to the differences in researchers' selection of TW, association relationships and test formats (Yanagisawa & Webb, 2020).

**sudden**

| beautiful quick surprising thirsty || change doctor noise school |

Figure 3. Example of a Word Associates Format test (Read, 1998, p. 46)

A more recent approach to measuring depth of knowledge is the *component approach* (also called the *dimensions approach*) (Nation & Webb, 2011; Yanagisawa & Webb, 2020). The multiple aspects of knowledge involved in knowing a word are broken down into further subcomponents, and each aspect of knowledge is measured to determine the extent of vocabulary learning (Nation & Webb, 2011). Following this approach, the most well-known and comprehensive specification of word knowledge was proposed by Nation (2001), as shown in Table 1. Nation suggests three general aspects of knowing a word: form, meaning, and use. Three sub-aspects are further identified, with each containing two types of knowledge, receptive and productive.

Table 1. Nation's (2001) Aspect of Word Knowledge Framework (p. 27)

|         |                                                 |   |                                                               |
|---------|-------------------------------------------------|---|---------------------------------------------------------------|
| Form    | spoken                                          | R | What does the word sound like?                                |
|         |                                                 | P | How is the word pronounced?                                   |
|         | written                                         | R | What does the word look like?                                 |
|         |                                                 | P | How is the word written and spelled?                          |
|         | word parts                                      | R | What parts are recognisable in this word?                     |
|         |                                                 | P | What word parts are needed to express the meaning?            |
| Meaning | form and meaning                                | R | What meaning does this word form signal?                      |
|         |                                                 | P | What word form can be used to express this meaning?           |
|         | concept and referents                           | R | What is included in the concept?                              |
|         |                                                 | P | What items can the concept refer to?                          |
|         | associations                                    | R | What other words does this make us think of?                  |
|         |                                                 | P | What other words could we use instead of this one?            |
| Use     | grammatical functions                           | R | In what patterns does the word occur?                         |
|         |                                                 | P | In what patterns must we use this word?                       |
|         | collocations                                    | R | What words or types of words occur with this one?             |
|         |                                                 | P | What words or types of words must we use with this one?       |
|         | constraints on use<br>(register, frequency ...) | R | Where, when, and how often would we expect to meet this word? |
|         |                                                 | P | Where, when, and how often can we use this word?              |

Note: In column 3, R = receptive knowledge, P = productive knowledge.

The receptive and productive knowledge (also referred to as passive and active mastery) distinction is one of the most widely used conceptualizations of vocabulary depth (Nation, 2013; Read, 2000; Schmitt, 2010). These two terms refer to skills-based vocabulary usage. In most cases, *receptive knowledge* is related to listening and reading skills, and it has been defined as “perceiving the form of a word while listening or reading and retrieving its meaning” (Nation, 2013, p. 47), while *productive knowledge* concerns speaking and writing skills and entails “wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form” (Nation, 2013, p. 47). Receptive knowledge is easier to gain than productive knowledge (Nation, 2013, 2020). However, as pointed out by Nation (2013), there is no clear-cut distinction between receptive and productive knowledge but their features are believed to overlap.

One notable advantage of the component approach is that each aspect can be more thoroughly investigated by using tests of different levels of sensitivity to measure the *strength* of knowledge. The importance of differentiating *strength* of knowledge from

*depth* of knowledge has been underscored by vocabulary researchers (Nation & Webb, 2011; Webb, 2012). While *depth* of knowledge concerns “the quality of multiple aspects of word knowledge” (Yanagisawa & Webb, 2020, p. 377), *strength* of knowledge focuses on learning one aspect of word knowledge, and it refers to “how well a single aspect is known” (Webb, 2012, p. 3). Measuring one specific aspect of vocabulary knowledge through different tests can demonstrate different degrees of knowledge of a particular aspect more accurately. Strength of knowledge can also refer to fluency or automaticity of vocabulary knowledge, which can be assessed by how quickly one aspect of knowledge is accessed (Godfroid, 2020b).

Measuring multiple aspects of vocabulary knowledge has been advocated for its comprehensive examination of word knowledge, its high construct validity when measuring different components by separating them for analysis, and the comparability of test scores (Nation & Webb, 2011; Schmitt, 2010; Yanagisawa & Webb, 2019). However, it has limitations in that it is time-consuming and only a limited number of words can be measured in each individual study (Nation & Webb, 2011; Schmitt, 2010). Due to the comprehensiveness of this approach, it is often neither practical nor necessary to examine all vocabulary aspects, therefore, the assessment of these aspects should be based on the purpose of teaching, learning, or research (Yanagisawa & Webb, 2020).

### ***2.1.2. Assessing the Form-Meaning Connection***

Among the different components of vocabulary knowledge, the form-meaning connection is the one most commonly measured in vocabulary research, given that it is the first and most important lexical aspect which must be acquired (Laufer & Goldstein, 2004; Nation, 2020; Schmitt, 2010). The emphasis on the form-meaning connection is

reasonable since words are basically units of meaning that are used to convey information (Laufer, Elder, Hill, & Congdon, 2004), and the form-meaning connection is central to all vocabulary tests.

In Nation's (2001) description of vocabulary knowledge (see Table 1), he distinguishes the form and meaning aspect at two levels, *receptively* as "what meaning does this word form signal?", and *productively* as "what word form can be used to express this meaning?" (Nation, 2001, p. 27). Laufer and Goldstein (2004) proposed four degrees of knowing the form-meaning link to assess learners' strength of word knowledge. These degrees are based on one's competence in "supplying the *form* for a given meaning versus supplying the *meaning* for a given form" (p. 405), and "being able to *recall* versus only being able to *recognize* (whether form or meaning)" (p. 406), as shown in Table 2. The hierarchy of difficulty for the four categories has been validated as (> = more difficult than): active recall > passive recall > active recognition > passive recognition. This hierarchy was confirmed by Laufer et al. (2004) but with active and passive recognition being equally easy to acquire.

Table 2. Laufer and Goldstein's (2004) Degrees of Vocabulary Knowledge (p. 407)

|                                | Recall             | Recognition        |
|--------------------------------|--------------------|--------------------|
| Active (retrieval of form)     | Supply the L2 word | Select the L2 word |
| Passive (retrieval of meaning) | Supply the L1 word | Select the L1 word |

When proposing the four degrees of form-meaning link, as shown in Table 2, Laufer and Goldstein (2004) originally used the terms *active* and *passive* to distinguish the retrieval of form and meaning. To make the terms more transparent when describing the construct being measured, Schmitt (2010) relabelled the four categories as

measuring the *aspects* required (form vs. meaning) and the *degree of mastery* (recognition vs. recall). Schmitt (2010) advocates the use of the following terms: form recall, meaning recall, form recognition, and meaning recognition, as replacements for Laufer and Goldstein's (2004) original terms (i.e., active recall, passive recall, active recognition, and passive recognition) to assess the degrees of knowledge of form-meaning connections. For the sake of transparency, Schmitt's (2010) terminology (as shown in Table 3) is adopted in the present thesis.

Table 3. Schmitt's (2010) Degrees of Vocabulary Knowledge (p. 86)

| Word knowledge | Word-knowledge tested                                         |                                                                    |
|----------------|---------------------------------------------------------------|--------------------------------------------------------------------|
| <i>Given</i>   | <i>Recall</i>                                                 | <i>Recognition</i>                                                 |
| Meaning        | Form recall<br>(supply the L2 item)                           | Form recognition<br>(select the L2 item)                           |
| Form           | Meaning recall<br>(supply definition/L1<br>translation, etc.) | Meaning recognition<br>(select definition/L1<br>translation, etc.) |

These four form-meaning link categories are believed to capture the specificity of vocabulary knowledge. Instead of referring to vocabulary simply as being learned, they can better reveal the incremental nature of the vocabulary learning process and show learners' degree of mastery (Read, 2000; Schmitt, 2010). In vocabulary assessment, recognition and recall ability are also commonly used, especially in discrete/selective/context-independent vocabulary tests (Schmitt, 2010). *Recognition* is operationalised as "test-takers are presented with the TW and are asked to show that they understand its meaning" (Read, 2000, p. 155), and *recall* is when "they [test-takers] are provided with some stimulus designed to elicit the TW from their memory" (Read, 2000, p. 155). The adoption of these categories also allows vocabulary research to be more comparable across different studies (Schmitt, 2010).

## 2.2. Approaches to Vocabulary Learning

Given the large vocabulary targets required for L2 learners, an important question in vocabulary research is: how can L2 learners learn so many words? Vocabulary can be learned intentionally or incidentally (Schmitt, 2000; Webb, 2020a; Webb & Nation, 2017). *Intentional vocabulary learning*, often referred to as *deliberate*, *instructed*, or *explicit learning* in the literature (Webb, 2020a), refers to learning which takes place when the primary intention of learners/activities is the development of vocabulary knowledge (Hulstijn, 2003; Webb, 2020b). It can occur during the explicit teaching of vocabulary in classroom settings. Schmitt (2000, pp. 144-145) proposed four categories of words that should be taught explicitly: 1) the first 2,000 most frequent words; 2) words particularly useful in a specific topic area for students; 3) words that students want to learn (learner-centred); and 4) the vocabulary necessary for classroom management. In addition, intentional learning also occurs when learners are engaged in vocabulary-learning activities or use various learning strategies to intentionally boost their vocabulary learning (Webb, 2020a; Webb & Nation, 2017). For example, memorising vocabulary using bilingual vocabulary lists (Hulstijn, 2001), applying verbal, visual or combined mnemonic devices, exploring the similarities and differences between new words and known words (Schmitt & McCarthy, 1997), among others. Empirical evidence showing the effectiveness of different approaches to the intentional learning of vocabulary abounds (e.g., Laufer & Shmueli, 1997; Nakata, 2008; Peters, 2007).

However, intentional learning is time-consuming, and it might not be enough to support L2 learners achieving the large vocabulary size targets mentioned in section 2.1.1 (Schmitt, 2000). It has also been suggested that only a fraction of words can be

acquired through formal study. After mastering the first 2,000–3,000 word families through a more intentional and explicit teaching approach, L2 vocabulary acquisition occurs incidentally through substantial informal language input (Coady & Huckin, 1997). Learners are believed to gradually build up and strengthen their knowledge of new words incidentally through numerous exposures (Nation, 2001; Schmitt, 2000). The study presented in this thesis is situated within the incidental vocabulary learning approach. The theoretical and empirical support for incidental vocabulary learning is discussed in the next section.

### ***2.2.1. What is Incidental Vocabulary Learning?***

Despite the numerous studies on incidental vocabulary learning, there is no consensus for the definition of this construct. Incidental vocabulary learning is often defined dichotomously with intentional vocabulary learning. From a *cognitive perspective*, the focus in the conceptualisation of incidental learning is on learners' intention. From this perspective, incidental vocabulary learning is learning vocabulary without the intent to learn (Hulstijn, 2003; Laufer & Hulstijn, 2001), whereas intentional vocabulary learning is learning vocabulary with deliberate intention and in an attempt to commit a specific set of words to memory (Hulstijn, 2003; Schmitt & McCarthy, 1997). However, researchers have argued that these definitions can be problematic, since there may be degrees of intention within an incidental learning situation due to the impossibility of ascertaining learners' exact attentive process in vocabulary learning (Bruton, Lopez, & Mesa, 2011; Gass, 1999; Hulstijn, 2001; Webb, 2020a), resulting in great difficulty in distinguishing between these two terms.

Therefore, instead of focusing on degrees of intention and attention, these two terms have been more commonly defined from a *pedagogical perspective* by

emphasising the purpose of the activity (Webb, 2020a). From a pedagogical perspective, incidental vocabulary learning refers to the vocabulary learning that occurs as a by-product of meaning-focused activities or tasks (Hulstijn, 2003), while intentional learning occurs in language-focused activities where learners' primary objective is to focus on the form of language rather than understanding the meaning (Hulstijn, 2003; Schmidt, 1994). Following this perspective, Webb (2020a) points out that the importance lies in the purpose of the activity rather than where intention and attention are located during the activity. Hence, to avoid the issues of attention and intention, and to bring more transparency to the terminology, Webb (2020a) advocates the use of two alternative terms: "meaning-focused learning" and "language-focused learning" (p. 226) to address definition issues.

A third approach to the definition of incidental and intentional learning is *methodological*. This approach is considered appropriate for researchers aiming to design a vocabulary learning experiment (Hulstijn, 2001). Similar to the pedagogy-oriented perspective, this approach does not differentiate between the two terms based on learners' intention or attention, but rather on the research design. To operationalise incidental vocabulary learning, learners are typically required to accomplish a task involving the processing of some information without being informed in advance of a forthcoming vocabulary test (Laufer & Hulstijn, 2001). This differs from an intentional learning design where learners are told in advance that they will be tested on their recall of new vocabulary afterwards. A great number of empirical studies have also pursued incidental vocabulary learning from this methodological perspective (e.g., Jelani & Boers, 2018; Montero Perez et al., 2018; Peters, Heynen, & Puimège, 2016; Peters & Webb, 2018; Webb et al., 2013). Following previous studies, *incidental vocabulary learning* in the present research is defined as a research tool and

operationalised as the absence of any test announcement in a meaning-focused activity. However, it should be noted that a subjective intention to learn can occur in an incidental learning research design, but intention of itself does not result in learning (Schmidt, 1995).

### ***2.2.2. Theoretical Support for Incidental Vocabulary Learning***

Although the notion of incidental learning is not firmly rooted in any particular theory (Hulstijn, 2003), vocabulary learning without intention is believed to occur during the process of comprehending and extracting information from language input (MacFadden, Barrett, & Horst, 2009). Krashen's (1989) Input Hypothesis, Schmidt's (1990) Noticing Hypothesis, and Tomlin and Villa's (1994) Model of Attention may shed some light on this learning process.

#### **2.2.2.1. Krashen's Input Hypothesis**

The Input Hypothesis assumes that "we acquire language by understanding messages" (Krashen, 1989, p. 440). It is based on the natural order hypothesis, which was first put forward by Corder (1967) and suggests that the rules of language are acquired in a predictable order. Krashen (1982) uses the formula " $i + 1$ " to represent the language learning process (p. 23), where " $i$ " refers to current competence, and " $i + 1$ " represents the next level. The " $i + 1$ " hypothesis (Krashen, 1982, p. 23) proposes that unknown language structures can be acquired when meaningful information is presented, using learners' already acquired language structures plus a language structure that is slightly beyond learners' previous knowledge. A prerequisite of learning is to understand the new input. Krashen maintains that acquisition can only occur when learners understand language by focusing on the meaning rather than the form of the

message. Krashen (1985) thus emphasises the importance of context, which includes “extra-linguistic information, our knowledge of the world, and previously acquired linguistic competence” (p. 2) and that can help learners understand language containing unacquired language structures. Input is thus regarded as “the essential environmental ingredient” (Krashen, 1985, p. 2).

Through an extensive review of empirical studies on L1 and L2 incidental vocabulary learning, Krashen (1989) concluded that learners can acquire vocabulary efficiently from comprehensible input. However, a notable shortcoming of the Input Hypothesis is that it does not clearly explain how comprehended input leads to acquisition. This has been pointed out by Lawson and Hogben (1996), who emphasize that researchers often did not distinguish well “between comprehension of word meaning in context and the acquisition of word meaning from context” (p. 105).

#### **2.2.2.2. Schmidt’s Noticing Hypothesis**

Schmidt’s (1990) Noticing Hypothesis attempted to further explore the transformation of comprehended input into acquisition. It denies the possibility of subliminal language learning and emphasises the necessity of *noticing*, claiming that “if noticed, it becomes intake” (p. 139). Schmidt (1990) distinguishes *intake* from comprehensible input, which are considered synonyms by Krashen. Schmidt (1990) claims that “*intake* is that part of the input that the learner notices” (p. 139), which is a prerequisite for learning. Learners select particular parts of the input for further processing, which is necessary for input to become intake and can finally result in learning. Schmidt (1990) also points out that, regardless of whether the linguistic form is noticed deliberately or purely accidentally, it can be regarded as intake. Importantly,

incidental learning is considered to be one type of intake process that requires spontaneous noticing (Schmidt, 1990).

Noticing has thus been seen as a “necessary and sufficient condition for the conversion of input into intake” (Schmidt, 1993, p. 209), and it is believed to be crucial to generate intake and finally result in learning. According to Schmidt (1990), *noticing* is a private cognitive experience in which stimuli are subjectively experienced. It is operationalised as the availability of self-reporting that takes place during and immediately after exposure to input (Schmidt, 1990, p. 132). However, it is debatable whether noticing is a necessary condition for learning to occur (Gass, Svetics, & Lemelin, 2003; Robinson, 2003). In his later publications, Schmidt (2001) also makes a weaker argument by claiming that noticing, if not necessary, is at least facilitative of learning.

In addition, Leow (1997, 2015) points out that Schmidt’s (1990, 1993) definition of noticing is a combination of focal attention and a low level of awareness. Therefore, the term *noticing* has also been criticised as a “hybrid concept” (Godfroid, Boers, & Housen, 2013, p. 485), entailing both attention and awareness, which are often studied separately in the cognitive psychology and bilingualism fields (Baddeley, 2007). This mixed use of terms may cause theoretical confusion which is unhelpful in disentangling the learning process (Godfroid et al., 2013).

#### **2.2.2.3. Tomlin and Villa’s Model of Attention**

Different from Schmidt’s Noticing Hypothesis, Tomlin and Villa (1994) distinguished *attention* theoretically and empirically from *awareness*, claiming that “awareness requires attention, but attention does not require awareness” (p. 194). They define *awareness* as “a particular state of mind in which an individual has undergone a

specific subjective experience of some cognitive context or external stimulus” (p. 193). In this definition, the role of awareness in the preliminary processing of input into intake during exposure is lessened, but attention is considered necessary in this process. A fine-grained model of attention for input processing in SLA was thus proposed.

According to Tomlin and Villa (1994), *attention* contains three separate, but interrelated, principal components: alertness, orientation, and detection, as can be seen in Figure 4. *Alertness* represents “an overall, general readiness to deal with incoming stimuli or data” (p. 196). It relates to the speed of information selection. For example, having a warning before the task could activate the brain area and thus lead to faster detection of the target. *Orientation* is the action of directing attentional resources to a certain type of stimulus but excluding others. It can facilitate or restrain further processing depending on whether the information encountered is expected or not. It is proposed that these two components (i.e., alertness and orientation) can facilitate detection, which is the last but vital component of attention. *Detection* is “the process that selects, or engages, a particular and specific bit of information” (p. 192). Only if information is detected can further processing of input (i.e., hypothesis formation) and subsequent learning occur, as illustrated in Figure 4. Tomlin and Villa (1994) propose that none of the three components of attention require awareness. Although alertness and orientation can be augmented by exploiting awareness, which can further enhance detection, it is detection itself that is necessary for learning. In other words, information can be cognitively attended to without the learner being aware.

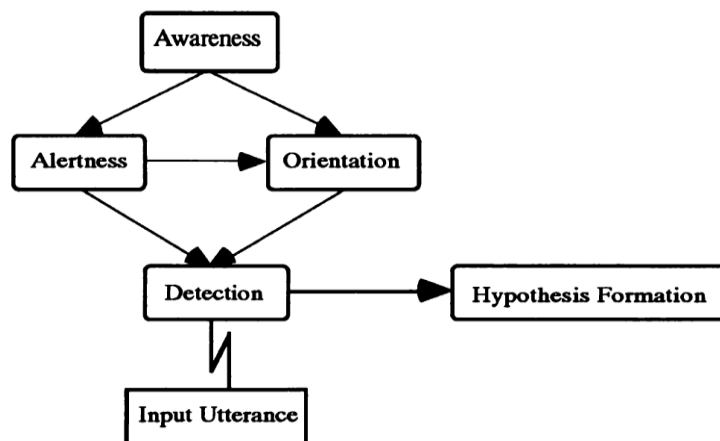


Figure 4. Relationship Between Awareness and the Three Components of Attention in Input Processing from Tomlin and Villa's (1994) Model of Attention (p. 197)

In summary, the aforementioned theories lend theoretical support to the possibility of learning new lexical items incidentally from language input. By pointing out the possibility of learning vocabulary from comprehensible input and highlighting the necessity of input, Krashen (1982, 1985, 1989) established the foundations of incidental vocabulary learning. Schmidt (1990, 1993) further elaborated the learning process by bringing up the notion of *noticing*, which is key for input to become intake that can lead to learning. However, researchers have criticised the compound definition of “noticing” entailing both awareness and attention (Godfroid et al., 2013; Leow, 1997), making it less clear which one is necessary for learning to occur. Drawing from work in cognitive and neuroscience, Tomlin and Villa (1994) distinguished attention and awareness by claiming that attention itself is sufficient for learning to occur. In addition, there seems to be a consensus in cognitive psychology and SLA that attention is crucial for L2 learning to occur (e.g., Schmidt, 1995, 2001; Tomlin & Villa, 1994). Therefore, incidental vocabulary learning is possible even without the intention to learn, it can occur when learners are exposed to comprehensible input and target items are attended to.

### ***2.2.3. Empirical Studies on Incidental Vocabulary Learning and Influential Factors***

Empirical evidence for incidental vocabulary learning has been extensively provided in the L1 context. Many studies have shown that L1 speakers develop their vocabulary incidentally through language exposure (e.g., Gampe, Liebal, & Tomasello, 2012; Lenhart, Lenhard, Vaahtoranta, & Suggate, 2018; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Suggate, Lenhard, Neudecker, & Schneider, 2013). A meta-analysis conducted by Swanborn and de Glopper (1999) found that L1 learners acquired about 15 per cent of unknown words incidentally during normal reading.

Incidental vocabulary gains have also been documented in the L2 context, although gains tend to be modest when compared to those reported by L1 studies. Previous research has shown that new words can be learned incidentally via different modes of exposure, including reading (e.g., Chang, 2019; Elgort et al., 2018; Godfroid et al., 2018; Horst, Cobb, & Meara, 1998; Hulstijn, 1992; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Pitts, White, & Krashen, 1989; Webb, 2005), listening (e.g., de Vos, Schriefers, Nivard, & Lemhöfer, 2018; Duquette & Painchaud, 1996; Pavia et al., 2019; Van Zeeland & Schmitt, 2013; Vidal, 2003), and reading-while-listening (e.g., Brown et al., 2008; Chang, 2019; Chen, 2021; Malone, 2018; Teng, 2018; Vu & Peters, 2020; Webb & Chang, 2012; Webb & Chang, 2015b; Webb et al., 2013).

In general, these studies have shown that all aforementioned modes of input can lead to incidental vocabulary learning, but at different rates. These different learning rates can be explained by the effect that different factors have on the process. Peters (2020, p. 125) classified three main types of factors that affect L2 learners' learning of

single words: word-related factors (i.e., word properties), contextual factors (i.e., the use of words in context), and learner-related factors (i.e., learners' individual differences). While conducting a comprehensive review of all these factors is beyond the scope of this thesis, this section briefly reviews those that have been studied in the context of incidental vocabulary learning.

Regarding word-related factors, part of speech and word length (i.e., number of word letters) are two of the factors that have received research attention. Regarding part of speech, studies exploring incidental vocabulary learning through listening (Van Zeeland & Schmitt, 2013) and reading (Horst & Meara, 1999) have shown that nouns seem to be more likely to be learned than verbs and adjectives. As for word length, shorter words seem to be better learned than longer words when reading (Godfroid et al., 2018).

Concerning contextual factors, one of the most important factors that accounts for different learning rates is the frequency of occurrence of unknown words in the input. A recent meta-analysis synthesizing 26 empirical studies conducted by Uchihara, Webb, and Yanagisawa (2019) revealed a medium correlation ( $r = .34$ ) between frequency of occurrence and L2 incidental vocabulary learning. Although different studies have suggested different numbers of word encounters as sufficient for vocabulary learning to occur, there is a consensus that the more encounters with unknown words, the higher the likelihood that those words can be learned (e.g., Brown et al., 2008; Pavia et al., 2019; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Van Zeeland & Schmitt, 2013; Vidal, 2003, 2011; Waring, 2003; Webb, 2007). Some reading studies have suggested that 4–5 times is considered sufficient for learning a word's meaning to take place (Pigada & Schmitt, 2006), while 8–10 times could lead to a real increase and acceleration in learning (Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006;

Waring, 2003; Webb, 2007). However, researchers have also pointed out that even just one encounter may lead to measurable vocabulary learning, especially in form and meaning recognition (e.g., Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Uchihara et al., 2019; Webb, 2007). Additionally, Uchihara et al. (2019) have also pointed out that it is not only the frequency of occurrence itself, but also its interactions with other factors, such as the differences in learners, treatment, and research methodology, that can affect L2 learners' incidental vocabulary learning.

Another vital contextual factor that has received considerable attention in the incidental vocabulary learning literature is the different presentation modes of words. Research comparing different modes shows that both reading-while-listening (e.g., Brown et al., 2008) and reading-only (e.g., Brown et al., 2008; Vidal, 2011) are superior to listening-only for incidental vocabulary learning. Moreover, reading-while-listening appears to be more beneficial than reading-only, but its superiority might be constrained by other factors, since the significant advantage of reading-while-listening has been reported in some studies (Chen, 2021; Malone, 2018; Webb & Chang, 2012), but not in others (Brown et al., 2008; Vu & Peters, 2020).

Concerning learner-related factors, learners with a higher proficiency level have been shown to acquire more vocabulary in both reading and listening studies (e.g., Malone, 2018; Vidal, 2003, 2011). Similarly, L2 learners with a larger vocabulary size have been found to achieve higher incidental learning gains in reading (e.g., Horst et al., 1998; Webb & Chang, 2015a).

In addition to the different factors influencing L2 learners' learning of single words, these different learning rates can also be attributed to the use of different vocabulary tests. Most of the empirical studies exploring incidental vocabulary learning have measured vocabulary knowledge by focusing on the four degrees of the form-meaning

link mentioned in section 2.1.2. In general, the greatest learning gains were noted in form recognition and meaning recognition, while the lowest gains were reported in meaning recall (e.g., Brown, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Van Zeeland 2013; Vu & Peters, 2020; Waring, 2003), echoing the difficulty levels proposed by Laufer and Goldstein (2004). Moreover, studies employing delayed posttests have also witnessed lower incidental learning gains than immediate posttests (e.g., Brown et al., 2008; Van Zeeland & Schmitt, 2013, 2011; Vidal, 2011; Vu & Peters, 2020; Waring, 2003).

It should be noted that the aforementioned factors are usually intertwined and interrelated to each other. For example, the benefits of different types of input modes may vary for learners with different proficiency levels. Vidal (2011) found that although learners could learn more vocabulary in a reading-only condition than in a listening-only condition, the advantage of reading-only was especially obvious for low-proficiency learners. This was potentially due to low-level learners' poor listening ability to segment real-time speech, which made learning in the listening-only condition more demanding. Research has shown that the role of frequency of occurrence is also modulated by the input mode. Due to the difficulty in separating aural real-time messages, learners in the listening-only condition were less likely to take advantage of the higher repetition of unknown words than learners in the reading-only condition, where written forms were available (Vidal, 2011).

This section has provided a brief review of the main findings of studies examining incidental vocabulary learning from reading, listening, and reading-while-listening. Of particular relevance to this thesis are those empirical studies that have examined incidental vocabulary learning while viewing. They are reviewed in detail in the following section.

### **2.3. Incidental Vocabulary Learning from Viewing**

With the rapid development of the Internet and globalization, recent decades have witnessed an unprecedented increase in the quantity and types of authentic audio-visual materials that are freely available. Television programmes, films and documentaries in different languages are widely available online for entertainment and study use, and learners can easily access these materials (Montero Perez, 2020b; Rodgers, 2013; Rodgers & Webb, 2011). Learners also seem to be motivated to watch foreign language films or videos, which may reduce their anxiety, increase their frequency of exposure to authentic L2 input, and raise their interest in language learning (Rodgers & Webb, 2011; Webb, 2011). Viewing has thus been advocated as a valuable type of out-of-class exposure that can increase learners' exposure to authentic L2 input and facilitate their vocabulary learning (Webb, 2020a). Researchers' interest in examining the potential of using audio-visual materials to facilitate language learning started in the 1980s and '90s (Vanderplank, 2010) and witnessed a second boost around 2009–2010 (Montero Perez, 2020b; Peters & Muñoz, 2020).

#### ***2.3.1. Theoretical Support for Learning from Viewing***

Audio-visual materials are a classic example of multimedia that contain dynamic pictures presented in pictorial form and authentic language input presented in verbal form (Mayer, 2005b). The combination of both verbal and non-verbal information is believed to facilitate viewers' understanding, information retention and language learning (Desmet & Cornillie, 2012; Kuppens, 2010; Niegemann & Heidig, 2012). This is supported by Dual Coding Theory (Paivio, 1986, 2007) and the Cognitive Theory of Multimedia Learning (Mayer, 2005a, 2009). Cognitive Load Theory (Chandler &

Sweller, 1991; Sweller, 1988, 1994, 2005a) and the redundancy principle (Sweller, 2005b) grounded in the Cognitive Theory of Multimedia Learning can also help to explain the potential pitfalls of using multimedia to facilitate learning. These relevant theories are now reviewed in turn.

### 2.3.1.1. Dual Coding Theory

The most general assumption of Dual Coding Theory (Paivio, 1986) is that there are two functionally independent but interconnected systems which handle different types of information. One is a *verbal system*, which handles language, and the other is an *imagery system*, which deals with non-linguistic objects and events (Paivio, 1986) (see Figure 5). These two systems can function independently (with only one active or both active in parallel) or cooperatively (interconnected to each other).

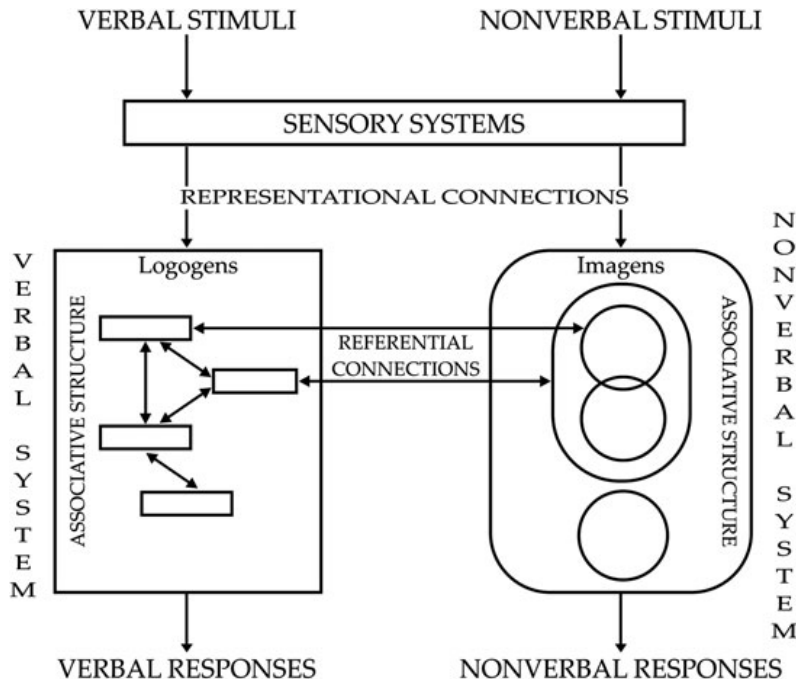


Figure 5. Dual Coding Theory (Paivio, 1986, p. 67)

According to Paivio (1986, 2007, 2014), activation happens after external stimuli are received by the corresponding *logogens* (i.e., verbal representations) or *imagens* (i.e., non-verbal representations). Relatively direct connections are made when the verbal stimulus or an object matches logogens or imagens correspondingly and they are available for further processing. *Referential connections*, i.e., indirect activation between imagens and logogens, are made where the representations of one system are activated in the other. Associative connections trigger indirect activation through within-system connections between either logogens or imagens. Finally, verbal or non-verbal responses are generated as output.

When information is dual-encoded, additive effects can be stimulated since information is activated in both systems, leading to better memory retention (Danan, 1992; Paivio, 1986). Pictures were found to contribute about twice as much as verbal codes to additive effects, which may help to alleviate the difficulty in remembering information (Paivio, 2007). Therefore, providing information in both verbal and pictorial forms can increase the opportunities to activate both verbal and imagery systems. The two systems are able to support each other, which is believed to facilitate information processing and augment information recall (Paivio, 1986, 2007).

A bilingual version of Dual Coding Theory has also been proposed (Paivio, 1986; Paivio & Desrochers, 1980) by separating the two verbal systems ( $V_1$  and  $V_2$ ) which correspond to a bilingual's two languages ( $L_1$  and  $L_2$ ). As illustrated in Figure 6, the imagery system and two verbal systems are assumed to be able to function independently but can also interconnect with each other through associative verbal connections ( $V_1 - V_2$ ) or through referential connections between verbal and non-verbal systems. According to this theory, it is believed that the use of translation can engage two separate memory stores and enhance the individual's memory recall (Danan, 1992;

Paivio, 2014). Also, by having an additional language which is interconnected to the first one, the benefits of combining verbal and non-verbal systems can be augmented (Paivio, 2007). It is thus believed that the use of L1 subtitled video should be very powerful for information retrieval and vocabulary learning (Danan, 1992; Danan, 2004). With images, soundtrack in one language and text in another conveying the same information, they provide a stronger connection for the individual's information processing and can offer additive effects of both image and translation (Danan, 1992, 2004).

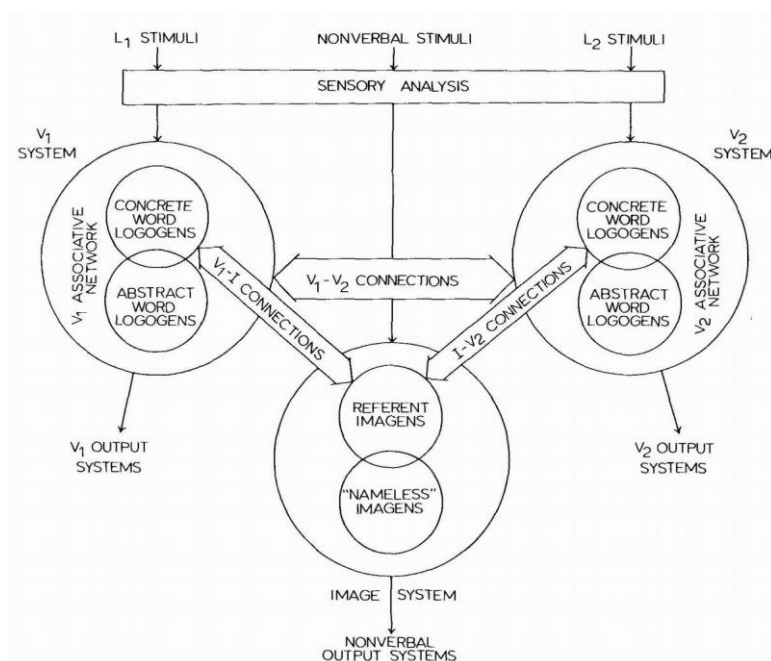


Figure 6. Bilingual Version of the Dual Coding Model (Paivio & Desrochers, 1980, p. 391)

### 2.3.1.2. Cognitive Load Theory

However, care should be taken when using multimedia to facilitate learning. When dealing with novel information, our working memory is severely limited in its capacity and duration (Sweller, 2005a). Focusing on how cognitive resources are used during learning and problem solving, Cognitive Load Theory suggests that effective

instructional material facilitates learning by not overloading our limited cognitive capacity (Sweller, 1988).

According to Sweller (2005b), there are three categories of cognitive load that need to be either reduced or increased to facilitate learning: intrinsic, extraneous, and germane. *Intrinsic cognitive load* is determined by the natural complexity of the information needing to be processed. *Extraneous cognitive load* refers to the appropriateness of instructional design for learners. It is determined by whether information is presented by considering learners' working memory limits and focusing working memory resources on schema construction and automation. Schema acquisition and automation are two principal learning mechanisms that can ultimately reduce the cognitive load and facilitate learning novel knowledge (Sweller, 1994). A *schema* is "a cognitive construct that organizes the elements of information according to the manner with which they will be dealt" (Sweller, 1994, p. 296). It determines how new information is processed, and increases the amount of information that can be stored in working memory. *Automation* refers to how effortlessly information can be processed without conscious control. Information processed with high automation requires less memory space and can free up cognitive capacity for other functions. *Germane cognitive load*, which is "effective" cognitive load, is "caused by effortful learning resulting in schema construction and automation" (Sweller, 2005b, p. 27). It is the working memory resources used to organise and integrate new information with pre-existing knowledge. For example, providing learners with a number of examples to demonstrate a point could increase their germane cognitive and facilitate schema construction (Sweller, 2005b).

Therefore, the design of multimedia learning materials (especially for information with a higher intrinsic cognitive load) should aim to lower the extraneous cognitive load

and increase the germane cognitive load to enable effective learning to occur. Cognitive Load Theory holds that the use of multimedia may not always be effective for learning. The presentation of learning materials should take learners' limited cognitive capacity into consideration by reducing their working memory load and actively engaging with learners' prior knowledge to facilitate learning.

#### **2.3.1.3. Cognitive Theory of Multimedia Learning**

Based on the multimedia principle that “people can learn more deeply from words and pictures than from words alone” (Mayer, 2005b, p. 1), the Cognitive Theory of Multimedia Learning explains how learners process information in a multimedia environment. This theory makes three basic assumptions: a dual-channel assumption, a limited capacity assumption, and an active processing assumption. These three assumptions incorporate elements from Paivio's (1986, 2007) Dual Coding Theory, Sweller's (1988, 2005a; Chandler and Sweller, 1991) Cognitive Load Theory, Baddeley's (1986, 1999) Model of Working Memory with an emphasis on limited processing capacity, and Atkinson and Shiffrin's (1968) memory stores and cognitive processes.

Similar to Dual Coding Theory, the dual-channel assumption is that the auditory/verbal channel is responsible for information entering the ears, and the visual/pictorial channel copes with information entering the eyes. The limited capacity assumption is that humans can only process a limited amount of information in each channel at a time. Therefore, instead of processing all available input, we are forced to make decisions about the allocation of our attention and the degree of organising and integrating information. The active processing assumption is that instead of passively receiving information, humans are active processors who attempt to make sense of

information (Mayer, 2005b). In this active process, learners first select relevant material, then organise it into verbal and non-verbal mental representations, and integrate the selected material with their prior knowledge in working memory.

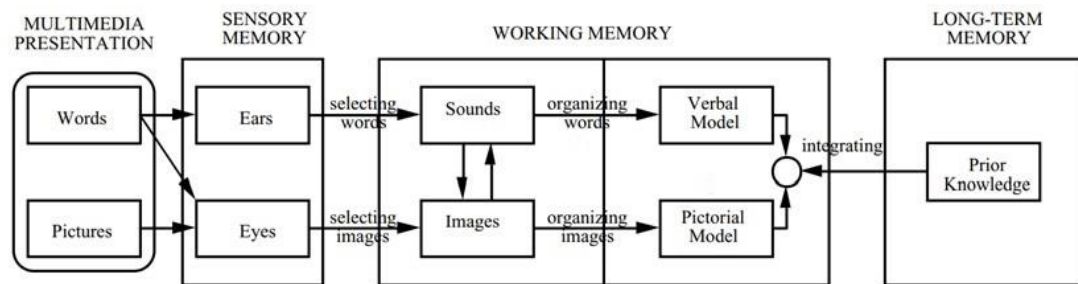


Figure 7. Cognitive Theory of Multimedia Learning (Mayer, 2005a, p. 37)

Figure 7 illustrates five cognitive processes that enable meaningful learning to occur in a multimedia environment: 1) selecting relevant words for processing in verbal working memory; 2) selecting relevant images for processing in visual working memory; 3) organizing selected words into a verbal model; 4) organizing selected images into a pictorial model; and 5) integrating verbal and pictorial representations with each other along with prior knowledge (Mayer, 2005a). These processes do not necessarily occur in a linear order, and they are more likely to occur segment by segment, many times, during a multimedia presentation.

Based on this theory, the use of both visual and verbal presentations of information can activate both systems, which potentially leads to deeper and greater cognitive processing. Once new knowledge is constructed in working memory, it is stored in long-term memory for later use to support new learning (Mayer, 2005a). Different from Paivio's (1986) Dual Coding Theory which mainly emphasises the benefits of presenting information in verbal and pictorial forms to facilitate learning, the Cognitive Theory of Multimedia Learning also points to the importance of considering learners' limited cognitive capacity and learners' active role in multimedia learning.

Mayer (2009) proposed 12 principles grounded in the Cognitive Theory of Multimedia Learning, these aim to boost the advantages of multimedia learning. They include five for reducing extraneous cognitive load – coherence, signalling, redundancy, spatial contiguity, and temporal contiguity; three for managing intrinsic cognitive load – segmenting, pre-training, and modality; and four for fostering germane cognitive load – multimedia, personalization, voice, and images. Of these principles, multimedia and redundancy are particularly relevant for learning from subtitled viewing, which are investigated in the present study.

According to Mayer (2009, Chapter 12), the basis of the multimedia principle is that learning and understanding are boosted more by presenting both verbal and pictorial information than presenting verbal information alone. By adding pictorial information to verbal information, learners' verbal and visual channels are both primed, which facilitates the establishment of mental connections between these two different representation systems. The integration of information can facilitate learning. Especially, this principle may apply more to learners with low prior knowledge than to learners with high prior knowledge of the information presented (Mayer, 2009, Chapter 12). Therefore, the use of audio-visual materials to support learning seems to be very promising, as both verbal input (spoken or written) and visual input (i.e., dynamic images) can activate both learners' representation systems and further enhance learning outcomes.

The redundancy principle suggests that redundant material impedes learning, and so better learning can be facilitated by the removal of redundant information (Mayer, 2009, Chapter 6). The redundancy effect occurs when “additional information presented to learners results in learning decrements compared to the presentation of less information” (Sweller, 2005b, p. 159). In other words, redundancy can occur in either

the presentation of identical information in different forms/media, or the presentation of additional information in the same form as elaborate information. A common scenario is adding redundant printed text to narrated audio/video, which might require extraneous processing resulting in cognitive overload and thus be detrimental for learning (Chandler & Sweller, 1991; Mayer, Lee, & Peebles, 2014; Sweller, 2005b) . However, when introducing these multimedia learning principles to L2 learning contexts, it has been repeatedly argued that care should be taken since those principles were originally put forward for domain learning in an L1 context (e.g., Montero Perez, 2020b; Mayer et al., 2014).

In the context of learning academic content in learners' L2, in contrast to the redundancy principle, the Redundancy Facilitation Hypothesis (Mayer et al., 2014) has recently been proposed and suggests “a reverse redundancy effect in scenarios where redundant material can support basic cognitive processing that is not yet automated in non-native speakers while minimizing extra cognitive load” (p. 654). An example is adding redundant video to audio which could help L2 learners to access word meanings (Mayer et al., 2014). It can also be argued that the provision of redundant written information, for example adding on-screen text to L2 viewing, can support L2 learners' cognitive processing of video if it is attended to. Moreover, on-screen text can be helpful for raising L2 learners' noticing of unfamiliar input and compensate for their less-advanced listening ability.

As pointed out by Sweller (2005b), whether the information is redundant should always be discussed in different situations and with different subjects, as there is no universally applicable standard for redundancy. The benefits of multimedia materials for L2 learning can only be achieved by taking learners' cognitive experience and ability into account, to ensure that multimedia can actually enhance the input and

increase learners' likelihood of noticing, rather than acting as a potential burden on learners' cognition.

In sum, Dual Coding Theory and the Cognitive Theory of Multimedia Learning offer theoretical explanations for the cognitive processing of multimedia information and point to the potential of learning from multimedia materials. Audio-visual materials containing images (i.e., non-verbal information) that are presented visually and verbal information that is presented aurally can simultaneously activate both imagery and verbal systems and thus further enhance learning outcomes. According to the bilingual version of Dual Coding Theory, viewing with L1 translations can potentially enhance the interconnections between the nonverbal and two verbal systems, enabling more access routes to be established for information retrieval. Information is thus more likely to be activated through the three systems which facilitate learning. However, according to Cognitive Load Theory and the redundancy principle, the L1 and L2 lines basically convey the same meaning in different forms, which may overload learners' cognitive capacity for information processing and thus impede their learning. Since these theories were originally put forward in an L1 context, care should be taken when extending them to an L2 context. The next section reviews relevant empirical studies to obtain an overview of the effects of L2 viewing and the use of on-screen text on L2 incidental vocabulary learning.

#### **2.3.1.4. Empirical Studies on Incidental Vocabulary Learning from Viewing**

Many empirical studies have attempted to explore the effectiveness of audio-visual materials for L2 learning, mainly focusing on vocabulary learning, comprehension, and grammar learning. They have found that images seen while viewing can facilitate learners' comprehension (e.g., Durbahn, Rodgers, & Peters, 2020), and the use of

on-screen text may further support this benefit (e.g., Montero Perez et al., 2013; Pujadas & Muñoz, 2020; Rodgers & Webb, 2017). Some studies have also found that viewing with on-screen text can facilitate L2 grammar learning (e.g., d'Ydewalle & Van de Poel, 1999; Lee & Révész, 2018, 2020; Pattemore & Muñoz, 2020).

Most relevant to this research are the effects of viewing on incidental vocabulary learning. The potential of using audio-visual materials to facilitate learners' vocabulary development first started to attract researchers' attention early in the 20<sup>th</sup> century. In the L1 context, Rice and Woodsmall (1988) found that preschoolers could comprehend an average of five new words after watching two short animated programmes, as revealed in a picture recognition test. The benefits of viewing for the development of young children's vocabulary were also later documented in a two-year longitudinal study conducted by Rice, Huston, Truglio, and Wright (1990).

This learning potential was later recognised in L2 vocabulary learning with several advantages being observed. First, the language used in authentic videos and films can be considered natural and original language input for L2 learners (Coady & Huckin, 1997; Rodgers & Webb, 2011), which can compensate for the insufficient L2 input that EFL learners in many contexts have. Second, watching audio-visual materials has been found to be favoured by a range of L2 learners (e.g., Peters, 2018; Rodgers & Webb, 2011; D. Wang, 2012; Webb & Rodgers, 2009; Yang & Chang, 2013). Viewing enjoyment can lower learners' affective filter and encourage acquisition to take place (Krashen, 1982). Third, different from text reading, viewing also provides the auditory form of new words with correct pronunciation and intonation (Coady & Huckin, 1997). In addition, images can help the viewer to make more accurate inferences of word meanings (Duquette & Painchaud, 1996; Peters, 2019; Rodgers, 2018), which may lead to better understanding and retention of words. Last but not least, the lexical coverage (i.e.,

percentage of known words in discourse) threshold for audio-visual materials to facilitate vocabulary learning seems to be lower than for reading. By analysing lexical coverage, genre, word types, and word occurrences in a variety of English television programmes, corpus-driven studies found that knowledge of the 3,000 most frequent word families plus proper nouns and marginal words resulted in 95% lexical coverage of television programmes, which could lead to adequate comprehension and might be sufficient for incidental vocabulary learning to occur (Rodgers & Webb, 2011; Webb, 2011; Webb & Rodgers, 2009). In other words, learners who master the 3,000 most frequent word families should be able to understand television programmes and benefit from viewing.

A number of empirical studies have documented the positive effects of viewing for L2 vocabulary learning. Vocabulary can be learned incidentally while viewing as little as one TV episode or programme. Puimège and Peters (2019) found that Dutch EFL learners could learn 13% of words at the level of form recall after watching a 30-minute excerpt from an English TV programme. Montero Perez (2020a) reported that Dutch-speaking learners of French learned 17% of target pseudowords, as measured by a spoken word recognition test, scoring 43% on form recognition, 20% on meaning recognition, and 2% on meaning recall after watching a 25-minute French documentary. These benefits also extended to studies employing longer viewing materials. After watching a full-length 54-minute English documentary, Chinese EFL learners in a study by Feng and Webb (2020) incidentally learned 15% of single words at the level of form recognition and 10% for meaning recognition on average. Peters and Webb (2018) revealed average learning gains for single words of 8% and 14% for meaning recall and meaning recognition, respectively, among Dutch-speaking EFL learners after watching a one-hour BBC documentary. In addition, Rodgers and Webb (2019) found that

Japanese EFL learners answered on average 11% of single words correctly in a meaning recognition test after watching ten episodes (7+ hours in total) of a television programme. Overall, these studies found that watching audio-visual material led to vocabulary gains despite the length of the material. In general, higher gains were reported for receptive knowledge than productive knowledge, and form knowledge was more likely to be acquired than meaning. It should be noted, however, that learning gains were in general inconsistent across different studies, they varied from 15% to 43% for form recognition, 10% to 20% for meaning recognition, around 10% for form recall, and below 10% for meaning recall.

In order to explain such variation, some studies on learning through viewing have also looked at the role of different factors in the process. Regarding word-related factors, Puimège and Peters (2019) found that words that were more concrete or with a higher corpus frequency were learned better. Moreover, contradicting previous findings in reading research (e.g., Godfroid et al., 2018), during uncaptioned viewing, words with more syllables were more likely to be learned (as measured by form recall) than words with fewer syllables. It can be argued that longer words might be more noticeable than shorter words, given the real-time nature of the input (Puimège & Peters, 2019). In addition, word frequency of occurrence also has a positive effect on vocabulary learning while viewing (e.g., Peters & Webb, 2018; Peters, 2019; Rodgers & Webb, 2019). Concerning learner-related factors, learners' vocabulary size also reveals significant positive effects on different aspects of vocabulary knowledge, including form recognition (e.g., Feng & Webb, 2020; Montero Perez, 2020a; Montero Perez, Peters, Clarebout, & Desmet, 2014), form recall (e.g., Puimège & Peters, 2019), meaning recognition (e.g., Feng & Webb, 2020; Montero Perez, 2020a; Montero Perez et al.,

2014; Peters & Webb, 2018), and meaning recall (e.g., Montero Perez et al., 2014; Peters & Webb, 2018).

Feng and Webb (2020) recently compared the effects of viewing with reading-only and listening-only conditions. They found that Chinese EFL learners in all conditions achieved significant incidental vocabulary learning gains, as measured by form and meaning recognition, but no significant differences were observed between groups. Their results suggest that viewing is at least as effective as unimodal input for incidental vocabulary learning. However, it should be noted that participants in the viewing condition watched video without on-screen text, which is not common for Chinese EFL learners, as also noted by the researchers. The benefits of viewing over reading and listening might be more salient with the presentation of on-screen text. Studies examining the effectiveness of viewing using on-screen text are reviewed in the following sections.

### ***2.3.2. Use of Captions and L1 Subtitles while Viewing***

Recent decades have witnessed an increase in the number of studies examining the use of captions and L1 subtitles to facilitate learners' vocabulary learning, as these forms have been claimed to be the ones most frequently encountered by foreign language learners (Lunin, 2015; Muñoz, 2017). *Captions* are transcripts written in the same language as the text spoken in the video and appear simultaneously at the bottom of the screen (Chung, 1999). They were originally designed for the deaf and hearing-impaired. Captions are also called *teletext subtitles*, *closed captions*, *same language subtitles* in different contexts (Vanderplank, 2010). Captions have also been labelled as *bimodal*, *unilingual*, *intralingual*, or *L2 subtitles* in academic works (Danan, 2004; Montero Perez et al., 2013; Muñoz, 2017). *L1 subtitles* are on-screen text

translated into the viewer's L1 (Markham, Peter, & McCarthy, 2001). They have also been called *subtitles*, *native language subtitles*, or *interlingual subtitles* in scholarly literature (Danan, 2004; Li, 2016; Markham et al., 2001). In the present thesis, the terms *captions* and *L1 subtitles* are used.

The effectiveness of captions and L1 subtitles has been examined in studies comparing a captions/L1 subtitles condition with a no-subtitles condition. Section 2.3.2.1 briefly reviews studies comparing captions with no subtitles. Then, research comparing L1 subtitles with no subtitles is summarised in section 2.3.2.2. Studies comparing different types of subtitles (i.e., captions, L1 subtitles, no subtitles) are reviewed in section 2.3.2.3.

#### **2.3.2.1. Use of Captions in Viewing**

Captions are believed to be helpful for facilitating vocabulary learning by visualising viewers' listening, aiding speech decoding and segmentation, and bridging the gap between learners' lagging listening comprehension skills and more developed reading comprehension skills (Danan, 2004, Garza, 1991; Lunin, 2015; Montero Perez et al., 2014). They may also increase learners' attention to unknown word forms which helps word recognition and vocabulary building (Danan, 2004; Markham, 1999; Winke, Gass, & Sydorenko, 2010). Furthermore, captions are believed to be beneficial for making form-meaning connections in the mental lexicon (Winke et al., 2010), and facilitating comprehension (e.g., Baltova, 1999; Garza, 1991; Markham & Peter, 2003; Markham et al., 2001; Montero Perez et al., 2013; Winke et al., 2010).

Price (1983) provided the earliest evidence showing the benefits of using captions to facilitate L2 learners' comprehension and language learning. Five-hundred English as a Second Language (ESL) students at a U.S. university with a variety of L1

backgrounds were assigned to a captions group or a no-captions group to watch four English excerpts once or twice. A comprehension test was administered after their viewing. The results showed that all participants benefited significantly from the use of captions, even after only one viewing. However, no details about the participants' proficiency, the stimuli selected, the design of the comprehension test, or the data analysis results are available. Thus, to what extent captions were better than no subtitles and how the benefits can be generalised to learners with different backgrounds and different stimuli remain unanswered. Moreover, the researcher only assessed comprehension by assuming that better comprehension contributed to language learning. Therefore, it is unclear which aspects of language learning might benefit from captioned viewing.

In the past three decades, many studies have been conducted to fill in those research gaps. In a meta-analysis, Montero Perez et al. (2013) analysed 18 empirical studies exploring the use of captions for L2 listening comprehension and vocabulary learning. Their findings revealed significant large effects of captioning on listening comprehension ( $g = 0.99$ ) and on vocabulary learning ( $g = 0.87$ ), despite the different vocabulary tests used across studies and differences in participants' proficiency. The authors thus claimed that L2 learners can benefit from captioned viewing for both comprehension and vocabulary learning, as long as the video materials used are appropriate for their proficiency level. However, as pointed out by the authors, their findings were only based on a limited number of research studies with uneven distributions of L2 learners' proficiencies, calling for more studies to investigate potential moderator variables affecting learning outcomes.

The benefits of captions for incidental vocabulary learning seem to be retained regardless of different viewing materials and L2 learners. The benefits have been

well-documented in both multiple viewing sessions where learners were asked to watch captioned videos each week for several consecutive weeks (e.g., Neuman & Koskinen, 1992; Vanderplank, 1988), and in single viewing sessions ranging from 10 to 40 minutes (e.g., Chai & Erlam, 2008; Jelani & Boers, 2018; Markham, 1999; Montero Perez et al., 2014, 2018; Sydorenko, 2010; Teng, 2018; Winke et al., 2010). Learners can also benefit from watching a range of captioned viewing materials, for example, TV programmes (e.g., Markham, 1999; Montero Perez et al., 2014, 2018; Neuman & Koskinen, 1992; Vanderplank, 1988), animations (e.g., Chai & Erlam, 2008; Teng, 2018), TV series (e.g., Sydorenko, 2010), TED talk (e.g., Jelani & Boers, 2018), and documentaries (e.g., Winke et al., 2010). The effectiveness of captioned viewing may also extend to learners with different L1 backgrounds and L2 proficiency levels (e.g., Vanderplank, 1988; Winke et al., 2010).

Most studies comparing the presence and absence of captions found that the use of captions tended to be more effective than no subtitles for form recognition (e.g., Markham, 1999; Montero Perez et al., 2014, 2018; Neuman & Koskinen, 1992; Teng, 2018), meaning recognition (e.g., Neuman & Koskinen, 1992; Teng, 2018), and meaning recall (e.g., Jelani & Boers, 2018; Sydorenko, 2010; Teng, 2018; Winke et al., 2010). Learning gains reported in the captions groups were significantly higher than those in the no subtitles groups, with a 5% to 43% increase in form recognition, 5% to 32% for meaning recognition, and 6% to 30% for meaning recall. The different learning gains reported in different studies could be attributable to the different research designs (e.g., one-time or two-time viewing), TW selection (e.g., single words or a combination of single words and multiword units; different word frequencies of occurrence), participants' profiles (e.g., children or adult learners), and the possible interaction between different factors. Moreover, studies also employed different scoring methods,

with some taking into account participants' answers to distractors in the analysis (e.g., Sydorenko, 2010), while others did not (e.g., Teng, 2018), which might also result in different learning gains. Despite these variations, in general, captions seemed to facilitate vocabulary learning, particularly at the level of form recognition.

Several studies reviewed in this section not only compared captions with no subtitles, but also examined the effects of other types of captions (e.g., keyword captions, glossed captions; Montero Perez et al., 2014, 2018; Teng, 2018). These studies are reviewed later in section 2.3.2.3.

#### **2.3.2.2. Use of L1 Subtitles in Viewing**

Similar to captions, the use of L1 subtitles is also believed to alleviate learners' anxiety over missing important information during viewing and boost learners' motivation (Danan, 2004). Compared to captions, whose benefits for vocabulary learning might be constrained when viewing materials are beyond the linguistic competence of learners, L1 subtitles seem to be more advantageous for lower level learners (Danan, 1992, 2004). Moreover, L1 subtitles can overcome the difficulty of having to rapidly process L2 lines in a video, which can lead to better comprehension than reading in L2 (Danan, 2004).

For vocabulary learning, translations can reduce the chances of misleading learners about the correct meaning of unknown words (Nagy & Herman, 1987; Nation, 2001), and promote the retention of correct word meanings at least among novice L2 learners (Grace, 1998). However, the benefits of L1 subtitles for L2 vocabulary learning are still controversial. Some researchers argue that learners using L1 subtitles may be more inclined to read L1 subtitles rather than actively make the link between L1 translations and auditory word forms (Lunin, 2015; Peters, 2019; Peters et al., 2016), whereas others

have pointed out that the automatic reading of L1 subtitles does not prevent the processing of a foreign language soundtrack and can still support vocabulary learning (e.g., d'Ydewalle & Pavakanun, 1997; Danan, 2004; Koolstra & Beentjes, 1999).

Empirical studies exploring the benefits of using L1 subtitles for L2 vocabulary learning are relatively scarce, and most of them have been conducted on children. d'Ydewalle and colleagues (1995, 1997, 1999) were pioneers in examining the effects of subtitled and dubbed viewing on L2 language learning. They conducted a series of studies examining the effects of different subtitling and dubbing types on incidental vocabulary learning and comprehension. Their findings showed that after watching subtitled or dubbed videos, adults and children without prior knowledge of the L2 could incidentally learn L2 vocabulary by choosing correct L2 word forms for L1 translations. In general, compared to viewing an L2 soundtrack without subtitles, the use of L1 subtitles was found to lead to a 9% to 20% increase in correctly matching an L2 form to an L1 translation.

Koolstra and Beentjes (1999) also found that after watching a 15-minute American documentary clip with L1 subtitles, Dutch-speaking children learned significantly more words than a no subtitles condition, as measured by form and meaning recognition tests. However, the results should be treated with caution since the control group (with L1 audio and no subtitles), who were not exposed to the L2, also made relatively high meaning recognition gains, reflecting the effect of guessing when completing a multiple-choice test. Moreover, no pretests were included to control for learners' prior knowledge of L2 TWs. Findings from earlier research seem to suggest an advantage for L1 subtitles over no subtitles for incidental vocabulary learning. However, the aforementioned studies were conducted on learners without or with very limited prior knowledge of the L2. Moreover, the L1 and L2 studied in most of these studies used

Latin scripts (e.g., English, French, Danish, Dutch). Thus, it is unclear whether the benefits of L1 subtitles are retained when the language of the video soundtrack and subtitles are from different writing systems.

In a more recent study, Fazilatfar, Ghorbani, and Samavarchi (2011) found that Iranian EFL learners who mastered the 2,000 most frequent word families could benefit from watching a 15-minute English animation with Persian subtitles. Participants' word meaning recognition and meaning recall gains increased by about 13% more in the L1 subtitles condition than in the no subtitles condition. However, no significant differences were found for form recognition, which casts some doubt on the effects of L1 subtitles facilitating form knowledge. Therefore, L1 subtitles seem to be beneficial for learning word meaning, even if the languages used in the subtitles and soundtrack are in different writing systems. However, due to the limited number of studies conducted on experienced L2 learners, it is arguable whether the advantages of L1 subtitles remain salient for L2 learners with different proficiency. Moreover, the viewing materials used in most studies comparing L1 subtitles and no subtitles were limited to short animations (e.g., Fazilatfar et al., 2011; d'Ydewalle & Pavakanun, 1995) or still motion-picture films (e.g., d'Ydewalle & Pavakanun, 1997; d'Ydewalle & Van de Poel, 1999). This potentially limits the generalisation of their research findings to other types of viewing materials of different lengths. The next section reviews studies comparing captions and L1 subtitles. Apart from examining the relative effectiveness of these two subtitling types, these studies also broadened the selection of stimuli and involved participants with different proficiencies, which can provide more insights into the effects of captions and L1 subtitles on incidental vocabulary learning.

### **2.3.2.3. Empirical Studies Comparing Captions and L1 Subtitles**

To date, empirical studies comparing the effects of captions and L1 subtitles have shown that L1 subtitles lead to better comprehension than captions (e.g., Bianchi & Ciabattini, 2008; Birulés-Muntané & Soto-Faraco, 2016; Lwo & Lin, 2012; Markham et al., 2001; Markham & Peter, 2003), which is reasonable given the advantage of learners' L1 proficiency. However, their comparisons in relation to incidental vocabulary learning have yielded inconclusive findings. While some studies found that captions were more effective than L1 subtitles for incidental vocabulary learning (e.g., Frumuselu, De Maeyer, Donche, & Colon Plana, 2015; Naghizadeh & Darabi, 2015; Peters, 2019; Peters et al., 2016), others found no significant difference between them (e.g., Birulés-Muntané & Soto-Faraco, 2016; Frumuselu, 2018; Peters et al., 2016; Pujadas & Muñoz, 2019).

Naghizadeh and Darabi (2015) conducted a multiple viewing session study on 27 15- to 17-year-old intermediate level Iranian EFL learners to compare the effects of captions and L1 subtitles on vocabulary learning. Participants were asked to watch a 90-minute English film in six sessions over four weeks in one of three subtitling conditions (e.g., captions, L1 subtitles, no subtitles). In each viewing session, participants watched a 15-minute film excerpt three times in class. Forty target words were selected based on participants' prior knowledge using a multiple-choice test before the treatment. Their learning gains were measured again after the last viewing. Posttest results showed that the captions group significantly outscored the other two groups, whereas no differences were found between the L1 and no subtitles groups. Their findings showed the superiority of captions over L1 subtitles for vocabulary learning. However, care should be taken when interpreting their findings. First, it was unclear which knowledge aspects the vocabulary test measured. Moreover, participants were asked to watch each film excerpt three times in class, which might poorly represent the

natural viewing process and lead to more attention being paid to the language rather than the content, resulting in higher gains. Additionally, word-level differences (e.g., frequency of occurrence, word class) were not controlled for in their analysis. Lastly, the findings were based on a very limited number of participants, which might thus not be generalisable to a larger population.

Later, Peters et al. (2016) conducted two exploratory studies with a more rigorous research design to compare the effectiveness of captions and L1 subtitles for EFL learners' incidental vocabulary learning. In the first experiment, 28 Dutch-speaking EFL learners from a secondary school who had mastered the most frequent 2,000 words were divided into captions and L1 subtitles groups to watch a 13-minute documentary twice. Pre- and posttests including a spoken form recognition test and a spoken meaning recall test were administered before and immediately after their viewing. Participants' vocabulary size and the frequency of occurrence of 39 target items (most of which were single words with two compounds and two phrasal and reflexive verbs) were taken into account in the analysis. Posttests revealed that the captions group significantly outperformed the L1 subtitles group on spoken form recognition. However, no obvious group difference was observed for spoken meaning recall, suggesting that captions were more beneficial than L1 subtitles for learning word forms, but they were similarly effective for learning meanings. In the second experiment, 18 low-proficiency to pre-intermediate level EFL learners with various L1 backgrounds watched a 20-minute English animated sitcom clip with either captions or L1 subtitles. Target items (16 single words and 2 compounds) were presented in written form in three posttests (i.e., form recall, form recognition, meaning recognition). The findings indicated that the captions group only outperformed the L1 subtitles group on form recall. The L1 subtitles group performed better than the captions group on meaning recognition but the

difference was not statistically significant. Learners' vocabulary size and the frequency of occurrence of each target item were found to be significant positive predictors of learning gains in both studies. Peters et al. (2016) concluded that the results showed a partial advantage for captions over L1 subtitles for learning word forms, but provided no evidence that L1 subtitles were more beneficial for learning word meanings. Although informative, this study was limited in its small sample size and the lack of a control group to demonstrate the relative effectiveness of different subtitling types by comparing to no subtitles.

Addressing these limitations, Peters (2019) conducted another study with 142 Dutch-speaking intermediate level EFL secondary school learners. Participants were asked to watch an 11 minute 25 second documentary clip twice with one of three subtitling types: captions, L1 subtitles, or no subtitles (control group). Pre- and posttests including form recognition and meaning recall were administered before and immediately after viewing. Findings showed that the captions group significantly outperformed the L1 subtitles and no subtitles groups on both tests, whereas no significant difference was found between the L1 and no subtitles groups. Similar to the study conducted in 2016, learners' vocabulary size and the frequency of occurrence of TWs also correlated positively with vocabulary scores. Moreover, this study also revealed positive effects of other word-related factors (e.g., on-screen imagery support, cognateness, corpus frequency) on learning gains. The researcher thus concluded that captions were more helpful than L1 subtitles for incidental vocabulary learning, especially for intermediate to advanced level learners. For learners with lower proficiency, as can be seen in Peters' study conducted in 2016, captions still seemed to be better than L1 subtitles for facilitating form knowledge, but not for facilitating meaning knowledge.

The potential advantages of captions over L1 subtitles also extend to learning colloquial, phrasal verbs, and informal expressions in studies conducted by Frumuselu and colleagues (2015, 2018) on L2 learners who varied in proficiency. However, this superiority seemed to be observable only after long-term viewing, but was not revealed in an immediate posttest after each short viewing session. Since these studies only focused on learning multiword items, their methods and findings are not discussed in depth.

While the aforementioned studies suggest an advantage for captions over L1 subtitles for incidental vocabulary learning, others did not report significant differences. Bisson, van Heuven, Conklin, and Tunney (2014) explored the viewing behaviour of 54 L1 English speakers without prior knowledge of Dutch while watching a 25-minute animation in one of four conditions: captions (Dutch soundtrack, Dutch subtitles), L1 subtitles (Dutch soundtrack, English subtitles), no subtitles (Dutch soundtrack only), and reversed subtitles (English soundtrack, Dutch subtitles). Participants' incidental vocabulary learning was also compared to a control group without viewing. Participants' knowledge of 78 Dutch single words was examined immediately after viewing via an auditory vocabulary test (hear an L2 word and decide if the L1 translation presented is correct). However, no significant group differences were found. The authors attributed this finding to the less sensitive vocabulary test used in this study which failed to capture small vocabulary gains. The lack of group differences can also be explained by participants' lack of L2 knowledge, which restricted their ability to make a form-meaning link for unknown words. Moreover, as pointed out by the authors, the participants did not have much experience of reading subtitles, which might have further limited the benefits of them.

Working with experienced L2 learners, Birulés-Muntané and Soto-Faraco (2016) compared the effect of captions, L1, and no subtitles on L2 learners' speech perception, word meaning recognition, and comprehension. Sixty EFL university students with high-intermediate proficiency level were pseudo-randomly assigned to one of three subtitling groups to watch one episode of a British TV series. Their vocabulary development was examined via a listening speech perception test (listening and filling in 24 word gaps embedded in a short English text) and a meaning recognition test with 15 single TWs before and after viewing. There were two sets of materials for each test, and each participant used one or other in a pretest and the other in a posttest. In other words, half of the participants were tested with one set of material in a pretest and the other set of material in a posttest, whilst the other half of the participants had the pre- and posttest materials reversed. Eight comprehension questions were also asked afterwards. The results showed that for listening speech perception, both the captions and no subtitles groups revealed significant improvements after viewing, while no improvement was found for the L1 subtitles group. The captions group also outperformed the other groups. For word meaning recognition, only significant but modest pre-post improvements were observed in the no subtitles groups, with no significant group differences revealed. In terms of comprehension, the L1 subtitles group significantly outperformed the other two groups, and the captions group also significantly outperformed the no subtitles group. The findings suggest that captions may help to improve learners' listening speech perception ability, and L1 subtitles led to better comprehension, but neither of them facilitated learning word meanings. However, the findings should be treated with caution since the items measured in the pre- and posttests were different for each participant, thus, the analysis might not have well captured participants' learning gains after the treatment.

A recent study conducted by Pujadas and Muñoz (2019) explored the effects of different subtitling types and types of learning instruction on vocabulary learning during extensive viewing. One-hundred-and-six secondary school EFL students at a beginner to low-intermediate level participated in a one-year pedagogical intervention. Participants were divided into four classes to watch an English TV series during a whole academic year. The groups differed in the language of on-screen text (captions vs. L1 subtitles) and the presence/absence of pre-teaching of vocabulary items. Pre- and posttests were conducted before and after each viewing session. Participants' knowledge of 120 single TWs was measured through form recall (listening to spoken forms and writing down words) and meaning recall. The findings showed that the groups with pre-teaching performed better than the no instruction groups on both tests. For the no pre-teaching groups, no significant difference was revealed between captions and L1 subtitles on any test. In spite of these non-significant results, the researchers pointed out that the L1 group performed slightly better than the caption group on form and meaning recall. The captions group only outperformed the L1 subtitles group when the TWs were pre-taught. The researchers thus suggested that L1 subtitles might have compensated for the lack of instruction in the no pre-teaching groups, which might have benefitted learners' vocabulary learning. This potential benefit of L1 subtitles might also be attributed to participants' low proficiency level.

As can be seen from the aforementioned studies, in those studies reporting a significant difference between L1 subtitles and captions, captions seem to have an advantage over L1 subtitles for the learning of word forms (e.g., Peters, 2019; Peters et al., 2016). The effects of captions for learning word meanings seem to be similar to L1 subtitles (e.g., Frumusel, 2018; Peters et al., 2016; Pujadas & Muñoz, 2019). Especially, the benefits of captions for incidental vocabulary learning tend to be more

salient for high-intermediate to advanced level L2 learners (e.g., Peters, 2019), whereas L1 subtitles are more likely to favour low-level L2 learners (e.g., Peters et al., 2016; Pujadas & Muñoz, 2019). However, the results are far from settled due to the limited number of studies. These inconclusive findings are probably caused by differences in participant profiles (L1 background, L2 proficiency level), the number of viewing sessions (one-off viewing, longitudinal viewing), and vocabulary test design. Although inconclusive, these studies show that learners' vocabulary size (e.g., Montero Perez, 2020a; Montero Perez et al., 2018; Peters, 2019; Peters et al., 2016; Puimège & Peters, 2019), the frequency of occurrence of target items (e.g., Peters, 2019; Peters et al., 2016; Teng, 2018), and word-related factors (e.g., imagery support, cognateness, and corpus frequency; Peters, 2019) can positively predict learners' vocabulary gains in subtitled viewing studies.

Studies reviewed in this section focused on comparisons between captions and L1 subtitles, as these are the ones claimed to be those most frequently encountered by foreign language learners (Muñoz, 2017), and this comparison is highly relevant for the study presented in this thesis. It is important to mention, though, that a few other studies have compared the effects of captions or L1 subtitles with other types of less-frequently investigated subtitles, including: *reversed subtitles* (i.e., L1 soundtrack, L2 subtitles; e.g., Bisson et al., 2014; d'Ydewalle & Pavakanun, 1995, 1997; d'Ydewalle & Van de Poel, 1999; Fazilatfar et al., 2011), *keyword subtitles* (i.e., only present L2 words that are essential to the video plot and TWs in the caption line; e.g., Montero Perez et al., 2014, 2018; Teng, 2019, 2020b), and *glossed captions* (i.e., captions including access to meaning; e.g., Montero Perez et al., 2018; Teng, 2020b).

Regarding the use of reversed subtitles, the findings from available studies demonstrate their advantages for incidental vocabulary learning and comprehension

among language learners without or with limited prior knowledge of the L2 (e.g., d'Ydewalle & Pavakanun, 1995, 1997; d'Ydewalle & Van de Poel, 1999; Fazilatfar et al., 2011). However, most of the studies were conducted on beginners. Thus, the effects of reversed subtitles for higher level L2 learners are still unclear.

Concerning keyword captions, they have been found to be less effective than full captions among lower level young L2 learners. Teng (2018, 2019, 2020b) found that full captions were significantly better than keyword captions for primary Chinese EFL learners' incidental vocabulary learning, as reflected in form, meaning, and word use tests. Teng (2018) explains that compared to keyword captions, full captions may provide more linguistic and content information which enables learners to derive the meanings of unfamiliar words. However, for adult learners of intermediate to advanced level, no significant difference was found between the use of full captions and keyword captions in terms of comprehension (e.g., Montero Perez et al., 2014) or vocabulary learning (e.g., Montero Perez et al., 2014, 2018). Therefore, keyword captions seem to be no better than full captions, since unknown words in full captions will still attract learners' attention and stimulate learners' noticing by themselves (Montero Perez et al., 2018).

Glossed captioning has recently gained attention in the field as an advantageous subtitling type to facilitate incidental vocabulary learning (e.g., Montero Perez et al., 2018, Teng, 2020b). Two types of glossed captions have been studied: *glossed keyword captions*, which are defined as "keyword captions with access to meaning: each keyword is linked to its corresponding L1 context-bound translation" (Montero Perez et al., 2018, p. 8), and *glossed full captions*, which include full captions with access to the meanings of one or two difficult words in the captions. Glosses can be triggered by tapping a specific key (Montero Perez et al., 2018) or clicking a TW (Teng, 2020b). By

comparing the effects of glossed keyword captions with full, keyword, and no captions, Montero Perez et al. (2018) found that glossed keyword captions were significantly more effective than the other three caption forms for meaning recall, and they were also significantly better than full and no captions for form recognition. Moreover, by calculating whether the L1 meaning of a word was looked up and how many times, a positive relationship was found between learners' look-up behaviour of L1 glosses and their form recognition and meaning recall gains. This advantage of glossed captions was further confirmed by Teng (2020b), who found that glossed full captions were the most beneficial for incidental vocabulary learning, followed by glossed keyword captions, and unglossed captions. Thus, it is believed that captions with access to meaning can direct learners' attention to unknown words as regards both form and meaning (Montero Perez et al., 2018; Teng, 2020b). Moreover, similar to their advantages in reading studies, glosses can overcome the pitfall of incorrect guessing and facilitate learners' comprehension of unknown words, which can lead to more learning gains (Teng, 2020a). However, the creation of glossed captions requires technological knowledge (Montero Perez et al., 2018), and they are not easily accessible online, which limits their application outside the classroom.

In summary, after providing theoretical support for incidental learning from viewing, the above three sections reviewed empirical studies comparing the use of different subtitling types in incidental vocabulary learning, with a particular focus on captions and L1 subtitles. Contrary to the redundancy principle (Chandler & Sweller, 1991; Sweller, 2005b), empirical evidence has demonstrated the benefits rather than detriments of on-screen text for L2 learners' vocabulary learning and comprehension. Especially, captions show their superiority over L1 subtitles and no subtitles for learning word forms (e.g., Peters, 2019; Peters et al., 2016), and the effectiveness of

captions and L1 subtitles for facilitating meaning knowledge seems to be equivalently more advantageous than no subtitles (e.g., Frumusel, 2018; Peters et al., 2016). Captions tend to be more helpful for higher-level L2 learners (Danan, 2004; Peters, 2019), while lower-level learners seem to benefit more from L1 subtitles in terms of incidental vocabulary learning (Danan, 2004; Pujadas & Muñoz, 2019). Presented written L2 can boost learners' noticing of unknown words, and presented L1 translations provide correct meanings of unknown words. Glossed captions, by presenting both L2 and some L1 equivalents, are found to contribute to greater learning gains than traditional subtitling conditions, showing the potential of including both written L1 and L2 forms in viewing conditions. A further subtitling type, which also presents both L1 and L2 text, and that has recently attracted researchers' attention, is bilingual subtitles. They are the focus of the present thesis and are discussed in detail in the next section.

### **2.3.3. Bilingual Subtitles**

*Bilingual subtitles*, also called *dual subtitles* (García, 2017; Gesa Vidal, 2019; Li, 2016; Liao et al., 2020) or *double subtitles* (e.g., Lazareva & Loerts, 2017), have been defined as a subtitling type “where each block is made up of two lines, each in a different language” (Bartolomé & Cabrera, 2005, p. 94). Researchers have claimed that the use of bilingual subtitles is very rare and only found in certain multilingual countries or areas where two or more languages are spoken, for example in Finland (Finnish + Swedish), Belgium (Flemish + French), Jordan (Arabic + Hebrew), Israel (Hebrew + Arabic), Singapore (English + Malay/Mandarin/Tamil), Malaysia (English + Malay), and Hong Kong (English + Mandarin/Cantonese) (Bartolomé & Cabrera, 2005; Gesa Vidal, 2019; Gottlieb, 2004; Liao et al., 2020).

Despite mainland China being a monolingual region with Mandarin Chinese as the official language, the use of bilingual subtitles has become increasingly popular in the last two decades. Anecdotal evidence suggests that bilingual subtitles are a product of the Internet and arrived along with imports of foreign language audio-visual videos and films, which are believed to have appeared in China at the beginning of the 21<sup>st</sup> century. The increasing numbers of imported foreign audio-visual materials and their rapid online circulation leads to high demand for subtitling translations to support viewers' comprehension. To meet this huge demand, *fansubbing*, which are free grassroots versions of subtitling translations made by non-professional amateur translators rather than officially licensed professionals, became predominant on the Internet (D. Wang & Zhang, 2017; Zhang, 2013). Bilingual subtitles in China originated as a popular type of fansubbing that presented both source-language information and its translation (D. Wang & Zhang, 2017; Zhang, 2013). Bilingual subtitles have become a very common subtitling type in China and are widely applied online and on TV for foreign audio-visual materials (Liao et al., 2020; Y. Wang, 2019; D. Wang & Zhang, 2017; Zhang, 2013). They are also advocated by the China's dominant TV broadcaster with the aim of attracting a wider audience (Liao et al., 2020). The most common type of bilingual subtitles in mainland China is a combination of Mandarin Chinese (L1) and English (L2), with L1 on the first line and L2 underneath, appearing simultaneously at the bottom of the screen.

#### **2.3.3.1. Controversy around Using Bilingual Subtitles**

The potential of bilingual subtitles to facilitate L2 learning has been noted recently in different regions (e.g., Dizon & Thanyawatpokin, 2021; García, 2017; Lazareva & Loerts, 2017; Li, 2016; Lunin, 2015). However, the use of bilingual subtitles can be

controversial in two different ways. From a cognitive theoretical perspective, the use of bilingual subtitles can be supported by the bilingual version of Dual Coding Theory (as reviewed in section 2.3.1.1; Paivio, 1986). By providing images, L1, and L2 in both aural and written forms, bilingual subtitles can enhance the interconnection between nonverbal and two verbal systems, enabling more access routes to be established for information retrieval. Information is thus more likely to be activated through three systems, leading to better memory recall. Bilingual subtitles have the potential to integrate the advantages of two monolingual subtitles by providing an accurate L1 translation and enabling easier matching of L1 and L2 words (Lunin, 2015). According to Li (2012), “they [Chinese EFL learners] rely on L1 for understanding and L2 for a deeper impression of words, such as spelling” (p. 35). They have been found useful as a reference, to confirm L2 learners’ listening, assist in adapting to different English accents, and also checking spellings and translation, as well as the authentic usage of words (García, 2017; Lazareva & Loerts, 2017; Li, 2012; Lwo & Lin, 2012). Bilingual subtitles also seem to be preferred by both EFL learners and English teachers who believe that they may reduce learners’ anxiety and boost their confidence in learning English (Li, 2012). However, the use of bilingual subtitles may also potentially impede learning. Cognitive Load Theory (Sweller, 1988) and the redundancy principle (Sweller, 2005b) propose that identical information presented in multiple forms might result in cognitive overload, which could then be detrimental for learning. The use of bilingual subtitles can be considered redundant, since they present the same verbal information in both aural and written forms, together with written L1 translations. Especially during fast-paced viewing, learners need to process dynamic video images along with two lines of subtitles, which may increase their cognitive burden and hamper their learning and

information processing. Bilingual subtitles have been found to increase the cognitive load for some L2 learners (e.g., Lwo & Lin, 2012).

The second controversy around the use of bilingual subtitles lies in the arguments about the use of L1 translation in L2 vocabulary learning. Using L1 equivalents to learn L2 vocabulary has long been criticised as encouraging learners' laziness and discouraging deeply engaging with L2 words, linking L2 words too exclusively to L1 equivalents, and failing to distinguish word meanings in different contexts (Joyce, 2018; Prince, 1996). According to the Depth of Processing Theory (summaries in section 2.5.2.1; Craik & Lockhart, 1972), it could also be argued that having translations of L2 unknown words might reduce learners' cognitive analysis of their meanings and lead to shallower memory traces which are then reflected in smaller gains. The hindrance of using L1 translation can even be exacerbated due to the lack of corresponding features in the L1, or mismatches between L2 and L1 semantic concepts, which can lead to learning difficulties and inaccurate establishment of meaning (Schmitt, 2010). In contrast, increasing evidence has also been put forward to advocate the use of L1 in L2 vocabulary learning. Psycholinguistic studies have demonstrated that for L2 learners with different proficiencies, the L1 is active while processing L2 words that share the same concept, and learners are likely to benefit from this lexical transfer and establish a form-meaning link (e.g., Jiang, 2002; Sunderman & Kroll, 2006). Vocabulary researchers have also reported greater learning gains using L1 translations compared to L2 definitions (e.g., Joyce, 2018; Ramachandran & Rahim, 2004), or inferring word meanings from L2 contexts (e.g., Prince, 1996). In spite of the increasing popularity of bilingual subtitles and the controversy around their potential benefits, very few studies have investigated their effectiveness for vocabulary learning.

### **2.3.3.2. Empirical Studies on the Use of Bilingual Subtitles for Incidental Vocabulary Learning**

In the context of engineering education, García (2017) examined users' opinions about the effectiveness of bilingual subtitles for incidental L2 vocabulary learning and comprehension. Participants' responses ( $N = 62$ ) to an online questionnaire showed that learners who used bilingual subtitles believed that they were helpful for the development of their L2 vocabulary in relation to form, meaning, and use. However, results were only based on participants' self-assessment, which may not reflect the actual effectiveness of bilingual subtitles.

Lazareva and Loerts (2017) compared the effectiveness of bilingual subtitles, L1, and no subtitles for incidental vocabulary learning. Forty-three Dutch speakers with no prior knowledge of Russian, from a Dutch university, were divided into three groups to watch an 8-minute Russian cartoon three times. After viewing, an audio sentence recognition test and a word meaning recognition test were administered to the two subtitled groups. No group differences were revealed in terms of audio sentence recognition. However, the bilingual subtitles group significantly outperformed the L1 subtitles group on word meaning recognition, suggesting an advantage of bilingual subtitles for facilitating meaning knowledge. However, care should be taken when interpreting the findings. This superiority of bilingual subtitles could be related to the test format. Since no audio access was available during the word recognition test, the test might have favoured the bilingual subtitles group which had seen the L2 written form during repeated viewings, leading to higher test scores. Moreover, the target items included in the meaning recognition test were a mixture of single words, multiword units, and short sentences. Thus, it is unclear whether the advantages of bilingual subtitles were consistent when learning different target items. Additionally, the

participants in this study did not have prior knowledge of the L2, making the findings less generalizable to other L2 proficiencies.

A very recent study conducted by Dizon and Thanyawatpokin (2021) compared the effects of captions, L1, and bilingual subtitles on Japanese EFL learners' incidental vocabulary learning and comprehension. Ninety-six beginners were asked to watch one episode of an American sitcom in one of three subtitling conditions. Their knowledge of 20 single TWs was measured four weeks before and immediately after viewing. The findings showed that for form recognition, no significant differences were revealed between the bilingual subtitles and monolingual groups, but the L1 subtitles group surprisingly outperformed the captions group. The authors attributed the better performance of the L1 subtitles group to the low proficiency of the participants, who might have encountered difficulty in following the captions. In terms of meaning recall, both the bilingual and L1 subtitles groups significantly outperformed the captions group. A similar pattern was observed for comprehension, with both the bilingual and L1 subtitles groups outperforming the captions group. However, in this case, bilingual subtitles also outperformed L1 subtitles. Studies by Lazareva and Loerts (2017) and Dizon and Thanyawatpokin (2021) seem to suggest an advantage for bilingual subtitles for facilitating meaning knowledge, but not for learning word forms. However, none of these studies have fully revealed the relative effects of bilingual subtitles, as a captions group was not included in Lazareva and Loerts's (2017) study and Dizon and Thanyawatpokin (2021) did not include a no subtitles group. For a proper understanding of the potential advantages of bilingual subtitles, it is important to compare bilingual subtitles with captions, L1, and no subtitles conditions. Importantly, their findings are restricted to beginner learners.

Particularly relevant for the present study are those studies that have examined the use of bilingual subtitles with more proficient Chinese EFL learners. Despite the support for using bilingual subtitles from both language learners and teachers in China (Li, 2012; D. Wang, 2012), empirical investigations examining the effects of bilingual subtitles on the incidental vocabulary learning of Chinese learners are scarce. Only a handful of studies have been conducted in this context, and most of the evidence is limited to academic theses and dissertations. Additionally, since the exploration of bilingual subtitles is still at a nascent stage, there are some important limitations in published studies, which questions the validity of their findings. These drawbacks are summarised at the end of this section.

The advantages of bilingual subtitles for incidental vocabulary learning were documented by Li (2016) in her doctoral thesis. A within-group design mixed methods study was conducted on 120 intermediate level Chinese EFL learners from three classes at a university. Each class was asked to watch three 20-minute BBC documentary clips in three weeks with one of the three subtitling conditions randomised in a counterbalanced order: captions, L1, and bilingual subtitles. A control group was also included with no subtitles. Thirty unfamiliar TWs that varied in word class (nouns, adjectives, verbs), frequency of occurrence (1 to 4), and level of frequency in the corpus (low or high) were chosen. Word recognition (i.e., choosing one sentence in which the TW was used correctly from three sentences) and meaning recall (i.e., providing the meaning/translation of the TW) tests were administered immediately after each viewing session and three weeks later. Participants' opinions on the use of different subtitles for their vocabulary learning and comprehension were also collected in a final questionnaire. Results showed that bilingual subtitles outperformed the other three groups for word recognition and meaning recall in both immediate and delayed posttests.

In addition, the L1 subtitles group significantly outperformed the captions group in immediate meaning recall, and the captions group outperformed the L1 subtitle group in delayed word recognition. TWs with higher frequencies of occurrence and corpus frequencies tended to be better learned, especially in the bilingual subtitles condition. Abstract nouns and verbs were less likely to be learned compared to concrete nouns and adjectives. The questionnaire results showed that the majority of participants chose bilingual subtitles as the best subtitling type for comprehension and vocabulary learning.

Two recent studies seem to suggest that the advantages of bilingual subtitles are inconsistent across different L2 proficiencies. In the study by Hao et al. (2021), intermediate level (i.e., non-English major) undergraduates ( $n = 147$ ) and advanced level (i.e., English major) undergraduates ( $n = 125$ ) were randomly assigned to one of four subtitling groups (i.e., captions, L1, bilingual, no subtitles) to watch four 5-minute TED talk videos in two weeks. Participants' knowledge of 36 academic words in the selected material was measured by a meaning matching test before and immediately after each viewing session. Comprehension was also checked using multiple-choice questions. The results revealed no significant group differences among the intermediate level learners in any tests. However, for advanced level learners, both bilingual and no subtitles groups significantly outperformed the captions group on meaning recognition. The L1 subtitles group outscored the no subtitles group on a comprehension test. The authors attributed the lack of group differences in the intermediate level learners to the challenging features of the selected material, since the L2 input might not have been comprehensible enough for them, making learning difficult even with the help of on-screen text. In addition, it should be noted that the presentation of L2 lines above L1 lines in the subtitles in this study does not reflect the most common presentation of

bilingual subtitles and might lead to learners' different use of subtitles and thus explain the different findings.

By conducting a within-subject study, Y. Wang (2019) also explored the relative effects of bilingual subtitles on incidental vocabulary learning and comprehension among intermediate to advanced Chinese EFL learners. Eighty university students from four classes (i.e., two freshman classes, one junior, one first year graduate) were asked to watch four excerpts of an American sitcom series in four subtitling conditions (i.e., captions, L1, bilingual, no subtitles) in a counterbalanced design. Participants' knowledge of five TWs in each excerpt was measured via meaning recall after each viewing session. Open-ended questions were used to measure participants' comprehension. There were mixed findings across different classes. In terms of vocabulary learning, L1 subtitles and captions seemed to be more beneficial for freshman students, while bilingual and L1 subtitles worked better for first year graduate students. However, no subtitles and captions were more effective for juniors. The authors explained the mixed findings as being a result of learners' different prior knowledge of the TWs in different video clips. Since the comparisons were analysed based on absolute gains, participants who knew more TWs in the pretest would have achieved lower gains. Moreover, it should be noted that only five TWs were tested in each video clip, which might not fully capture learners' learning gains. In terms of comprehension, the use of bilingual and L1 subtitles was in general more beneficial than captions and no subtitles, with only the junior class showing no group differences.

The effectiveness of bilingual subtitles for language learning by younger learners of English was explored in a study by Lwo and Lin (2012). Thirty-two junior high school students were invited to watch two animations in one of four conditions (i.e., captions, L1, bilingual, no subtitles), and their comprehension of the video was checked

while viewing (answering questions orally after each scene). After each viewing session, participants took a comprehension test and two vocabulary tests (a multiple-choice meaning recognition test and a fill-in-the-blank test). Semi-structured interviews were held after the viewing sessions to explore participants' attention allocation while viewing and their attitudes towards different subtitling types for comprehension. Results showed no significant advantages of bilingual subtitles over other subtitling types for vocabulary learning and comprehension. Although no significant differences were found between groups, lower-level learners seemed to benefit more from the use of bilingual subtitles, especially for recalling more complex sentences. Together with the interview data, it showed that lower-level learners reported more selective use of information while viewing to meet their needs for comprehension or language learning, while more proficient learners were found to be more easily distracted by L1 lines in bilingual subtitles. However, as the authors pointed out, the findings should be interpreted with caution, as the viewing material used was specifically designed for English language teaching and contained simple sentences and a high degree of similarity. Crucially, the procedure in this study required participants to stop and answer comprehension questions after each scene and sentence, which interrupted the viewing process and could have affected the findings.

Taken together, very few studies have explored the relative effects of bilingual subtitles on incidental vocabulary learning and the available studies have reported inconclusive findings. While some studies have found advantages for bilingual subtitles over L1 subtitles (e.g., Lazareva & Loerts, 2017; Li, 2016), captions (e.g., Dizon & Thanyawatpokin, 2021; Hao et al., 2021; Li, 2016; Y. Wang, 2019), and no subtitles (e.g., Lazareva & Loerts, 2017; Li, 2016; Y. Wang, 2019) for facilitating meaning knowledge, no significant group differences have been reported in other studies (e.g.,

Hao et al., 2021; Lwo & Lin, 2012; Y. Wang, 2019). Moreover, the effects of bilingual subtitles also varied within groups of similar proficiency, as some studies documented an advantage of bilingual subtitles over monolingual subtitles for lower-level learners (e.g., Dizon & Thanyawatpokin, 2021; Lazareva & Loerts, 2017), while others did not (e.g., Lwo & Lin, 2012). Their effectiveness for intermediate and advanced level L2 learners was also inconsistent across different studies (e.g., Hao et al., 2021; Li, 2016; Y. Wang, 2019). Nevertheless, more consistent findings have been reported for the benefits of bilingual subtitles for comprehension, showing that bilingual subtitles are as effective as L1 subtitles for facilitating learners' viewing comprehension and more helpful than captions and no subtitles (e.g., Dizon & Thanyawatpokin, 2021; Y. Wang, 2019).

The reason for these inconsistent findings for vocabulary learning may relate to important methodological limitations. It should be emphasised that most of the bilingual subtitle studies reviewed in this section were not published in top international journals. They are reviewed here to capture earlier attempts to explore bilingual subtitles and provide a comprehensive review covering the most relevant and recent findings in this area. Therefore, their findings need to be treated with caution. The majority of studies did not carefully check the suitability of viewing materials, which could have caused comprehension difficulties and affected learning outcomes (e.g., Dizon & Thanyawatpokin, 2021; Hao et al., 2021). Although with the help of images the required lexical coverage seems to be lower for viewing than for reading (Peters & Muñoz, 2020), it is still important to ensure the appropriateness of the viewing material selected for adequate comprehension (Webb & Rodgers, 2009). Additionally, most studies did not include information about the TWs, nor did they take into account different word-related factors (e.g., part of speech, word length) and word frequencies of occurrence in the analysis (e.g., Dizon & Thanyawatpokin, 2021; Hao et al., 2021;

Lazareva & Loerts, 2017; Lwo & Lin, 2012; Y. Wang, 2019). Moreover, despite the important role of learners' vocabulary knowledge in learning from viewing (e.g., Montero Perez, 2020a; Peters et al., 2016), no studies on bilingual subtitles have taken into account the potential effects of learners' vocabulary knowledge on comprehension or vocabulary learning. Additionally, most of these studies only used one vocabulary test to examine learners' learning gains (e.g., Hao et al., 2021; Lazareva & Loerts, 2017; Y. Wang, 2019), which is insufficient to capture the effects of bilingual subtitles on different aspects of vocabulary learning. Last but not least, the lack of standard reporting of test scores and learning gains also causes difficulties in making cross-study comparisons.

Seemingly contradictory findings may also be explained by differences in research design (incidental learning, intentional learning), participant profiles (different L2 proficiency levels, different viewing habits), viewing materials (animations, self-designed animations, documentary, TV series, TED talks), and the number of viewing sessions (one-off viewing, multiple viewing sessions). These differences can potentially trigger learners' different uses of bilingual subtitles which might account for different learning gains.

Importantly, inconclusive findings may also be explained by differences in the way learners make use of the sources of information available in bilingual subtitles. Some previous studies have attempted to use questionnaires and interviews to explore L2 learners' use of bilingual subtitles while viewing. Some participants have reported that bilingual subtitles gave them the freedom to switch their attention across different lines flexibly (García, 2017; Li, 2016), whereas others thought there was too much information presented at the same time, which could potentially cause cognitive overload (Lwo & Lin, 2012). These initial findings point towards interesting differences

in how learners make use of the input provided in bilingual subtitled viewing, which could indeed explain the contradictory findings that have been reported so far. However, none of the available studies systematically reviewed how learners actually process different sources of information when using bilingual subtitling in L2 learning settings. Therefore, it is still not clear how attention is split between the image and subtitling area in bilingual subtitles, and how learners' process unknown words during bilingual subtitled viewing. A better understanding of learners' processing of different sources of input in bilingual subtitles, as well as TWs and their translations in subtitles, would help us better understand the potential benefits of bilingual subtitles for incidental vocabulary learning, and help to clarify the conflicting findings reported so far.

To sum up, although informative, the available research findings are far from conclusive. The superiority of bilingual subtitles over other subtitling conditions has been revealed in some studies (e.g., Dizon & Thanyawatpokin, 2021; García, 2017; Hao et al., 2021; Lazareva & Loerts, 2017; Li, 2016; Y. Wang, 2019), but not in others (e.g., Dizon & Thanyawatpokin, 2021; Hao et al., 2021; Lwo & Lin, 2012). More evidence is needed to reach a better understanding of the benefits (and potential detrimental effects) of bilingual subtitles for comprehension and vocabulary learning. How learners make use of bilingual subtitles and what may cause inconclusive results remain unclear. A couple of studies exploring learners' use of bilingual subtitles relied on self-reporting questionnaires and interviews (e.g., Li, 2016; Lwo & Lin, 2012). While informative, they were only based on learners' subjective opinions, which might be limited by memory decay. Other techniques that provide more direct evidence for the way learners' process bilingual subtitles should be used.

One such technique that has the potential to inform research in this area is eye-tracking. In recognition of its benefits, recent viewing studies have started to use it.

Eye-tracking is chosen as one of the data collection techniques in the present study. The following sections review eye-tracking techniques and relevant eye-tracking studies that have been conducted on L2 reading and viewing.

#### **2.4. Use of Eye-tracking in SLA**

Eye-tracking allows the real-time, online, and direct recording of an individual's eye-movement behaviour, usually when he or she processes information on a computer screen (Conklin, Pellicer-Sánchez, & Carroll, 2018; Godfroid, 2020a). It is believed to be a versatile, unobtrusive, and sensitive methodology which has been widely applied in a range of applied linguistics domains (Conklin et al., 2018; Godfroid, 2020a). It allows researchers to know which areas of visual input have been focused on within a specific time period, how many times, and for how long (Pellicer-Sánchez & Siyanova-Chanturia, 2018). The premise of eye-tracking is the *eye-mind assumption* (Just & Carpenter, 1980) and the *eye-mind link* (Reichle, 2006), which assumes a close association between eye movements and the human mind. In other words, people's eye gaze reflects their covert attentional process.

According to Robinson (2003), "attention is the process that encodes language input, keeps it active in working and short-term memory, and retrieves it from long-term memory" (p. 361). When processing information, attention can be conceptualized as various cognitive functions to regulate our actions (e.g., selecting information, focusing on information, inhibiting distractions, etc.) and further facilitate our learning (Robinson, Mackey, Gass, & Schmidt, 2011). Since eye movements are direct measures of allocation of overt attention and are closely related to covert attentional processes (Leow, 2015; Rayner, 2009), eye movements are believed to be a robust physiological measure to study attention (Leow, 2015; Robinson et al., 2011). Despite the relationship

between attention and eye movements not always being straightforward, the link between them is close in complex information processing tasks such as reading (Rayner, 1998). In reading, it is assumed that what is fixated on is what people are attending to, and a longer fixation time relates to more cognitive processing (Conklin et al., 2018). Eye-movement data have been used for the operationalisation of attention in studies on L2 learning (Conklin et al., 2018; Godfroid, 2020a; Rayner, 2009).

#### **2.4.1. Eye-Tracking Measures**

Eye-tracking data are reported according to different *areas of interest* (AOI, also known as *interest area*, IA, or *region of interest*, ROI). Eye-tracking usually records the following three main types of eye movements in relation to a selected AOI. The first type of eye movement is *saccades*, i.e., rapid movements of the eyes from one point to another (Rayner, 2009). In text-based reading, saccades normally move from left to right (e.g., English, Chinese), but they can also be from right to left according to the language (e.g., Arabic, Hebrew). A second type of eye movement is *regressions*, i.e., saccades that move back to previously read text, which occur about 10–15% of the reading time in skilled readers (Rayner, 2009). Thirdly, *fixations* are periods of time when the eyes remain relatively still between saccades to process visual information (Rayner, 2009). Due to the rapid nature of saccades, when new information can hardly be acquired, fixations are the most useful measure for text-based researchers (Conklin et al., 2018; Rayner, 2009).

In the examination of reading, fixation durations are often classified as early or late measures, reflecting different stages of the reading process (Conklin et al., 2018). *Early measures* capture the initial stages of text processing, reflecting highly automatic word recognition and lexical access processes (Clifton, Staub, & Rayner, 2007; Conklin et al.,

2018), while *late measures* signal comparatively late stages of processing, which reflect more conscious and strategic processes and could indicate an interrupted reading process (Godfroid, 2020a; Conklin et al., 2018). Both early and late measures are typically reported in order to paint a full picture of the process being examined. In text-based studies, the most commonly reported eye-tracking measures in SLA are presented in Figure 8 (for a comprehensive review of eye-tracking measures, see Conklin & Pellicer-Sánchez, 2016; Conklin et al., 2018; Godfroid, 2020a).

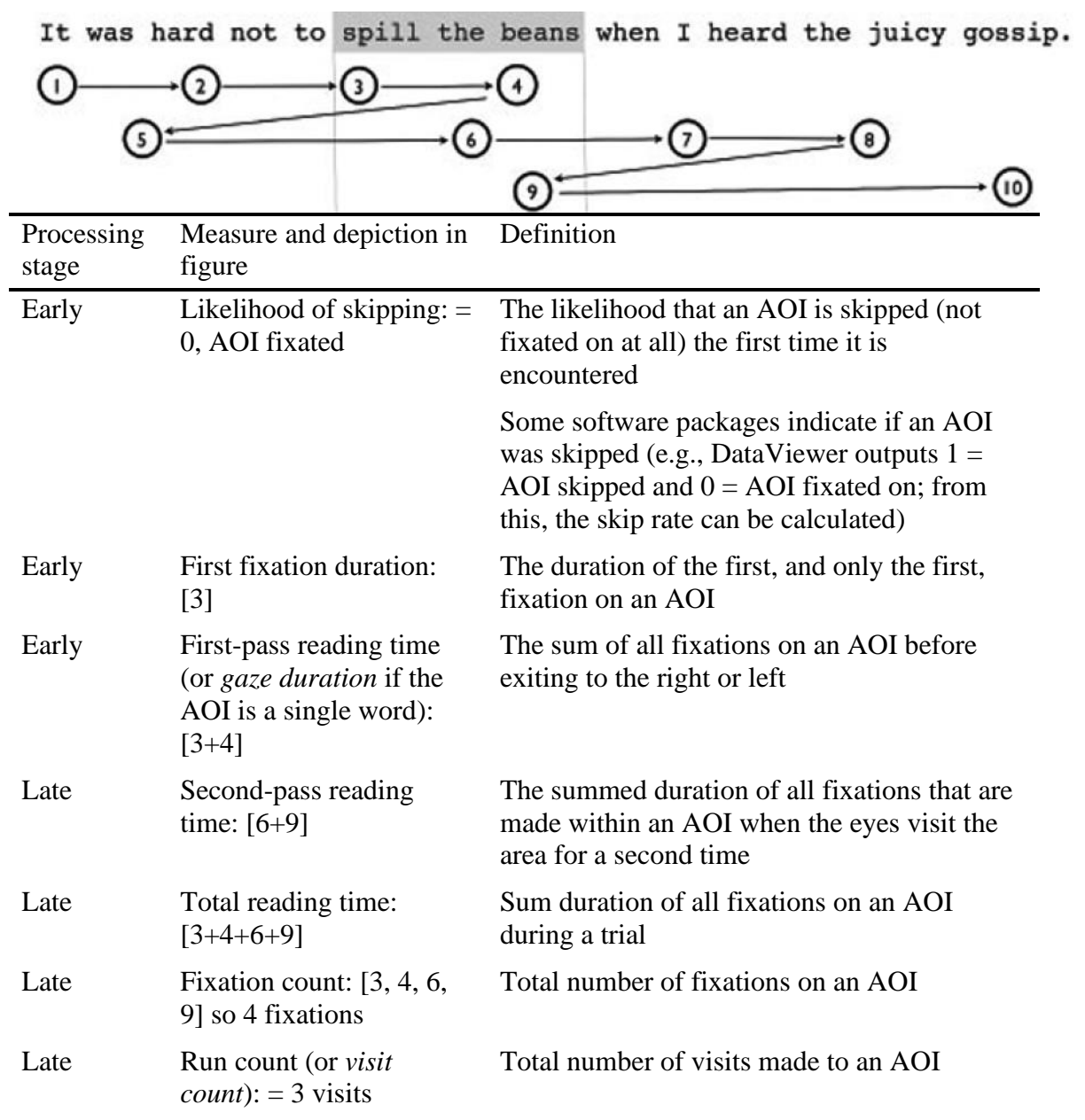


Figure 8. Illustration and Definition of Commonly Reported Eye-Tracking Measures in Text-Based Studies (adapted from Conklin and Pellicer-Sánchez (2016, p. 456), Conklin, Pellicer-Sánchez, and Carroll (2018, p. 68), and Godfroid (2020, p. 211))

Most text-based studies use eye movements as a processing load measure, and hypothesize that a longer reading time implies more difficult and complex underlying processing (Tanenhaus, 2007). Both first fixation duration and first-pass reading time are early measures, reflecting word-level processing. In other words, they represent how easily a reader can retrieve the meaning of a word from their mental lexicon, and are usually affected by word frequency and familiarity, the ambiguity of word meanings, word predictability and semantic associations (Conklin et al., 2018). Second-pass reading time is a pure late-processing measure that reflects reanalysis having encountered an initial processing difficulty. It is recommended to be used to clearly distinguish late from early processing (Godfroid, 2020a). Apart from being influenced by lexical factors, late measures may also relate to higher-level factors such as contextual, sentence or discourse factors (Conklin et al., 2018). Both total reading time and fixation count are aggregate late eye-movement measures. They subsume all time and visits and indicate the general processing of an AOI (Conklin et al., 2018; Godfroid, 2020). Although the measures are not independent but do to some extent correlate with each other, it is generally recommended to use multiple measures to paint a more complete picture of different cognitive processes (Conklin et al., 2018; Godfroid, 2020a; Rayner, 1998).

The processing of a visual scene is different from text-based reading. In image- or video-based studies, eye movements are more often used as a representational measure, which relates to “when and where people fixate as the utterance unfolds” (Tanenhaus, 2007, p. 318). More attention being paid to an area implies more saliency and attraction, or more consideration involved in processing the area (Conklin et al., 2018). In image-

or video-based studies, apart from late measures such as total reading time and fixation count, the proportion of fixations is believed to be especially useful, since this proportion can reveal the relative time spent on an area compared to other parts of the stimulus (Conklin et al., 2018). However, there is no standard as to which eye-tracking measures should be used in image- or video- based studies, as can be observed from the variety of measures in video-based studies summarised in Table 4. The measures reported in existing studies include fixation measures, proportion of fixations, skip rate, and run count. Their definitions are presented in Table 5. It has been emphasized that the selection of measures should always be based on the aim and task of each research (Conklin et al., 2018; Godfroid, 2020a). For example, if we want to know whether a subtitling area has been attended to, the use of skip rate may be sufficient. But if we want to further explore how long a subtitling area has been attended to compared to the image area, total reading time percentage can be more informative.

Table 4. Summary of Eye-Tracking Measures Reported in Video-Based Studies

|                                                               | Total<br>readin<br>g<br>time<br>% | Average/Me<br>an fixation<br>duration | Fixatio<br>n<br>count | Ski<br>p<br>rate | Total<br>fixatio<br>n<br>duratio<br>n | Run<br>cou<br>nt | First-pa<br>ss<br>reading<br>time | Rereadi<br>ng time |
|---------------------------------------------------------------|-----------------------------------|---------------------------------------|-----------------------|------------------|---------------------------------------|------------------|-----------------------------------|--------------------|
| Perego,<br>Del<br>Missier,<br>Porta, and<br>Mosconi<br>(2010) | ✓                                 | ✓                                     | ✓                     |                  |                                       | ✓                |                                   |                    |
| Kruger,<br>Hefer,<br>and<br>Matthew<br>(2014)                 | ✓                                 |                                       |                       |                  |                                       |                  |                                   |                    |
| Liao et al.<br>(2020)                                         | ✓                                 | ✓                                     |                       |                  |                                       |                  |                                   |                    |

|                                   | Total<br>readin<br>g<br>time<br>% | Average/Me<br>an fixation<br>duration | Fixatio<br>n<br>count | Ski<br>p<br>rate | Total<br>fixatio<br>n<br>duratio<br>n | Run<br>cou<br>nt | First-pa<br>ss<br>reading<br>time | Rereadi<br>ng time |
|-----------------------------------|-----------------------------------|---------------------------------------|-----------------------|------------------|---------------------------------------|------------------|-----------------------------------|--------------------|
| Gass et al. (2019)                | ✓                                 |                                       |                       |                  |                                       |                  | ✓                                 | ✓                  |
| Muñoz (2017)                      |                                   | ✓                                     | ✓                     | ✓                | ✓                                     |                  |                                   |                    |
| Winke et al. (2013)               | ✓                                 | ✓                                     |                       |                  |                                       |                  |                                   |                    |
| Bisson et al. (2014)              |                                   | ✓                                     | ✓                     | ✓                | ✓                                     |                  |                                   |                    |
| d'Ydewalle and De Bruycker (2007) | ✓                                 |                                       |                       | ✓                | ✓                                     | ✓                |                                   |                    |
| Sum                               | 6                                 | 5                                     | 3                     | 3                | 3                                     | 2                | 1                                 | 1                  |

Table 5. Definitions of Commonly Reported Eye-Tracking Measures in Video-Based Studies (adapted from Conklin, Pellicer-Sánchez, and Carroll (2018, p. 68) and Godfroid (2020, p. 211))

| Eye-tracking measure           | Definition                                                                                 |
|--------------------------------|--------------------------------------------------------------------------------------------|
| Total reading time %           | Proportion of all summed time spent on an AOI                                              |
| Average/Mean fixation duration | Average duration time of fixations on an AOI                                               |
| Fixation count                 | Total number of fixations on an AOI                                                        |
| Skip rate                      | Likelihood that an AOI is skipped (not fixated on at all) the first time it is encountered |
| Total fixation duration        | Duration of all summed fixations on an AOI                                                 |
| Run count                      | Total number of visits made to an AOI                                                      |
| First-pass reading time        | Sum of all fixations on an AOI before exiting to the area                                  |
| Rereading time                 | Summed duration of all fixations on an AOI except for those made during a first pass       |

In video-based studies, apart from AOIs which identify specific areas that researchers are interested in looking at, it is also vital to identify specific time points in the video to be investigated (Conklin et al., 2016; Godfroid, 2020a). Specific time points are called *interest periods* (IPs). After identifying both AOIs and IPs, learners' eye movements are only recorded within AOIs during active IPs. For example, in a captioned viewing study conducted by Winke et al. (2013), participants' eye movements were only recorded during those time periods when captions were present on the screen. In this case, IPs were defined as the presentation times of captions in the video.

#### **2.4.2. *Eye Movements in Visual Scene Perception and Reading***

Since the focus of the present study is on subtitled video, which is a special type of stimulus consisting of dynamic images, soundtrack, and text, this section briefly summarises the main findings on eye-movements concerning visual scene perception and text-based research. It then moves on to the application of eye-tracking in SLA research.

Rayner (2009) claims that compared to reading, fixation durations in visual scene perception tend to be longer (with an average close to 300 ms) and the size of saccades tends to be larger (with an average 40–50 ms). Compared to watching static pictures, watching videos has been found to have even longer fixation durations and smaller saccades (Dorr, Martinetz, Gegenfurtner, & Barth, 2010). Image-based studies have revealed that viewers do not fixate on every part of a scene and most of their fixations fall on the more informative or salient areas of the scene (Rayner, 2009). Moreover, a viewer can extract enough information to understand the gist of an image with a present time as short as 42 ms (Castelhano & Henderson, 2008). In video-based studies, empirical evidence has also demonstrated a systematic relationship between what a

viewer hears and where (and when) their eyes tend to move (Coklin et al., 2018; Godfroid, 2020a; Rayner, 2009).

Psycholinguists have used eye-tracking extensively to investigate text processing while reading (for a review, see Rayner 1978, 1998, 2009). Decades of eye-tracking research on reading have shown that many factors affect reading behaviour. Words with high frequency or familiarity to the reader tend to receive fewer fixations (e.g., Williams & Morris, 2004). In addition, longer words and those with lower predictability are more likely to receive longer fixations (Kliegl, Grabner, Rolfs, & Engbert, 2004).

Concerning different writing systems, for L1 readers of English, the average fixation duration on a single word lasts for around 200–250 ms and the mean saccade length is about 7–9 letter spaces (Rayner, 1998). For L1 readers of Chinese, since linguistic information is more densely packed in Chinese characters, average saccades are much shorter in Mandarin Chinese (about 2.6 characters) than in English (about 7–8 letters) (Rayner, Li, Juhasz, & Yan, 2005; Rayner, Li, & Pollatsek, 2007). However, it seems that the eye movements of L1 Chinese readers share more similarities than differences with L1 English readers (Rayner et al., 2005). Average fixation durations, reading rates, and regression rates when reading L1 Chinese do not seem to differ much from L1 English (Rayner, 2009; Rayner et al., 2007; Sun & Feng, 1999). Moreover, similar to reading in English, Chinese words with high frequency and predictability tend to be fixated on less and skipped more often (Rayner et al., 2005; 2007). These findings suggest that eye movements during reading are controlled by the linguistic content rather than by the visual form of the particular language. Apart from the aforementioned factors, readers' individual differences, such as executive control, age, language learning background, and reading speed, also play important roles in reading (Godfroid, 2020a; Kliegl et al., 2004; Rayner, 1998).

While extensively used in psychology and psycholinguistics for decades to explore language processing, eye-tracking has only recently started to be used in SLA to examine the process of language learning (Conklin & Pellicer-Sánchez, 2016). To date, research using eye-tracking in SLA has involved two main strands: spoken-based (i.e., studying auditory processing) and text-based studies (i.e., studying textual processing), with the latter one being dominant (Godfroid, 2020a). Most spoken-based studies are more psycholinguistically oriented, where eye-tracking is taken as a representational measure, revealing the activations of certain linguistic representations in the listener's mind (Tanenhaus, 2007; Tanenhaus & Trueswell, 2006). For text-based studies, eye-tracking is taken as a processing load measure (Tanenhaus & Trueswell, 2006), which is based on the eye-mind link assumption, believing that longer eye fixation durations may signal more processing or more complex task demands (Godfroid, 2020a).

Text-based eye-tracking studies have been classified into five main strands by Godfroid (2020a, p. 86): grammar, vocabulary and the bilingual lexicon, instructed SLA, captions and subtitles processing, and assessment. The present research is situated within the captions and subtitles processing and vocabulary strands and uses eye-tracking to investigate: 1) learners' attention allocation to captions/subtitles and images during subtitled viewing, and 2) learners' processing of unknown words during subtitled viewing and its potential relationship to their vocabulary gains. Therefore, eye-tracking studies on captions/subtitles processing and text-based studies focusing on L2 vocabulary acquisition have the closest relevance to the present study and are reviewed in the following sections.

#### ***2.4.3. Empirical Eye-Tracking Studies on Captions/Subtitles Viewing***

Eye-tracking is a very convenient tool when exploring learners' allocation of attention while watching captioned/subtitled videos. Learners might not be aware of their own attentional process trajectory, however, their unconscious behaviour can be captured by eye-tracking equipment (Conklin et al., 2018). In SLA, most of the studies using eye-tracking in on-screen text viewing have investigated L2 learners' processing of different on-screen texts, with some of them also exploring the relationship between the processing of on-screen text and comprehension or vocabulary learning. As reviewed in section 2.4.1, there is considerable variation in the eye-tracking measures reported in these studies. While the studies reviewed in this section used a variety of measures, only the results of those most relevant to the focus of the present study (i.e., those pertaining to the attention distribution between subtitling and image areas) will be reviewed. These measures include: total reading time percentage, fixation count, skip rate, and run count.

d'Ydewalle and colleagues (2007, 1991, 1992) were pioneers in eye-tracking research on on-screen text reading. They found that for both adults and children, the reading of on-screen text seemed to be more or less spontaneous, and viewers could switch effortlessly between images and the subtitling area. L1 viewers would read on-screen text regardless of whether they had the habit of using subtitles (e.g., d'Ydewalle et al., 1991), or whether they had limited knowledge of the L2 (e.g., d'Ydewalle & De Bruycker, 2007). d'Ydewalle et al. (1991) found that when the soundtrack and subtitles were both in the viewer's L1, participants spent about 25% of the time looking at the subtitles despite their habit of using subtitles, and more time was spent on two-line than one-line subtitles. d'Ydewalle and De Bruycker (2007) also compared the eye movements of Dutch-speaking adults and children (without prior knowledge of Swedish) while viewing a 15-minute excerpt from a Swedish cartoon

with one/two-line L1 subtitles or reversed subtitles (i.e., Dutch soundtrack and Swedish subtitles). Eye-tracking data showed that, in general, less time was spent on one-line subtitles than two-line ones. Moreover, participants using L1 subtitles skipped subtitles less often ( $M = 4\%$  vs.  $21\%$  in reversed subtitles), spent more total reading time on subtitles ( $M = 41\%$  vs.  $26\%$  in reversed subtitles), and had higher run counts between images and subtitling areas ( $M = 0.49$  vs.  $0.24$  in reversed subtitles) than when using reversed subtitles. Therefore, it seems that L1 subtitles are read more often when the soundtrack is in an unfamiliar language regardless of the viewer's age.

Perego et al. (2010) explored Italian-speakers' viewing of a well- or ill-segmented L1 subtitled 15-minute Hungarian drama. No participants had any prior knowledge of Hungarian or the habit of using subtitles. A multiple-choice word recognition test and comprehension tests were administered after the viewing. The offline tests revealed that participants could benefit from subtitled viewing in both well- or ill-segmented conditions. Eye-tracking data revealed that, in general, participants spent on average  $67\%$  of fixation time and  $172.81$  fixation counts on the L1 subtitling area. No relationships were found between participants' eye movements and their offline test scores. However, it should be noted that participants in the studies by d'Ydewalle & De Bruycker (2007) and Perego et al. (2010) were not familiar with the L2 in the soundtrack, which might have resulted in heavier reliance on L1 subtitles while viewing.

Eye-tracking has also been used to explore L2 learners' caption-reading behaviour. Winke et al. (2013) triangulated eye-tracking data with interview data in order to explore the caption-reading behaviour of English-speaking learners of different L2s, and the differential effects of captions across languages. L1 English learners of Arabic, Chinese, Russian, and Spanish watched two documentary clips (dubbed and captioned

in the L2s) that differed in content familiarity while their eye movements were recorded. The findings showed that, in general, participants spent approximately 68% of fixation time on captions during their presentation. However, this percentage varied across learners of different L2s, with learners of Arabic spending a significantly higher percentage of time reading captions (75%) than learners of Spanish (63%) or Russian (67%). Content familiarity only mattered to learners of Chinese, who spent significantly more time reading captions with unfamiliar (74%) than familiar content (62%). Based on interview findings, the researchers pointed out that participants relied on captions for comprehension, especially when the soundtrack was difficult to understand. A longer processing time spent on captions might indicate learners' efforts to derive meaning from the captions, which might be affected by the logographic distance between learners' L1 and L2 and learners' different L2 proficiency levels.

Gass et al. (2019) investigated L2 learners' captioned reading behaviour and its relationship to their working memory and comprehension. In the first study, 46 learners of Spanish at a U.S. university were asked to watch a short documentary clip in Spanish using either captions or no captions twice. Participants' comprehension and verbal working memory were measured after the second viewing. The results showed that captions significantly improved participants' comprehension. In addition, on average, participants in the captions group spent 74% of total reading time on captions in both viewing sessions. Learners with low comprehension spent more time rereading captions than learners with high comprehension scores. In the second study, 24 English learners with various L1 backgrounds at the same U.S. university were asked to watch the same video but with English audio and captions. Participants were found to spend 55% and 51% of total reading time on the captions during first and second viewings, respectively. Participants demonstrated similar caption reading behaviour for their first viewing

despite their comprehension scores. However, learners with high comprehension scores spent less time on captions on a second viewing than learners with low comprehension. In both studies, learners' verbal working memory did not significantly affect their caption-reading behaviour, but there was a trend demonstrating that learners with high verbal working memory tended to use captions less on a second viewing.

Some other eye-tracking studies have compared the reading of different subtitling types. In a study by Bisson et al. (2014), 54 English L1 speakers without any Dutch knowledge were divided into four groups and watched four short animation video clips in one of four conditions: captions, L1, reversed, and no subtitles (control group). An auditory incidental vocabulary test was administered after their viewing, but no significant group differences were found. Eye-tracking data showed that all the experimental groups spent significantly longer time on the subtitling area than the control group. No significant differences were observed between the captions and L1 subtitles groups, as revealed by fixation count (5.65 vs. 5.90), normalized total duration (0.43 vs. 0.51), and normalized skipped subtitles (0.09 vs. 0.04). However, the learners in the two groups using L2 audio paid more attention to the subtitling area compared to the reversed group, as revealed by all the eye-tracking measures. This may have been due to the participants' unfamiliarity with the soundtrack language. Significant differences were revealed in the time spent processing the image area across groups, as revealed by total fixation duration and the number of fixations: control group > reversed group > captions group > L1 subtitles group. However, it should be noted that the participants in their study did not have any prior knowledge of the L2, thus they might have had to rely more on L1 subtitles and less on captions for comprehension, making their findings less comparable to studies with more proficient L2 learners.

A comparison of the processing of different subtitles by more proficient EFL learners was conducted by Kruger et al. (2014). Sixty-eight South African EFL learners who used English as the main language in academic settings were asked to watch a 14-minute recorded lecture in English with either captions, L1, or no subtitles. Participants' comprehension of the lecture and potential cognitive load were examined after viewing. No significant differences were found among different subtitling conditions in terms of comprehension. A self-reporting cognitive load questionnaire suggested the benefits of using captions and L1 subtitles for lowering learners' frustration levels compared to no subtitles. Eye-tracking data showed that participants in the captions condition spent significantly more time (42.9% of total reading time) on the subtitling area than those in the L1 subtitles condition (20.3%). L1 subtitles were also skipped more often than captions. The authors explained that this might indicate participants' preference for English in an academic setting and their attempts to lower their cognitive effort by reducing the use of redundant L1 when L2 audio was understandable. The findings can also be attributed to more proficient L1 reading which required less processing time.

Muñoz (2017) further compared the reading of L1 subtitles and captions, with a focus on the potential effects of age and participants' L2 proficiency on their subtitling reading patterns. Forty Spanish-Catalan learners of English who varied in age and L2 proficiency were asked to watch two short English cartoon clips, one with L1 subtitles and the other with captions in a counterbalanced way. The findings suggested that participants who were younger or with a lower proficiency level seemed to rely more on on-screen text for comprehension. However, as noted by the researchers, there was a large overlap between the age and proficiency groups, which might make it difficult to attribute effects to just one factor. In general, eye-movement data comparing the reading

of captions and L1 subtitles revealed that captions were less likely to be skipped and were fixated on more often compared to the L1 subtitles condition, as revealed by the skip rate and total fixation duration. However, only median values for each age and proficiency group were reported, which makes the results less comparable to other studies reporting mean values. Moreover, no statistical analysis was done to explore differences in subtitling type.

To the best of my knowledge, Liao et al. (2020) have conducted the only eye-tracking study so far exploring the processing of bilingual subtitles. Given its close relevance to the present study, this study is reviewed in more detail. A within-group study was conducted to investigate viewers' attention allocation to the subtitling area and visual images, their comprehension, and the potential cognitive overload caused by using bilingual subtitles as compared to other subtitling types (i.e., captions, L1, and no subtitles). Twenty intermediate level Chinese postgraduates (with an average IELTS score of 6.74) at an Australian university were assigned to one of four groups and asked to watch four 5-minute BBC documentary clips in four subtitling conditions with their eye movements recorded. A 12-item cognitive load questionnaire adapted from Leppink (2014) was administered after each viewing session. It examined learners' intrinsic, extraneous, and germane cognitive load during viewing by using a self-reporting rating scale with 12 statements. Participants' comprehension was also tested using a free recall test after each viewing session. Total overall reading time % (as a measure of visual attention allocation) and average fixation duration (as an indirect measure of extraneous cognitive load) on the subtitling and image areas were calculated and compared across groups.

Eye-tracking findings showed that the participants spent the longest total reading time on bilingual subtitles (33.62%), which was significantly longer than the time spent

on reading L1 subtitles (21.55%), but not significantly different from reading captions (32.15%). When using bilingual subtitles, no significant differences were revealed between the time spent on the L1 and L2 lines. However, a closer inspection of eye-movement data showed that instead of evenly distributing attention to read both lines, participants were found to choose either Chinese or English as a dominant source to receive visual-verbal information. Comparisons were then made between the processing of L1 and L2 lines in bilingual subtitles separately from monolingual subtitles. Participants were found to spend significantly less time on L2 lines when using bilingual subtitles (15.29%) than using captions (32.15%), but a similar amount of time was spent on L1 lines (18.33%) and when using L1 subtitles (21.55%).

Researchers thus suggested that the use of bilingual subtitles resulted in a significant reduction in the use of L2 lines, but participants' reliance on L1 seemed to be stable, when compared to L1 subtitles. The researchers provided three explanations for participants' reliance on L1 lines. The first reason was due to the dominant role of the participants' L1, on which they could more easily rely for comprehension. Moreover, L1 lines were presented on the first line in bilingual subtitles which were more salient and attracted the viewers' attention. Additionally, L2 subtitles were more redundant than L1 subtitles as they repeated information presented in the audio and were more likely to be ignored. However, the authors did not explain why some participants did not choose L1 lines as their dominant visual-verbal input when using bilingual subtitles, given that they were theoretically easier to understand. In terms of the processing of image areas, less time was spent when using the three subtitling conditions (67.28% on L1 subtitles, 64.58% on captions, and 64.48% on bilingual) than no subtitles (73.29%), but no significant differences were found between the three subtitling groups, suggesting participants' stable reliance on images.

Regarding cognitive load, no significant differences concerning average fixation duration on the subtitling area were found between the three subtitling conditions (159 ms on L1 subtitles, 144 ms on captions, 150 ms on bilingual L1 lines, 140 ms on bilingual L2 lines), implying similar cognitive demands when reading the three subtitling types. Based on the similar average fixation durations on L2 lines in the captions and bilingual subtitles groups, the researchers also suggested that similar depth of processing of L2 lines was involved. The cognitive load questionnaire indicated that the use of bilingual subtitles did not induce more cognitive load than monolingual subtitles but had the lowest intrinsic and extraneous cognitive load across conditions, suggesting significant benefits for reducing cognitive load than no subtitles. This aligns with Li's (2016) questionnaire findings that bilingual subtitles did not cause cognitive overload, but rather could ease learners' learning burden and facilitate vocabulary learning. Liao et al. (2020) argue that when using bilingual subtitles, participants can actively select and integrate different information based on their comprehension needs. Comprehension results also showed that the bilingual subtitles group achieved the highest mean scores, although no significant differences were revealed across groups. However, the lack of significance may be due to the free recall test that was conducted in English, which might have interfered with the participants' English writing skills and thus could not fully capture learners' comprehension differences. While informative, there were several important limitations of this research which might constrain the generalisation of its findings. First, the findings were only based on eye-movement data from 16 participants. Moreover, the viewing material was only 5 minutes long for each subtitling condition, which might not be sufficient to capture participants' natural processing behaviour of on-screen text. Importantly, the within-subject design required each participant to watch four video clips in four different subtitling conditions in a

counterbalanced way. Potential order effects might have affected participants' viewing behaviour. For example, in three out of four groups, participants used bilingual subtitles immediately after using captions. It may be that after captioned viewing, participants were more familiar with reading L2, which increased their use of L2 lines in the following bilingual subtitled viewing. Order effects could be particularly influential given the use of short stimuli, which left limited time for participants to adapt to a different subtitling type.

In sum, the research reviewed above provides interesting findings concerning L2 learners' attention allocation during captioned/subtitled viewing. All the studies show that viewers processed on-screen text regardless of the subtitling condition. On-screen text did not prevent viewers processing images, but rather served as a support to aid viewers' understanding of the video. The use of captions/subtitles did not seem to add to cognitive load but actually seemed to ease L2 viewers' frustration levels compared to no subtitles (Kruger et al., 2014). Even with seemingly redundant written information, the use of bilingual subtitles was helpful in lowering L2 viewers' intrinsic and extraneous cognitive load (Liao et al., 2020). Viewers of different ages and proficiencies all seemed to be able to make use of the on-screen text based on their own needs by ignoring redundant input and paying more attention to information that could support their understanding (e.g., Kruger et al., 2014; Liao et al., 2020; Muñoz, 2017; Winke et al., 2013). As for the processing of different subtitling types, viewers were found to spend 20–41% of total reading time on L1 subtitles, while the percentage rises to 32–74% for reading captions. Captions were also less likely to be skipped than L1 subtitles (e.g., Kruger et al., 2014; Liao et al., 2020; Muñoz, 2017). Among L2 learners, less processing time was spent on L1 subtitles than on captions (e.g., Kruger et al., 2014; Liao et al., 2020; Muñoz, 2017) and bilingual subtitles (e.g., Liao et al., 2020).

Although the overall reading time on bilingual subtitles was similar to captions, significantly less time was allotted to L2 lines when using bilingual subtitles compared to using captions, whereas the time spent reading L1 lines was similar to using L1 subtitles (Liao et al., 2020). Previous studies have also shown that the processing of on-screen text can be affected by various factors. More advanced L2 learners seem to rely less on captions/subtitles while viewing compared to less proficient L2 learners (e.g., Muñoz, 2017; Winke et al., 2013). Content familiarity with the viewing material and the language distance between L1 and L2 might also trigger different processing behaviours (e.g., Winke et al., 2013).

Despite the important contribution of these findings to our understanding of how subtitled/captioned videos are processed, there are some important limitations and gaps that need to be acknowledged. First, overall, the number of eye-tracking studies exploring L2 learners' use of captions/subtitles is still limited. Crucially, only one study has explored the use of bilingual subtitles, with a limited number of participants ( $N = 16$ ) and short viewing material (5 minutes). Second, the comparability of findings across studies is limited due to the different eye-tracking measures reported, various types of viewing materials used (e.g., recorded lecture, cartoon, documentary, TV series), and the diverse profile of the viewers (e.g., viewers with/without the habit of using subtitles, viewers with/without prior knowledge of the L2 in the video soundtrack, viewers of different ages). Third, all the studies reviewed above used relatively basic statistical analysis (e.g., one-way ANOVA, t-tests, Kruskal–Wallis test) to analyse participants' average eye-movement data without controlling for other important covariates and random effects, which to some extent limits the generalisation of their findings. Last but not least, these studies only investigated L2 learners' attention allocation to the overall subtitling area rather than focusing on learners' processing of lexical items. Thus, the

relationship between the processing of vocabulary items while viewing and vocabulary learning gains is still understudied. To the best of my knowledge, to date, only one eye-tracking study has explored the processing of unknown vocabulary while viewing. It is reviewed in the next section after a review of eye-tracking studies focusing on reading.

#### ***2.4.4. Empirical Eye-Tracking Studies on Incidental Vocabulary Learning from Reading and Viewing***

Most of the eye-tracking studies investigating vocabulary acquisition have focused on incidental vocabulary learning from reading (Godfroid, 2020a). With the help of eye-tracking, these studies not only examined learning gains using offline tests, but also expanded their investigation to learners' online processing of unknown TWs. Their main findings are first summarised in this section, followed by a review of eye-tracking studies on incidental vocabulary learning from viewing.

First, a great number of eye-tracking reading studies have compared learners' online processing of unknown words to the processing of familiar words (e.g., Elgort et al., 2018; Godfroid et al., 2018; Godfroid et al., 2013; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez, Conklin, & Vilkaitė-Lozdienė, 2021; Williams & Morris, 2004). L1 reading studies have demonstrated that reading unknown words showed longer initial fixations, longer total reading times, and more regressions compared to reading familiar words (e.g., Williams & Morris, 2004). Similar findings have been reported in L2 reading studies, showing that novel L2 words were skipped less often than familiar words (e.g., Mohamed, 2018). Moreover, novel vocabulary items tended to attract longer reading times than known words, as measured by first fixation duration (e.g., Elgort et al., 2018; Godfroid et al., 2013; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021; Williams & Morris, 2004),

first-pass reading time (e.g., Elgort et al., 2018; Godfroid et al., 2013; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021; Williams & Morris, 2004), second-pass reading time (e.g., Godfroid et al., 2013; Williams & Morris, 2004), total reading time (e.g., Elgort et al., 2018; Godfroid et al., 2013, 2018; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021), fixation count (e.g., Elgort et al., 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021), go-past time (i.e., regression path duration) (e.g., Elgort et al., 2018), and number of regressions (e.g., Elgort et al., 2018; Mohamed, 2018; Williams & Morris, 2004). This difference in processing time is particularly salient in early encounters with words in the text (e.g., Elgort et al., 2018; Mohamed, 2018; Pellicer-Sánchez et al., 2021).

Second, many reading studies have also investigated how L2 learners' eye-movement patterns on unknown words change across multiple encounters in a text (e.g., Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021). Results from these studies have shown that the reading time on novel lexical items decreased with subsequent exposures, suggesting more fluent and automatic reading processes. Unknown words were found to be read significantly faster after 3–4 encounters (e.g., Godfroid et al., 2018; Pellicer-Sánchez, 2016), 5–7 encounters (e.g., Elgort et al., 2018), and 10–12 encounters (e.g., Mohamed, 2018) in different studies. Pellicer-Sánchez (2016) also found that L2 learners started to read nonwords in a similar manner to known words after eight encounters in the text, while in a more recent study, Pellicer-Sánchez et al. (2021) found that only after pre-teaching instructions was the processing of nonwords in the text similar to known words after eight encounters by L2 readers. Other researchers have suggested that more encounters are needed for unknown items to be processed like known words. Mohamed (2018) proposed a figure of 30, whereas Elgort et al. (2018) found that even after 40

encounters, differences in fixations and reading times were still noticeable between novel and familiar words. These inconsistent findings might be caused by the different degrees of support offered by contextual cues for unknown words (Elgort et al., 2018). While the specific number of encounters needed for unknown items to be processed like known words varies across studies, they all point to a general decrease in processing repeated unknown words in reading.

Most relevant to the present study is the investigation of the potential relationship between the online processing of unknown TWs and the learning outcomes of these TWs after reading (e.g., Godfroid & Schmidtke, 2013; Godfroid et al., 2013; Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021; Williams & Morris, 2004) and viewing (e.g., Montero Perez et al., 2015). By recording the online processing of novel lexical items during reading and viewing, eye-tracking enables us to further explore the relationship between learners' attention and their lexical gains. As argued by Pellicer-Sánchez (2020a): "the real drive to use eye-tracking in incidental vocabulary learning research is to find out whether eye-movement patterns can predict vocabulary learning gains" (pp. 140-141). This section now focuses on reporting the findings on this relationship.

Despite the increasing number of studies exploring learners' online processing behaviour in incidental vocabulary learning from reading, the relationship between readers' eye movements and vocabulary learning gains is far from settled. The first study to explore this potential relationship was conducted in an L1 context. Williams and Morris (2004) explored L1 English speakers' reading of tens of sentences containing either a familiar, a less familiar (with high or partial knowledge), or an unfamiliar TW for comprehension, while their eye movements were recorded. Multiple-choice meaning recognition tests were administered afterwards. A variety of

early and late measures were analysed. The findings showed that second-pass reading times on unfamiliar TWs positively correlated with meaning recognition gains, while first-pass reading times correlated negatively. The researchers attributed this negative relationship between initial processing and meaning recognition to the fact that little information could be used to guess novel words' meanings from the preceding context, whereas reanalysis of the following context might be useful for meaning inferences. However, their findings were limited to the L1 context.

Godfroid et al. (2013) were the first to explore the relationship between attention and vocabulary gains among L2 learners. Twenty-eight upper intermediate or lower advanced Dutch L2 learners of English were asked to read 20 different English paragraphs with 12 of them containing control words (real known words) and pseudowords. An unannounced multiple-choice gap-filling test was conducted immediately after their reading to examine participants' form recognition knowledge of the pseudowords. All the eye-tracking measures (i.e., first fixation duration, gaze duration, second-pass reading time, and total reading time) of the pseudowords showed positive correlations with the vocabulary test scores. However, only total reading time achieved statistical significance. Their findings suggested an overall positive relationship between L2 readers' attention and their learning gains, at least at the form recognition level. However, the two studies above used sentences or short paragraphs as reading material instead of longer reading texts, which might not be a good representation of natural reading.

Focusing on the reading of a short English story, Pellicer-Sánchez (2016) further tapped into this relationship by looking at different aspects of incidental vocabulary learning and word retention. Thirty-seven EFL learners with various L1 backgrounds and 36 L1 speakers of English took part in the experiment. A 2,300-word story was

written for this study containing six nonwords and six control words. A comprehension test, form recognition (select the correct spelling of TWs), meaning recall (via individual interview), and multiple-choice meaning recognition tests were completed after the reading. Delayed vocabulary posttests were administered with L2 participants two weeks later. The findings showed that only the total reading time for nonwords was positively related to immediate meaning recall gains, but the relationship did not hold for form or meaning recognition or other eye-tracking measures (i.e., first fixation duration, gaze duration, and fixation count).

In a very recent study focusing on the effect of pre-reading instruction, Pellicer-Sánchez et al. (2021) also explored the relationship between learners' total reading time for unknown words and their vocabulary learning gains. Ninety-two English L1 speakers and 88 English L2 speakers with various L1 backgrounds were asked to read a 2,290-word English text in one of four conditions (e.g., pre-reading instruction + reading, reading-only, instruction-only, and reading-baseline). Six pseudowords repeated eight times were embedded in the text, these were replaced by known words in the reading-baseline condition. Participants' comprehension and vocabulary learning were assessed immediately after reading. The findings showed that while controlling for different conditions, cumulative total reading time for the pseudowords significantly predicted meaning recognition, with longer processing times leading to higher learning gains. However, no significant results were obtained for form recognition or meaning recall.

The predictive role of total reading time in incidental vocabulary learning has also been documented by Mohamed (2017) and Godfriad et al. (2018). Following earlier studies, Mohamed (2017) invited 42 advanced L2 English learners with various L1 backgrounds to read a modified short novel containing 20 pseudowords and 20 known

words for comprehension while their eye movements were recorded. Vocabulary form recognition, meaning recall, and meaning recognition tests were administered after reading. The findings of generalized mixed models showed that total reading time was a strong positive predictor for all three vocabulary tests, while first fixation durations only showed a positive correlation with form recognition, and first-pass reading time only correlated positively with meaning recall. The author explains that different cognitive processes were involved in different eye-movement measures. Early measures could largely reveal learners' attention to word form, which related more to form recognition gains. Total reading time marked the total attention paid to a word, which could be more helpful to predict learning in both form and meaning aspects.

It is important to note that all the studies reviewed so far used modified, unauthentic, and relatively short texts, which can differ from the types of materials learners engage with. In order to address this limitation, Godfroid et al. (2018) compared L1 and L2 learners' incidental vocabulary learning from reading five chapters of an authentic English novel with 29 naturally embedded Dari words. Nineteen L1 speakers and 35 advanced L2 learners were asked to read the chapters for comprehension with their eye movements recorded. Form recognition, meaning recall, and meaning recognition tests were used to measure their learning gains. The findings revealed that the summed total reading time for TWs related positively to meaning recognition and recall gains, suggesting that a longer processing time might reflect participants' word inference effort. However, it should be noted that compared to other studies using nonwords as replacements for known words in the text, the meanings of the Dari words which were naturally embedded into this English novel could be more easily inferred from the context. English translations or explanations were directly

provided for some of the Dari words in their research, which might have accounted for higher incidental learning gains.

In contrast to the positive relationship reported in the studies reviewed so far, other studies have revealed no significant relationship between L2 learners' processing time for unknown words and learning gains. In a recent study by Ouyang, Huang, and Jiang (2020), 45 high-intermediate Chinese learners of English were asked to read a 671-word English text from a past TOEFL reading examination containing 17 unfamiliar words. Participants read with either no glosses or with L1 glosses while their eye movements were recorded. Immediate meaning recall and meaning recognition tests were administered after reading. Eye-movement data revealed that participants in the non-glossed group spent significantly longer time processing the TWs than the glossed group. However, no significant relationship between participants' processing time for unknown single TWs during reading (as measured by first fixation duration, gaze duration, total fixation counts, and total fixation duration) and their vocabulary scores in the non-glosses condition was revealed. The predictive role of processing time for TWs was only significant in the L1 glosses condition. The researchers thus attributed the findings to the fact that participants in the non-glossed group did not notice the TWs, but the presence of the L1 glosses enhanced learners' noticing of the TWs during reading, which further contributed to greater learning gains. However, it should be noted that instead of using original eye-tracking data, this study used Z-scores (i.e., a standard score measures the standard deviation between the mean and a raw score) in their data analysis, which might have less statistical power for capturing any potential relationship. Moreover, although care was taken in selecting the TWs, no pretests were included to ensure participants' unfamiliarity with those words, which might account for different reading patterns.

The relationship between L2 learners' processing of unknown words in a text and their vocabulary knowledge was also explored by Elgort et al. (2018), using different types of vocabulary measures. Forty high-intermediate to advanced level L2 learners of English were asked to read two English novel chapters containing 14 low-frequency L2 TWs varying in frequency of occurrence. Incidental vocabulary learning gains were measured after reading using a sentence-reading posttest (i.e., reading the TWs in 14 semantically neutral sentences while eye movements were recorded) and a meaning recall test. Participants' prior familiarity with the TWs was examined following the vocabulary posttests. Participants' processing of the TWs in the sentence-reading posttest was compared to their processing of the last occurrence of each TW in the main text to detect any processing differences. The results demonstrated the establishment of orthographic representations of the TWs during the text reading (as revealed by similar first-fixation duration and regression path duration in processing), but the lexical-semantic representations of the TWs were weak and contextually-dependent (as revealed by more fixations and longer reading time on the TWs in the sentence-reading posttest). No significant relationship was revealed between the processing time for the TWs in the sentence-reading posttest and their meaning recall gains, indicating a lack of relationship between learners' processing of the TWs (after exposure and potential learning) and their learning gains. However, care should be taken when comparing these findings to other eye-tracking studies since the relationship between the participants' initial processing of the TWs in the experimental text and their meaning recall gains was not reported in this study.

In sum, research exploring the relationship between eye movements in reading and vocabulary gains has demonstrated inconclusive findings. Apart from Ouyang et al. (2020) and Elgort et al. (2018), who reported no relationship, most studies have

reported some sort of relationship but varied in the different eye-tracking measures and vocabulary tests that showed a link. Total reading time seems to be a positive predictor of incidental vocabulary learning gains, especially for meaning (e.g., Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021). However, its correlation with form recognition is inconsistent across studies since some of them have reported a positive correlation (e.g., Godfroid et al., 2013; Mohamed, 2018) while others have not (e.g., Godfroid et al., 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021). Contradictory findings have been reported concerning the predictive role of early eye-tracking measures (e.g., first-pass reading time and first fixation durations) in vocabulary gains. Mohamed (2018) found a positive effect, whereas Williams and Morris (2004) reported a negative effect. Second-pass reading time also seems to relate positively to form recognition (Godfroid et al., 2013) and meaning recognition gains (Williams & Morris, 2004). These findings, though inconclusive, seem to suggest a positive predictive role of learners' attention and their vocabulary learning gains, but it is not always strong enough to be captured in different studies, and it is likely to be affected by different factors.

The incongruence in previous findings could relate to three potential causes. First, the aforementioned research applied various experimental designs. While some of them used authentic reading texts (e.g., Godfroid et al., 2018; Mohamed, 2018; Ouyang et al., 2020), others used texts designed for the purposes of the study (e.g., Godfroid et al., 2013; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021; Williams & Morris, 2004). These texts also varied in length and text difficulty, potentially leading to differences in processing and word inferencing. Moreover, participants in some of the studies shared a similar L1 background (e.g., Godfroid et al., 2013; Ouyang et al., 2020), while in others they had diverse L1 backgrounds (e.g., Godfroid et al., 2018; Mohamed, 2018;

Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021), which might also have affected learners' L2 reading behaviour. Second, when exploring the potential relationship between learners' eye movements and learning scores, some studies only used relatively basic statistical approaches such as non-parametric statistics (Kruskal-Wallis, Mann-Whitney *U* tests) and correlation (e.g., Ouyang et al., 2020; Pellicer-Sánchez, 2016; Williams & Morris, 2004). In these analyses, an averaged mean for each participant or each item was used, instead of taking both subject variables and item variables into account. Since the eye-movement data in these designs are nested within each participant and encounters are nested within each item, it is important to account for variance at both the participant-related and item-related levels in the analysis in order to produce more powerful and generalisable findings (Cunnings, 2012; Godfroid, 2020a; Mohamed, 2018).

To the best of my knowledge, only one eye-tracking study has explored the processing of vocabulary in captioned/subtitled viewing. Montero Perez et al. (2015) explored the effects of different types of captioning and test announcements on vocabulary learning and on their viewing processes. Fifty-one high-intermediate learners of French at a Flemish university were asked to watch two short authentic French audio clips (in total 9 minutes) while their eye movements were recorded. Participants were divided into four groups which differed in: 1) types of captioning: full captions (FC) or keyword captions (KC), and 2) vocabulary test announcements: intentional (INTEN) or incidental (INCID). Participants' vocabulary size and prior knowledge of the 18 target items (13 single words and 5 multiword units) were tested before the treatment. Form recognition, clip association, meaning recall, and multiple-choice meaning recognition posttests were conducted after the viewing session. In terms of vocabulary gains, the KC groups significantly outperformed the FC groups

on form recognition, and INTEN groups performed significantly better than INCID groups on meaning recall. Meaning recognition scores were not analysed due to potential guessing effects. Participants' vocabulary size positively correlated with their meaning recall gains. Eye-tracking data, measured by first-pass reading time, second-pass reading time, and total fixation duration, revealed that, in general, the INTEN groups had longer fixations on the target items than the INCID groups, as revealed by second-pass reading time and total fixation duration. The KC groups had longer first-pass reading time on the target items than the FC groups. In the incidental groups, second-pass reading time and total fixation duration on target items did not differ much in the KC and FC groups, indicating that the learners paid similar amounts of attention to the target items regardless of the captioning condition. However, in the intentional groups, the KC group spent significantly more time on the TWs than the FC group. The results suggest the influence of both enhancement techniques on learners' attention allocation to the target items. The relationship between eye-tracking patterns and form recognition test scores was also explored. Eye movements could to some limited extent predict word learning in the FC groups, but not for the KC groups. To be specific, the predictive role of eye movements was more salient in the FCINTEN group as longer total fixation duration and second-pass reading time could predict form recognition scores, suggesting that the reanalysis and attention paid to the target items indicated learners' effort to commit the words to memory. However, for the FCINCID group, only first-pass reading time could predict form recognition gains, while longer second-pass reading times led to lower learning gains. This finding suggests that a longer processing time might indicate processing problems instead of learning intention. Despite its importance in uncovering the relationship between word processing and learning in subtitled viewing, it should be noted that the predictive role of attention was

only explored at the level of form recognition and did not extend to meaning aspects due to the low meaning recall gains ( $M = 1.98$ ). As revealed in eye-tracking studies on reading, L2 learners' attention to unknown words tend to be closely related to the learning of word meanings (e.g., Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016). Therefore, it is also worth exploring this potential relationship further in subtitled viewing.

As reviewed in both eye-tracking studies on reading and viewing, it can be noticed that the link between processing time of unknown vocabulary and learning gains is far from settled. Montero Perez et al.'s (2015) findings seem to support previous reading studies (e.g., Mohamed, 2018) showing a potential predictive role for learners' early processing of unknown words facilitating knowledge of word form, but the lack of significance for total fixation duration and the negative role of second-pass reading time in learning gains cast more doubts on the potential relationship between learners' attention and incidental vocabulary learning gains. The findings also point to the potentially complex cognitive processes involved in subtitled viewing, which may account for the less straightforward relationship between attention and learning. Interpretations of the inconsistent findings in previous studies were largely based on speculation about learners' cognitive behaviour without empirical evidence. Eye-tracking is a versatile, unobtrusive, accurate, and objective method to probe learners' attention (Godfroid, 2020a). Eye movements are indicative of learners' cognitive processes, but different cognitive processes may underlie eye movements (Godfroid, 2020a; Montero Perez et al., 2015; Rayner, 1998) in these learning from reading/viewing conditions. As Pellicer-Sánchez (2020a) and Montero Perez et al. (2015) have mentioned, a longer processing time for a novel word could reflect learners' intention to learn and learners' effort to commit the word to memory, leading

to larger learning gains. However, a longer processing time for a word could also imply learners' processing difficulty or unsuccessful attempt to decode the novel word, which may be reflected in lower learning gains. As pointed out by Montero Perez et al. (2015), "[eye-tracking] data only provide information on the amount of attention involved in the learning process but do not inform us about learners' 'engagement' (Schmitt, 2008, p, 338) with the TWs" (p. 325). Research needs to examine the different subprocesses that are involved in these vocabulary learning conditions and explore how learners engage with unknown vocabulary. Although attention is defined in this thesis as one element of engagement (as discussed in the following section), the importance of not only investigating learners' attention allocation to a word but also exploring other types of engagement has been underscored by eye-tracking researchers (e.g., Godfroid & Schmidtke, 2013; Montero Perez et al., 2015; Pellicer-Sánchez, 2020a).

## **2.5. Engagement**

The investigation of engagement in language learning is important since it is believed to be the key to drive learning (Philp & Duchesne, 2016). Exploring learners' engagement with language is helpful to explain "why some linguistic or language-related behaviours and attitudes seem to facilitate language learning and learning about language/s more than others" (Svalberg, 2009, p. 243). Regarding vocabulary learning, Schmitt (2008) also claims that "anything that leads to more and better engagement should improve vocabulary learning" (p. 339). Scholars seem to agree that engagement plays an important role in learning in general, and in vocabulary learning in particular, hence it is vital to first understand what engagement is and how it has been defined in the literature.

### 2.5.1. Definitions of Engagement

Due to its wide use in everyday language, *engagement* is considered a multifaceted construct with various acceptable meanings (Fredricks, Blumenfeld, & Paris, 2004; Hiver, Al-Hoorie, Vitta, & Wu, 2021; Mercer & Dörnyei, 2020). It has been defined differently in various research fields (Hiver et al., 2021; Svalberg, 2018). In education, *engagement* has been used contextually as “school engagement” (Fredricks et al., 2004) or “academic engagement” (Skinner, Kindermann, & Furrer, 2009), where “engagement” has been referred to as “the quality of a student’s connection or involvement with the endeavour of schooling and hence with the people, activities, goals, values, and place that compose it” (Skinner et al., 2009, p. 494). It can be divided into *behavioural*, *cognitive*, and *emotional* engagement, with a focus on students in school settings (Fredricks et al., 2004). In language learning and teaching, *engagement* is also considered to be “a meta-construct that unites many separate lines of research within the field [language learning]” (Zhou, Hiver, & Al-Hoorie, 2020, p. 78). In language classroom settings, *engagement* has been seen as “student engagement”, defined as “effortful learning through interaction with the teacher and the classroom learning opportunities” (Christenson, Reschly & Wylie, 2012: vi). However, it should be noted that this definition of engagement concerns its motivational dimension by fixating on learners’ engagement in learning language, instead of learners’ engagement with the language itself (Mercer & Dörnyei, 2020, p. 19).

Svalberg (2009) offered the first systematic discussion of the construct of *engagement* in the context of language learning and use. She adopted a language awareness perspective by focusing on *engagement with language*. Svalberg (2009) argued that the *engagement* construct involves *cognitive*, *affective*, and *social* aspects which to some extent overlap and are likely to affect each other. She defines

“engagement [with language]” as “a cognitive, and/or affective, and/or social state and a process in which the learner is the agent and language is object (and sometimes vehicle)” (p. 244). The three aspects have been illustrated as follows:

- Cognitively, the Engaged individual is alert, pays focused attention and constructs their own knowledge.
- Affectively, the Engaged individual has a positive, purposeful, willing, and autonomous disposition towards the object (language, the language and/or what it represents).
- Socially, the Engaged individual is interactive and initiating. (Svalberg, 2009, p. 247)

Svalberg’s definition of engagement seems to be the most elaborate to explain learners’ engagement with language. In line with Svalberg, Philp and Duchesne (2016) acknowledge the multidimensional and overlapping nature of engagement in tasks in the context of language learning classrooms. Apart from the three aspects identified by Svalberg (2009), Philp and Duchesne (2016) added a *behavioural* aspect, which they define as “time on task or participation” (p. 55). Also, they point to the necessity of not only capturing a single dimension of engagement (usually the cognitive aspect) but also taking the other three dimensions into account to understand the full complexity of engagement. Besides, they emphasize that, due to the multifaceted nature of engagement, it is important to define and study this concept in different contexts.

The need to clearly define engagement in empirical studies was echoed by Hiver et al. (2021), after having systematically reviewed 112 empirical studies published in the past 20 years exploring engagement in language learning. Hiver et al. (2021) argue that “engagement is a dynamic, multidimensional construct comprising situated notions of cognition, affect and behaviours – including social interactions – in which action is a requisite component” (p. 25). They point out that *action* is the central characteristic of engagement in learning and also emphasize the dynamic and malleable nature of engagement in learning. However, Svalberg (2009), Philp and Duchesne (2016), and

Hiver et al. (2021) consider language as a whole without looking at any specific components of it.

Focusing on vocabulary learning in particular, Schmitt (2008) adopted the term *engagement* with the purpose of encompassing all possible types of involvement during vocabulary learning. As he argues, “overall, it seems that virtually anything that leads to more exposure, attention, manipulation, or time spent on lexical items adds to their learning” (p. 339). Schmitt (2008) emphasizes the importance of students’ motivation, attitudes, and strategic behaviour in the vocabulary learning process. He also lists other factors that could facilitate vocabulary learning, including frequency of exposure, attention, noticing, intention to learn, requirement to learn, need to learn/use, manipulation, time and interaction spent on lexical items. Schmitt (2008) points out that these factors can be facilitated by teachers, materials writers and learners themselves. Although he put forward a range of factors that can facilitate vocabulary learning and uses engagement as an umbrella term, there is still no clear definition of engagement in L2 vocabulary learning and studies seem to have adopted different definitions. In addition, including the importance of other agencies (i.e., teachers and materials writers) in the definition of engagement, rather than focusing on L2 learners’ own subjective engagement with lexical items, might have made this notion even harder to define or operationalise in research.

### ***2.5.2. Other Relevant Theories About Engagement in Language Learning***

Without explicitly using the term engagement, there are a number of theories that have also attempted to investigate learners’ cognitive or affective engagement in language learning more precisely, including Craik and Lockhart’s (1972) Depth of Processing Hypothesis and Laufer and Hulstijn’s (2001; Hulstijn and Laufer, 2001)

Involvement Load Hypothesis (Schmitt, 2008, 2010). Both theories have been used frequently in the SLA field in an attempt to explain how different cognitive processes can contribute to information storage in memory or learning outcomes. Both of them discuss the different levels or components of learners' cognitive processing which is key for engagement in language learning.

#### **2.5.2.1. Depth of Processing Hypothesis**

In cognitive psychology, Craik and Lockhart (1972) put forward the Depth of Processing Hypothesis, which focuses on the processes involved in memory, and how they may relate to information retention. This hypothesis posits that the depth of the initial analysis of new information determines the durability of the memory trace. It proposes that new information can be stored better and longer in memory with deeper processing than with shallow levels of analysis (Craik & Lockhart, 1972). Here, greater *depth* indicates “a greater degree of semantic or cognitive analysis” (Craik & Lockhart, 1972, p. 675), in which the stimuli has been fully analysed, attended to, and more deeply encoded (Craik & Tulving, 1975). In contrast, stimuli that do not receive full attention or are only superficially processed may only account for temporary memory traces (Craik & Lockhart, 1972; Craik & Tulving, 1975). For example, processing the meaning of a new word leads to a deeper level compared with processing the word's phonological form (Craik & Tulving, 1975).

However, the Depth of Processing Hypothesis has been criticised for its vagueness in defining *level* and *depth*, resulting in difficulties in operationalisation (Hulstijn 2001; Laufer & Hulstijn, 2001). Leow (2015) later provided a more comprehensible definition of Depth of Processing: “the relative amount of cognitive effort, level of analysis, and elaboration of intake, together with the usage of prior knowledge, hypothesis testing,

and rule formation employed in decoding and encoding same grammatical or lexical item in the input” (p. 204). Leow (2015) applied this notion of engagement to vocabulary learning. As illustrated in Table 6, a lexical item is considered to be processed more deeply when a form-meaning connection is made, or a longer time and greater cognitive effort are involved to get the meaning of the target item. This may result in better storage in long-term memory. Empirical studies in SLA that have explored depth of processing in L2 development mainly used two types of research methods: 1) indirect methods, i.e., experiments with assumed conditions to trigger deeper processing, and 2) direct methods, i.e., using concurrent verbal reports or think-aloud protocols (Leow, 2015).

Table 6. Leow’s (2015) Operationalisation of Depth of Processing with Lexical Items (pp. 227-228)

Operationalization of Depth of Processing (DOP): Lexical Items

|             | <i>Level 1</i>                                                                                                                                                                                                                                                                                                                                  | <i>Level 2</i>                                                                                                                                                                 | <i>Level 3</i>                                                                                                                                                                                              |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|             | <i>Low depth of processing</i>                                                                                                                                                                                                                                                                                                                  | <i>Medium depth of processing</i>                                                                                                                                              | <i>High depth of processing</i>                                                                                                                                                                             |
| Description | Shows no potential for emerging form-meaning connection                                                                                                                                                                                                                                                                                         | Provides some evidence of processing target item                                                                                                                               | Provides evidence of making accurate form-meaning connection                                                                                                                                                |
| Descriptors | Reads target quickly<br>Translates the phrase to English but leaves the target in Spanish<br>Says s/he isn’t sure what it is<br>Says s/he will click something<br>Repeats the target item<br>Carefully pronounces target word<br>Does not spend much time processing target item<br>Low level of cognitive effort to get meaning of target item | Spends a bit more time processing target item<br>Makes a comment that indicates some processing of target item<br>Some level of cognitive effort to get meaning of target item | Spends time processing target item<br>Provides an accurate translation of target item or finds a different way to say almost the same thing<br>High level of cognitive effort to get meaning of target item |

### 2.5.2.2. Involvement Load Hypothesis

Building on the Depth of Processing Hypothesis, Laufer and Hulstijn (2001) introduced the Involvement Load Hypothesis with the aim of operationalising the construct of Depth of Processing for L2 vocabulary learning. The Involvement Load Hypothesis was initially proposed to explain vocabulary learning in incidental learning conditions. Laufer and Hulstijn (2001) introduced the term *involvement*, which is defined as “a motivational-cognitive construct which can explain and predict learners’ success in the retention of hitherto unfamiliar words” (p. 14). *Involvement* has three components, *need*, *search*, and *evaluation*. Instead of focusing on cognition merely in the information processing aspect, the *need* component tackles learners’ motivation and need to complete a specific task, which equates more to the affective aspect in Svalberg’s (2009) construct of engagement with language. Depending on the degree of drive, a moderate need is imposed by external agents (e.g., tasks or teachers’ requirements), which is distinguished from a strong need that is generated by learners themselves. The *search* and *evaluation* components are concerned with the cognitive processing of lexical items, with a particular focus on the form-meaning relationship. *Search* is the attempt to find the meaning of an unfamiliar word with external assistance (e.g., consulting a dictionary or teacher). *Evaluation* entails making comparisons of an unknown word with other words or its other meanings to assess the appropriateness of a word in its context. This hypothesis posits that better learning and retention can be achieved with higher involvement load, and tasks with more involvement components included are believed to be more effective for vocabulary retention.

Although the Involvement Load Hypothesis acknowledges both the cognitive processing and motivational aspect of L2 vocabulary learning, this hypothesis was originally put forward to evaluate task design, rather than focusing on L2 learners’

subjective, cognitive or affective engagement with lexical items during tasks (Schmitt, 2008). Therefore, like other indirect methods for measuring the Depth of Processing, it does not take learners fully into account (Leow, 2015; Schmitt, 2010). A task with high involvement load may still trigger varying degrees of subjective involvement when being completed by different L2 learners. Consequently, this hypothesis is less useful to explore learners' subjective engagement, which is the focus of the present research.

### **2.5.3. *Measuring Engagement***

Researchers have suggested the use of self-reports as a direct method to elicit learners' thoughts and tap into their internal processes during their engagement with language input (Hiver et al., 2021; Svalberg, 2018; Zhou et al., 2020). It is believed that "self-report methods are especially useful for measuring emotional and cognitive engagement, which tend to be elusive and less easily observable or inferred from external behaviors" (Zhou et al., 2020, p. 83). The two most common introspective research methods in SLA research are *think-aloud protocols* and *stimulated recall* (also called *retrospective interview*) (Dörnyei, 2007). The former requires collecting data concurrently with language production, and the latter is conducted after a language event with a prompt to support learners' memory retrieval (Gass & Mackey, 2017). Both of them are helpful in exploring the quality of learners' engagement with target items (Gass & Mackey, 2017; Philp & Duchesne 2016).

Think-aloud protocols are frequently used in reading studies. Its advantage is to provide valid data on participants' spontaneous task-related thoughts without corrupting or changing their memory (Ericsson, 2002; Leow, Grey, Marijuan, & Moorman, 2014). However, they are not suitable for research using eye-tracking during viewing due to the reactivity issue. In other words, verbalising one's thoughts increases the time on task,

resulting in the distortion of eye movements (Godfroid & Schmidtke, 2013). In addition, in natural viewing settings, the real-time nature of watching audio-visual material, where new information is continuously provided, does not allow participants to pause and verbalise their thoughts.

*Stimulated recall*, or *stimulated recall interview*, is an introspective research method to elicit the thought processes that occur while a learner is doing a task or an activity by asking the learner to verbalise those processes after the events with a prompt to stimulate their memory (Gass & Mackey, 2017). This research method has been frequently applied in cognitively oriented research but is not limited to exploring the cognitive aspects of L2 learning. It has been used, for example, to investigate attention/awareness/noticing, strategy use, motivation, processing, interaction, and reading/writing, etc. (Gass & Mackey, 2017).

Stimulated recall is suggested as a suitable research method to capture learners' cognitive engagement (Philp & Duchesne, 2016; Zhou et al., 2020). Its biggest advantage compared to think-aloud protocols lies in its unobtrusive characteristic. It is believed that if stimulated recall is conducted shortly after the task, then thoughts are still in short-term memory, and so more valid information can be attained by cuing subjects with specific material used in the experiment (Ericsson & Simon, 1993; Gass & Mackey, 2017; Rose, McKinley, & Briggs Baffoe-Djan, 2020). In practice, in order not to add to the cognitive load of participants and to collect more accurate data on their thoughts, it is important not to ask participants *why* they responded in a certain manner but let participants continue expressing their thoughts (Ericsson, 2002; Ericsson & Simon, 1993; Gass & Mackey, 2017). Also, it is recommended to use participants' L1 when possible, to avoid linguistic difficulties, so that participants can fully recall their thoughts without any language constraints (Gass & Mackey, 2017).

The use of stimulated recall to explore learners' cognitive processes has been challenged because of its validity and reliability (Gass & Mackey, 2017; Rose et al., 2020). In other words, whether stimulated recall can reflect the real thought processes of participants and whether participants can report their thoughts accurately. Therefore, to address the issues of mistaken memory retrieval and memory deterioration, stimulated recall should always be conducted as soon after the actual event as possible, and participants should be asked about their thoughts rather than explanations of their behaviour (Gass & Mackey, 2017; Rose et al., 2020). Also, it is recommended that stimulated recall should be conducted before any posttests in order to mitigate the impact of reactivity (Gass & Mackey, 2017). It is believed that with care in the design, collection and interpretation of data, stimulated recall can elicit veridical and reliable data on learners' cognitive processes (Gass & Mackey, 2017). In the present study, stimulated recall was used to examine learners' awareness and processing strategies as part of their engagement with unknown words. Replays of video intercepts and participants' own recorded eye movements were used to prompt their memory.

#### ***2.5.4. Empirical Studies Exploring L2 Learners' Engagement with L2 Vocabulary in Reading and Viewing***

Despite the multifaceted nature of engagement, most studies on L2 learners' engagement in vocabulary learning have mainly focused on L2 learners' cognitive engagement during reading and viewing. They have also used different methodologies and different conceptualisations of the construct of engagement. Among the many different elements of engagement identified by Schmitt (2008), the majority of vocabulary learning studies have focused on the examination of learners' attention to unknown items, awareness, and processing/learning strategies used to engage with unknown items in an L2 context.

The eye-tracking studies reviewed in section 2.4.4 are one of these main branches tapping into learners' attention to TWs. As reviewed in section 2.4.4, those studies examined the amount of attention learners paid to processing unknown lexical items in different learning conditions, such as reading (e.g., Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016) and viewing (e.g., Montero Perez et al., 2015). By triangulating eye-tracking data with self-reports, Godfroid and Schmidtke (2013) investigated not only the amount of learners' attention to unknown lexical items, but also their awareness of novel words. Other researchers have explored "word/lexical inferencing strategies" (Hu & Nassaji, 2012; Huckin & Bloch, 1993; Nassaji, 2003; Rott, 2000), "vocabulary learning strategies" (Lawson & Hogben, 1996; Sydorenko, 2010), and "word/lexical processing strategies" (Fraser, 1999; Rott, 2005), as part of learners' cognitive engagement with unknown words in L2 input. These studies on awareness and processing/learning strategies are now reviewed in turn.

Godfroid and Schmidtke (2013) explored L2 learners' noticing of novel words in reading and its relationship with learners' incidental learning gains by separating learners' attention (examined with eye-movement data) and awareness (examined with stimulated recall interviews). It is the only study to date that has triangulated data from eye movements, stimulated recall, and vocabulary test scores to explore the relationship between L2 learners' attention, awareness, and incidental vocabulary learning gains. In this study, 29 advanced EFL learners were asked to read 20 English paragraphs containing 12 pseudowords while their eye movements were recorded. Participants' knowledge of the pseudowords was assessed using an unannounced fill-in-the-gap vocabulary posttest, presenting the original sentences with the pseudowords missing and possible answers immediately after reading. After the posttest, participants were asked to report their awareness of the pseudowords. Using Tulving's (1983, 2002) framework,

awareness in this study was coded as three categories: *no awareness* (participants did not consciously remember the TW), *noetic awareness* (participants remembered the TW was somewhere in the text), and *autonoetic awareness* (participants remembered the TW in a particular sentence). Mixed-effects models revealed a significant, positive relationship between awareness and attention, with autonoetic awareness associated with an average of 306 ms extra total processing time on the pseudoword compared to unawareness. Moreover, both attention and awareness positively predicted word recognition, with awareness being the strongest predictor. Noetic and autonoetic awareness both predicted learning gains, with learning of vocabulary at 26.8% and 66.5%, respectively. Although informative, this study only focused on learners' attention and awareness, and only two levels were distinguished within the "with awareness" category. This study provides interesting insights about two components, amount of attention and level of awareness, that could be considered elements of engagement. It explored whether learners had attended to novel vocabulary, for how long, and whether they were aware of having encountered those words. However, the specific strategic behaviours that learners implemented to process novel words were not examined.

A number of vocabulary studies have investigated L2 learners' engagement with novel words in reading by exploring learners' processing/learning strategies and their effectiveness for vocabulary learning. Some studies have used think-aloud protocols to explore L2 learners' vocabulary inferencing strategies by asking learners to verbalize their thoughts and guess the meanings of TWs while reading. Learners' knowledge of TWs has also been examined using different vocabulary tests. As shown in Table 7, Huckin and Bloch (1993) list six types of word inferencing strategies used by three intermediate-level Chinese EFL learners. These strategies were further grouped as

successful and unsuccessful guessing, based on posttest results. Using context clues was found to be the strategy most relied on and led to the largest number of cases of successful guessing. Participants also frequently detoured around the word without making a guess (labelled as potholes). Nassaji (2003) identified six strategies and five knowledge sources used to infer word meanings by 21 intermediate ESL learners (for details see Table 7). Repeating (i.e., repeating out loud any portion of the text including the TW) was the most frequently used strategy, whereas verifying and self-enquiry contributed to the greatest inferencing success. World knowledge and morphological knowledge were the most frequently mentioned knowledge sources that participants used to infer word meanings, and they also led to the most inferencing success. Hu and Nassaji (2012) found four general word inferential strategies were used by 11 advanced ESL learners: form-focused, meaning-focused, evaluating, and monitoring strategies (see Table 7). Meaning-focused strategies were the most frequently used, followed by form-focused, evaluating, and monitoring strategies. TWs seemed to be better retained when using form-focused strategies, especially when they were combined with meaning-focused strategies. A negative relationship between the ease of word inferencing and learning retention was suggested. The three studies reviewed above paint a general picture of the most frequent types of strategies learners used to infer meanings of unknown words and how they might relate to vocabulary learning.

A few studies have also investigated the strategies used during reading with meaning support, for example using L1 and a dictionary. Lawson and Hogben (1996) asked 15 adult English learners of Italian to complete a vocabulary learning task by reading 12 Italian sentences each containing an unknown Italian noun. Half of the TWs were accompanied by L1 translations. Participants were asked to verbalise all their thoughts during their learning of unknown TWs, and do a meaning recall test afterwards.

Four main categories emerged from the think-aloud data with 15 subcategories: repetition, word feature analysis, simple elaboration, and complex elaboration (see Table 7 for details). The findings showed that repetition and simple elaboration were the strategies most frequently used and also contributed to better meaning recall. Findings also showed that when L1 clues were available, participants relied more on L1 translations to obtain word meanings and used the context in a more complex way than when no L1 translations were available. The frequency of using strategies positively related to learning gains. However, words that could be easily understood from the context required less attention and simple elaboration, resulting in less retention.

Fraser (1999) explored the lexical processing strategies used in reading with a dictionary. Eight French learners were trained to use different vocabulary processing strategies and read English texts with the help of dictionaries. Participants' engagement with unfamiliar vocabulary during reading was measured through stimulated recall by asking the thoughts they had at the first encounter with each word. Their meaning recall gains were examined one week after each reading session using a 5-point VKS. As shown in Table 7, participants' lexical processing strategies were generally grouped as consult (a dictionary), ignore, infer (word meaning), and no attention (not noticed). The findings showed that participants used strategies alone and also in combination. Findings revealed that, in general, participants were found to infer word meaning more than they consulted a dictionary, with both being more frequently reported than the ignored and no attention cases. Moreover, the combination of inferring and consulting strategies led to more correct inferences and high retention rates than using either strategy alone.

The findings reported by Lawson and Hogben (1996) and Fraser (1999) indicate that learners actively make use of meaning support (i.e., L1 translations or dictionaries)

to engage with unknown words in reading, and the use of meaning support is also combined with other strategies. However, it should be noted that the above-reviewed studies were set in a deliberate vocabulary learning context where participants were deliberately asked or trained to infer the meaning of TWs during reading. Therefore, the types and frequency of use of participants' strategies might be different from those implemented in an incidental learning setting, as previous studies have shown that informing L2 learners of the aim of vocabulary learning could influence their learning gains and their dictionary look-up behaviours while reading (e.g., Peters, 2007, 2009).

In an incidental learning setting, Rott (2000) listed 12 word inferencing strategies (seven local and five global strategies, for details see Table 7) used by eight low- and mid-intermediate English learners of German when they first encountered unknown TWs during natural reading. However, the frequency of use of each strategy was not calculated, possibly due to the small number of participants. Rott (2005) later investigated L2 learners' vocabulary processing strategies in different glossed reading conditions. Ten L1 English learners of German were randomly assigned to either a multiple-choice glosses (MCGs) condition or a single-translation glosses (STGs) condition to read a German passage containing four unknown TWs, each occurring four times. Think-aloud protocols were used to record participants' word processing strategies. A surprise VKS test and a multiple meaning recognition test were administered immediately after their reading and four weeks after. Two main types of strategies emerged from the think-aloud protocols data: *meta-cognitive strategies*, which represented relatively shallower processing and demonstrated learners' processing of the orthographic aspects of TWs. They included the simple use of glosses or without form-meaning connections. And *semantic elaboration strategies*, by involving need, search, and evaluation according to the Involvement Load Hypothesis

(Laufer & Hulstijn, 2001), showed learners' attempts to assign meaning to TWs by accessing and retrieving existing knowledge sources (see Table 7). The results showed that MCG readers integrated both strategies to establish form-meaning connections, while STG readers only used meta-cognitive strategies. Both groups performed similarly on an immediate vocabulary posttest, but the MCG group demonstrated better retention in a delayed posttest. Therefore, the authors suggested that words could be better retained when semantic elaboration strategies were used, which were deeper and more elaborate than meta-cognitive processing strategies for vocabulary learning.

The studies reviewed above reveal the types of L2 learners' engagement with unknown words while reading by looking at their attention and awareness (e.g., Godfroid & Schmidtke, 2013) and their most frequently used vocabulary learning/processing strategies (e.g., Fraser, 1999; Hu & Nassaji, 2012; Huckin & Bloch, 1993; Lawson & Hogben, 1996; Nassaji, 2003; Rott, 2000, 2005). Although informative, the different research designs should be considered when interpreting the findings. Only three studies have been conducted in incidental learning conditions (Godfroid & Schmidtke, 2013; Rott, 2000, 2005), with participants not being asked to learn novel words or infer their meaning, which is the condition examined in the present research. It should also be noted that learners' engagement with unknown words can also be affected by other factors, for example, participants' different proficiency levels and L1 backgrounds, the difficulty and length of reading materials, and the research design. Most importantly, these studies explored L2 learners' vocabulary processing strategies in a reading context, which may be different from the findings in L2 viewing.

Most studies exploring learners' engagement with vocabulary during viewing examined learners' attention to novel words via eye-movement data, as reviewed in section 2.4.4. To the best of my knowledge, Sydorenko (2010) is the only empirical

study that has explored learners' vocabulary learning strategies during viewing. Beginner L2 learners of Russian were asked to watch three 2-minute Russian video twice (the first time focusing on comprehension and the second time on vocabulary learning) in one of three viewing conditions: video + audio + captions, video + audio, and video + captions. A comprehension test and vocabulary tests (form recognition and meaning recall in both aural and written form) were administered after viewing. Participants' vocabulary learning strategies during viewing were measured in a final questionnaire. Two types of general strategies emerged: modality-specific strategies (including matching visual images with words, reading captions) and common vocabulary guessing strategies (including recognizing words that are similar to L1, using the roots of known words, paying attention to the verbal context, paying attention to grammar) (see also Table 7). Using visual images was the most frequently used strategy in all viewing conditions. Participants in the video + audio + captions group were found to use fewer general guessing strategies than participants in the other two groups. However, as the researcher noted, this study deliberately selected target items with well-matched visual support, which might not be the case in other videos. Therefore, it is still not clear how learners engage with novel words that vary in image support in a more natural viewing setting. In addition, this research only investigated the overall strategies used by participants during viewing using an open-ended questionnaire rather than focusing on the strategies for each individual unknown word. It is possible that their general strategies did not represent what happened with each individual word. Most importantly, this study only investigated strategy use with/without captions, and no previous studies have investigated learners' vocabulary processing strategies in viewing using L1 or bilingual subtitles. It is not clear whether, when learners' L1 is also available in the subtitling area, L2 learners still use the same

strategies, as mentioned by Sydorenko (2010), to engage with unknown words, or if they actively engage with L1 translations as reported in reading studies (e.g., Lawson & Hogben, 1996; Rott, 2005).

In sum, by reviewing the existing literature exploring L2 learners' use of processing/inferencing strategies during reading and viewing, an important aspect of engagement with unknown L2 words, it can be concluded that learners are likely to notice unknown words in L2 input (e.g., Godfroid & Schmidtke, 2013; Hu & Nassaji, 2012; Rott, 2000, 2005; Sydorenko, 2010). Moreover, they use different types of processing strategies to engage with unknown words, individually or in combination (e.g., Fraser, 1999; Hu & Nassaji, 2012; Huckin & Bloch, 1993; Rott, 2000). As can be observed in Table 7, in the reading context, the seemingly most frequently used processing strategy is inferring word meanings based on context (e.g., Fraser, 1999; Hu & Nassaji, 2012; Huckin & Bloch, 1993; Lawson & Hogben, 1996). However, when L1 is available during reading (e.g., L1 translations or glosses), learners tend to make use of L1 translations to aid their understanding and learning (e.g., Lawson & Hogben, 1996; Rott, 2005), which might potentially lead to shallower processing of unknown words in an incidental learning setting (e.g., Rott, 2005). In the viewing context, images tend to be the most frequently used strategy to engage with unknown words (Sydorenko, 2010). Studies exploring the relationship between learners' strategy use and word learning also emphasize the distinction between successfully inferring word meanings in context and the acquisition of word meanings from the context (e.g., Hu & Nassaji, 2012; Huckin & Bloch, 1993; Lawson & Hogben, 1996; Rott, 2000). No particular strategy has been found that always results in correctly inferring outcomes (Huckin & Bloch, 1993; Nassaji, 2003; Rott, 2000). As Table 7 shows, the types of strategies that contributed to the largest learning gains varied across different studies.

These findings are informative in revealing L2 learners' engagement, especially their strategic behaviours to engage with unknown words in reading and viewing. However, there are still several important gaps that need to be acknowledged. First, it is still unclear which type(s) of vocabulary processing strategies are most frequently used by L2 learners in incidental vocabulary learning settings. As mentioned above, most of the above-reviewed studies explored L2 learners' strategy use in deliberate vocabulary learning conditions (e.g., Fraser, 1999; Hu & Nassaji 2012; Sydorenko, 2010), which may be different from the processing strategies applied in incidental learning. Second, no previous studies have to date systematically explored L2 learners' engagement with unknown words in viewing, especially under different subtitling conditions. It is thus important to collect participants' self-reported data to better understand how they make use of different modes of input to engage with unknown words during subtitled viewing, which should also be useful in explaining their vocabulary learning gains. Third, as can be seen in Table 7, think-aloud studies exploring L2 learners' strategy uses were only based on very limited numbers of participants, ranging from three to 21. Fourth, as Table 7 shows, due to the inductive coding approach applied in all studies, the coding of strategies and terms used varied dramatically across studies. Researchers also adopted different approaches to categorise strategies by focusing on the success/failure of outcomes (Huckin & Bloch, 1993), the use of knowledge sources (Nassaji, 2003), vocabulary components/features (Hu & Nassaji, 2012), the use of different modalities (Sydorenko, 2010), or the complexity of elaboration (Lawson & Hogben, 1996). Some of the categories seem to have oversimplified learners' engagement with unknown words by defining strategies as simple/elaborate or successful/unsuccessful. Additionally, it is controversial to draw conclusions based dichotomously defined strategies, and makes it difficult to make comparisons drawn across studies. Crucially,

the studies reviewed in this section either focused on learners' attention, awareness or learners' processing strategies, without examining all three aspects together. A more comprehensive examination of L2 learners' engagement during viewing is therefore needed.

Table 7. Summary of Contextual Word Processing Strategies Reported in Previous Research (in chronological order)

|                          | <b>Research purpose &amp; Research method of engagement</b>                                                                                                                                        | <b>Coding of strategies</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <b>The most frequent strategies</b>                                                                                                                                                                                 | <b>Strategies contributed to higher word learning gains</b>                                                                                        |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Huckin and Bloch (1993)  | To explore the guessing strategies of three intermediate Chinese EFL learners during English reading<br><br>Think-aloud protocols                                                                  | Successful strategies:<br><ul style="list-style-type: none"> <li>• Used context</li> <li>• Latent word knowledge</li> <li>• Morphological analysis</li> </ul> Unsuccessful strategies:<br><ul style="list-style-type: none"> <li>• Mistaken ID (mistook the word for another that resembles it)</li> <li>• Potholes (where the subject simply avoided the word in his written translation)</li> <li>• Incomplete knowledge (relied on partial knowledge of a word and were unable to guess the full meaning of the word)</li> <li>• Morphological analysis</li> <li>• Used context</li> </ul> | <ul style="list-style-type: none"> <li>• Used context (especially using local clues)</li> <li>• Mistaken ID (mistook the word for another that resembles it)</li> </ul>                                             | <ul style="list-style-type: none"> <li>• Used context (especially using local clues)</li> </ul>                                                    |
| Lawson and Hogben (1996) | To investigate the vocabulary learning strategies used by 15 adult English learners of Italian during the reading task of L2 sentences (with and without salient L1 clues) for vocabulary learning | <ul style="list-style-type: none"> <li>• Repetition: reading of related words, simple word rehearsal, writing word and meaning, cumulative rehearsal, testing</li> <li>• Word feature analysis: spelling, word classification, use of</li> </ul>                                                                                                                                                                                                                                                                                                                                              | <ul style="list-style-type: none"> <li>• Repetition (especially Reading of related words and Simple rehearsal)</li> <li>• Simple elaboration (especially Sentence translation and Simple use of context)</li> </ul> | <ul style="list-style-type: none"> <li>• Repetition (especially Simple rehearsal)</li> <li>• Simple elaboration (especially Appearance)</li> </ul> |

|               | <b>Research purpose &amp; Research method of engagement</b>                                                                                                               | <b>Coding of strategies</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <b>The most frequent strategies</b>                                                                                                                                                  | <b>Strategies contributed to higher word learning gains</b>                                                                                     |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
|               | Think-aloud protocols                                                                                                                                                     | <p>suffixes</p> <ul style="list-style-type: none"> <li>• Simple elaboration:<br/>sentence translation, simple use of context (i.e., learners provided a possible meaning for the word before checking the L1 clues or no specific reference is made to any other word(s) in the sentence), appearance similarity, sound link</li> <li>• Complex elaboration:<br/>complex use of context, paraphrase, and mnemonic use (i.e., learners made a serious attempt to derive word meaning from the sentence by making reference to meaning or features of other words in the sentence, perhaps suggesting possible alternative meanings for the TW)</li> </ul> |                                                                                                                                                                                      | <p>similarity and Sound link)</p> <ul style="list-style-type: none"> <li>• Complex Elaboration (especially Paraphrase and Mnemonic)</li> </ul>  |
| Fraser (1999) | To investigate the lexical processing strategies used by eight French learners of English when they encounter unfamiliar vocabulary while reading and the impact of these | <ul style="list-style-type: none"> <li>• Consult (the dictionary)</li> <li>• Ignore</li> <li>• Infer (the word meaning):<br/>L1/L2 word identification (i.e.,</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>• Infer (with “Sense creation” used more than “L1/L2 word identification”) &gt; Consulted &gt; Ignored or Paid no attention to the</li> </ul> | <ul style="list-style-type: none"> <li>• Infer (especially L1/L2 word identification)</li> <li>• Using both inferring and consulting</li> </ul> |

|                | <b>Research purpose &amp; Research method of engagement</b>                                                                                                                                                                                                          | <b>Coding of strategies</b>                                                                                                                                                                                                                                                                                                                                                                                               | <b>The most frequent strategies</b>                                                    | <b>Strategies contributed to higher word learning gains</b>                        |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
|                | strategies on vocabulary learning<br><br>Stimulated recall                                                                                                                                                                                                           | associations based on the phonological or orthographic form of the word), sense creation (i.e., using cues from the context)<br><br>• No attention (not noticed)                                                                                                                                                                                                                                                          | unfamiliar words<br>• Learners were likely to combine the infer and consult strategies | strategies had high retention rate than using either inferring or consulting alone |
| Rott (2000)    | To explore the relationship between processing of reading, word inferencing strategies used, and the incidental vocabulary gains during reading a short article containing multiple occurred TWs among eight English learners of German<br><br>Think-aloud protocols | Local strategies:<br>inferences using immediate context, demonstrates awareness of TWs but does not infer meaning, breaks TW into its two components, skips TW, tried different word categories, use of grammatical knowledge, monitor<br><br>Global strategies:<br>use of background knowledge, elaborating on the context, lexically correct inferences, conceptual inferences, circumlocution of the meaning of the TW | Not mentioned due to the limited number of participants                                | Not mentioned                                                                      |
| Nassaji (2003) | To examine 21 ESL learners' use of strategies and knowledge sources used in L2 lexical inferencing and their                                                                                                                                                         | Strategy types:<br>• Repeating (repeating out aloud any portion of the text)<br>• Verifying (examining the inferred                                                                                                                                                                                                                                                                                                       | Strategies:<br>Repeating > Analogy > Verifying > Monitoring > Self-inquiry > Analyzing | Not mentioned                                                                      |

|  | <b>Research purpose &amp; Research method of engagement</b>               | <b>Coding of strategies</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <b>The most frequent strategies</b>                                                                                                                 | <b>Strategies contributed to higher word learning gains</b> |
|--|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
|  | <p>relationship with inferential success</p> <p>Think-aloud protocols</p> | <p>meaning based on the wider context)</p> <ul style="list-style-type: none"> <li>• Analyzing (analyzing the word into various parts/components)</li> <li>• Monitoring (showing a conscious awareness of the problem or the ease/difficulty of the task)</li> <li>• Self-inquiry (asking oneself questions about the parts that already inferred)</li> <li>• Analogy (based on word sound/form similarity with other words)</li> </ul> <p>Knowledge sources:</p> <ul style="list-style-type: none"> <li>• Grammatical knowledge (using knowledge of grammatical functions or syntactic categories)</li> <li>• Morphological knowledge (using knowledge of word formation and structure)</li> <li>• Knowledge of L1 (translating or finding a similar word in the L1)</li> <li>• World knowledge (using knowledge of the content/topic beyond the text)</li> <li>• Discourse knowledge (using knowledge about the relation between/within sentences and the devices that make connections between</li> </ul> | <p>Knowledge sources:</p> <p>World knowledge &gt; Morphological knowledge &gt; Grammatical knowledge &gt; Discourse knowledge &gt; L1 knowledge</p> |                                                             |

|                       | <b>Research purpose &amp; Research method of engagement</b>                                                                                                                                                                                                                                  | <b>Coding of strategies</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | <b>The most frequent strategies</b>                                                                                                                                                         | <b>Strategies contributed to higher word learning gains</b>                                                                                                                                           |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                       |                                                                                                                                                                                                                                                                                              | the different parts of the text)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                             |                                                                                                                                                                                                       |
| Rott (2005)           | To explore word processing strategies used by 10 English learners of German during text-enhanced reading (i.e., with multiple-choice glosses and single-translation glosses) and the relationship between processing strategies and word learning and retention<br><br>Think-aloud protocols | <ul style="list-style-type: none"> <li>• Meta-cognitive strategies: referring to the glosses, monitoring learners' own comprehension of the TWs by showing their uncertainty of the meaning of the TWs saying, "I am not sure", without the attempts to provide a meaning</li> <li>• Semantic elaboration strategies: referring to the gloss during the non-glossed encounters, searching for meaning in the context of the TW, using existing knowledge sources to retrieve a synonym of the L1 gloss, and accessing background knowledge to make meaning of the TW</li> </ul> | <p>Multiple-choice glosses condition:<br/>Integrated both meta-cognitive and semantic-elaborative strategies</p> <p>Single-translation glosses condition:<br/>Meta-cognitive strategies</p> | <p>For initial learning gains: meta-cognitive strategies were similar as semantic elaboration strategies</p> <p>For learning retention (4-week delayed posttest): semantic elaboration strategies</p> |
| Hu and Nassaji (2012) | To examine the relationships between 11 advanced ESL learners' ease of inferencing word meanings from context, the inferential strategies they use, and their retention of these words                                                                                                       | <ul style="list-style-type: none"> <li>• Form-focused strategies: analysing, analogy, repeating</li> <li>• Meaning-focused strategies: using textual clues, discourse, paraphrasing</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                  | Meaning-focused strategies > Form-focused strategies > Evaluating strategies > Monitoring strategies                                                                                        | A positive and significant relationship between meaning-focused strategies and inferencing success;                                                                                                   |

|                  | <b>Research purpose &amp; Research method of engagement</b>                                                                                                                                                                             | <b>Coding of strategies</b>                                                                                                                                                                                                                                                                                                       | <b>The most frequent strategies</b> | <b>Strategies contributed to higher word learning gains</b>                                       |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|---------------------------------------------------------------------------------------------------|
|                  | Think-aloud protocols                                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>• Evaluating strategies: self-inquiry, confirming/disconfirming, commenting</li> <li>• Monitoring strategies: stating the failure/difficulty, re-attempting</li> </ul>                                                                                                                     |                                     | A positive and strong significant correlation between form-focused strategies and word retention. |
| Sydorenko (2010) | To examine the effects of input modality (i.e., video, audio, and captions) on vocabulary learning, attention to input, and vocabulary learning strategies during viewing among 26 beginner L2 learners of Russian<br><br>Questionnaire | <ul style="list-style-type: none"> <li>• Modality-specific strategies: matching visual images with words, reading captions</li> <li>• Common vocabulary guessing strategies: recognizing words that are similar to L1, using the roots of known words, paying attention to verbal context, paying attention to grammar</li> </ul> | Matching visual images with words   | Not mentioned                                                                                     |

### 2.5.5. Operationalisation of Engagement in the Current Research

As reviewed in section 2.5.1, to date, no elaborate definition of engagement has been provided in vocabulary learning research. Besides, *engagement* is a multidimensional and complex construct which involves various but overlapping aspects. As introduced in section 1.2, the present research aims to explore how learners' attend to unknown words and how their processing of those words may relate to their vocabulary learning gains. Svalberg's (2009) *engagement with language* with a focus on the cognitive aspect is the clearest and most relevant definition of engagement for the present research, because it: 1) focuses on learners' engagement with language items in particular; 2) provides an explicit explanation of the components of engagement which can serve as a framework to analyse the different aspects of engagement, unlike Schmitt's (2008) vaguer definition of engagement with vocabulary; and 3) in particular focuses on learners' subjective experience with language.

As explained in section 2.5.1, according to Svalberg (2009), *cognitive engagement* is defined as: "cognitively, the Engaged individual is alert, pays focused attention and constructs their own knowledge" (p. 247). With a particular focus on vocabulary, this definition can be expanded in the present research to answer: how much focused attention participants paid to unknown words, how alert they were to unknown words (operationalised as attention and awareness), and how participants constructed their knowledge of unknown words during viewing (operationalised as vocabulary processing strategies).

As reviewed in section 2.2.2.3, researchers have advocated a distinction between attention and awareness. Eye-tracking is an unobtrusive method that has been widely used in a number of empirical studies in SLA to explore attention (Godfroid, 2020a; Montero Perez, 2020b). It is believed to be a valid and accurate operationalisation of

attention, in particular the process of covert attentional orienting (Godfroid, 2020a; Wright & Ward, 2008). The high temporal and spatial resolution of eye-tracking enables researchers to distinguish early and late measures of processing and thus sensitively capture learners' attention allocation (Godfroid, 2020a; Leow, 2014). However, eye-tracking has been criticised for not probing the quality of cognitive operations. Consequently, it has been suggested that eye movement data should be triangulated with measures of learners' awareness and other cognitive behaviours to explore cognitive processes (Leow, 2014; Godfroid et al., 2013).

In this research, the definition of *awareness* is only used at the *noticing* level, which is a surface level phenomenon referring to “the conscious registration of the occurrence of some event” (Schmidt, 1995, p. 29). This should be distinguished from awareness at the level of *understanding*, which is a higher level of awareness, and “implies recognition of a general principle, rule or pattern” (Schmidt, 1995, p. 29). Since the present study only focuses on awareness at the noticing level, the terms *awareness* and *noticing* are used interchangeably in this thesis. According to Leow (2014), the key characteristic of awareness is *reportability*. Awareness at the noticing level is operationalised as the availability for self-reporting either during or immediately after exposure to input (Leow, 2015; Schmidt, 1990). Think-aloud protocols, which offer insights into learners' cognitive processes and the strategies employed (Leow, 2014), have been adopted by most of the above-reviewed studies to measure learners' awareness and strategic behaviours. However, as mentioned in section 2.5.3, two major disadvantages of think-aloud protocols obstruct its application in the current research – its intrusive nature and the reactivity issue. Therefore, in order not to interrupt participants' viewing process or distort eye-movement data, stimulated recall is considered a suitable and reliable method to check learners' awareness of each TW and

explore learners' processing strategy use when encountering each unknown word.

Vocabulary processing strategies refer to the strategies that learners use to construct their knowledge of a word when exposed to L2 input, irrespective of their learning intention. This is operationalised as learners' verbalized reports in the stimulated recall interviews in the present study. This term agrees with Cohen's (1990, p. 5) definition of *learning strategies* as "learning processes which are consciously selected by the learner", emphasising the elements of choice and consciousness. However, this term does not underscore the intentional learning feature, which should be distinguished from *strategic vocabulary learning*, defined as "an intentional, dynamic and iterative process for the effective, efficient, and even enjoyable learning of vocabulary" (Gu, 2020, p. 271).

In sum, cognitive engagement is operationalised in the present study as: attention, awareness, and vocabulary processing strategies. Attention is investigated using an eye-tracking method, and participants' awareness and processing strategies are explored using stimulated recall.

## **2.6. Summary and Research Questions**

To summarise, vocabulary research has shown that both breadth and depth of vocabulary knowledge are essential for L2 learners, and a considerable number of studies have revealed the potential of incidental vocabulary learning from different modes of L2 input to develop learners' vocabulary knowledge. While the majority of studies to date have focused on learning from reading, recently, an increasing number of studies have explored learning from viewing. These studies have shown that using different types of subtitles can foster incidental vocabulary learning. Bilingual subtitles, a widely used and popular subtitling type among Chinese learners of English, however,

have received very little attention. The literature reviewed in this chapter has shown some notable gaps in the use of bilingual subtitles for incidental vocabulary learning. First, studies examining the effectiveness of bilingual subtitles for incidental vocabulary learning are very scarce, and inconsistent findings have been revealed across studies. These inconsistent reported could be due to learners' differential use of the sources of input available in bilingual subtitles. So far, only one study has explored L2 learners' general processing of different subtitling areas when using bilingual subtitles with the help of eye-tracking. Thus, it is still not clear how L2 learners' pay attention to available subtitles and unknown words during bilingual subtitled viewing. Moreover, eye-tracking studies have shown that the amount of attention paid to unknown words during reading and viewing seems to relate to their learning gains, but no previous studies have explored this relationship with bilingual subtitled viewing. In addition, no empirical studies so far have investigated learners' different types of cognitive engagement with unknown words during viewing with different subtitling conditions. By combining offline vocabulary tests to examine learners' learning gains with eye-tracking and stimulated recall to probe learners' cognitive engagement with unknown words, this research can let us paint a comprehensive picture of the effects and potential of bilingual subtitles. The combination of different research methods can also help to further unravel the complex and unclear relationship between processing measures and vocabulary gains. To address these gaps, the present research aims to answer the following questions:

RQ1: To what extent does the use of bilingual subtitles increase learners' vocabulary knowledge (as measured by form recognition, meaning recall, and meaning recognition tests), compared to captions, L1, and no subtitles?

RQ2: To what extent does the use of bilingual subtitles facilitate learners' viewing comprehension, compared to captions, L1, and no subtitles?

RQ3: How do learners allocate their attention to different areas (i.e., images, subtitling areas, unknown TWs and/or their L1 translations) during bilingual subtitled viewing, compared to captions, L1, and no subtitles, as revealed by eye-tracking data?

RQ4: Do learners' online processing of unknown TWs and their corresponding L1 translations predict their vocabulary gains in different subtitling conditions (i.e., captions, bilingual, and L1 subtitles)?

RQ5: How do learners engage with unknown TWs, as measured by their level of awareness and use of vocabulary processing strategies, during bilingual subtitled viewing, compared to captions and L1 subtitles as reported in stimulated recall interviews?

RQ6: Do participants' awareness and processing strategies of unknown TWs at the group level corroborate their vocabulary learning gains and their attention allocated to those words?

## **Chapter 3. Methodology**

The purpose of this study is to explore L2 learners' incidental vocabulary learning through subtitled viewing, with a particular focus on bilingual subtitles. A QUAN + qual mixed methods design (Dörnyei, 2007; Johnson & Christensen, 2012) was applied to answer the research questions presented in section 2.6. Learners' learning outcomes were examined through offline tests, and their processing behaviours were investigated via eye-tracking methods and stimulated recall. In this chapter, I first report results of an initial online questionnaire study conducted to confirm the significance and widespread use of bilingual subtitles in the context being investigated in this thesis. Then, the rationale for using a mixed methods design, as well as the results of two pilot studies are presented. The chapter finalises with a description of participants, research design, instruments, procedures, and data analyses used in the main study.

### **3.1. Initial Online Questionnaire Study**

As stated in section 1.1, the investigation presented in this thesis was based on the observation that bilingual subtitles are indeed a very common subtitling type in the Chinese context and thus they deserve more research attention. In order to further confirm this and provide a stronger rationale and justification for the present study, an online questionnaire study was conducted to understand the use and preference of different types of subtitles among Chinese learners of English. Ethical approval was obtained to collect the online questionnaire data as part of the main study (explained in section 3.4.2). An online questionnaire was designed using a Chinese online questionnaire builder *WJX.cn* (<https://www.wjx.cn/>) and was administered using

snowball sampling before the main study. Participants provided their consent through the online platform before accessing the questions. Participants' basic personal information, English learning background, self-rated English proficiency level, and habits of watching English audio-visual materials were collected using closed-ended questions. Their frequency of use and preferences of using different subtitling types were asked using nine Likert scale questions (see Appendix S1). The internal consistency reliability of the Likert scale was examined, with Cronbach's Alpha coefficient at .56, which is common for short scales with limited number of items (Pallant, 2016).

Two-hundred-and-six Chinese learners of English completed the questionnaire, and 202 valid questionnaire results were received. The participants varied in gender (34.16% male, 65.84% female), age (61.88% ages 18–25, 24.75% ages 26–30, 8.91% ages 31–40, 3.47% ages 41–50, 0.99% ages 50+), and self-rated proficiency level (6.93% at beginner level, 34.16% at low intermediate, 40.10% at high intermediate, 11.88% at low advanced, and 6.93% at high advanced). The results showed that most of the participants enjoyed watching English audio-visual materials both as an entertainment (79.21%) and in the English classroom (80.20%). Their average rate for the frequency of watching audio-visual materials was 4.25 (Max = 6) for entertainment, and 2.75 (Max = 6) for learning in English classroom settings. Participants also reported to use on-screen text while watching English audio-visual materials frequently ( $M = 4.85$ , Max = 6). Moreover, 99.00% of the participants had heard about bilingual subtitles. In terms of the preference of subtitling types, in general, bilingual subtitles had the highest average score at 5.07 (Max = 6), followed by captions ( $M = 4.11$ ), L1 subtitles ( $M = 3.63$ ), and no subtitles ( $M = 2.10$ ). In addition, bilingual subtitles were also rated as the most frequently used subtitling type ( $M = 4.58$ , Max = 6), followed by

L1 subtitles ( $M = 3.73$ ), captions ( $M = 3.25$ ), and no subtitles ( $M = 2.45$ ). Bilingual subtitles were the most popular among beginners to low advanced level Chinese EFL learners. However, for participants who rated their proficiency as high advanced, bilingual subtitles were rated as the second frequently used ( $M = 3.29$ ,  $\text{Max} = 6$ ) and the second preferred ( $M = 4.00$ ,  $\text{Max} = 6$ ) subtitling type following captions ( $M = 3.71$  and  $4.43$ , respectively). However, this finding should be treated with caution due to the limited number of participants who fell into the high advanced category (6.93%).

Overall, the results of this initial questionnaire study confirmed that: 1) EFL learners in China are familiar with bilingual subtitles; 2) bilingual subtitles are indeed the preferred option among EFL learners in China; and 3) bilingual subtitles are the most frequently used subtitling type by EFL learners in China, at least among the learners who regarded themselves as beginner to low advanced level EFL learners. Having provided empirical evidence to further confirm the need to examine the effectiveness of bilingual subtitles in the Chinese EFL context, I now move to the discussion of the research design followed in the main study reported in this thesis.

### **3.2. Research Design**

A mixed methods design was chosen for the present investigation. Mixed methods research is one of the three main methodological approaches, i.e., quantitative, qualitative, and mixed methods, often dated back to the late 1980s. It has been widely used in the social and behaviour sciences (Creswell & Plano Clark, 2011; Teddlie & Tashakkori, 2009) and has gained popularity in applied linguistics over the past two decades (Dörnyei, 2007; Hashemi & Babaii, 2013). The central premise of mixed methods research is that the combination of quantitative and qualitative research can

offset the weaknesses of each approach and complement the strengths of each other (Creswell & Plano Clark, 2011; Dörnyei, 2007).

In this thesis, I adopt an early definition of *mixed methods* provided by Greene, Caracelli, and Graham (1989):

In this study, we defined mixed methods designs as those that include at least one quantitative method (designed to collect numbers) and one qualitative method (designed to collect words), where neither type of method is inherently linked to any particular inquiry paradigm. (p. 256)

Different categorisations of mixed methods designs have been proposed. Two typological principles have been widely used to distinguish the various mixed methods designs, i.e., the *sequence* (concurrent or sequential) and the *dominance* (equally dominant or unequally dominant) (Dörnyei, 2007; Johnson & Christensen, 2012). *Sequence* refers to whether the qualitative and quantitative phases of the study take place at approximately the same time (i.e., concurrent) or they are occurred over time (i.e., sequential). Also, in a sequential study, the second phase is often designed by addressing the findings of the first phase to facilitate its development, which is different from the concurrent design (Creswell & Plano Clark, 2011). *Dominance* refers to whether the qualitative and quantitative parts of the study have approximately equal priority or weight (i.e., equally dominant) in answering the research questions and interpreting the results, or one part has more emphasis than the other (i.e., unequally dominant). Consequently, for a mixed methods study containing only two components (qualitative and quantitative), the aforementioned two typological principles result in nine possible combinations:

Table 8. Mixed Methods Design Matrix (adapted from Johnson and Christensen, 2012, p. 435)

|                  | Concurrent  | Sequential  |
|------------------|-------------|-------------|
| Equally dominant | QUAL + QUAN | QUAL → QUAN |

|                    | Concurrent  | Sequential  |
|--------------------|-------------|-------------|
| Unequally dominant | QUAL + quan | QUAN → QUAL |
|                    |             | QUAL → quan |
|                    | QUAN + qual | qual → QUAN |
|                    |             | QUAN → qual |
|                    |             | quan → QUAL |

In Table 8, “qual/QUAL” and “quan/QUAN” stand for qualitative and quantitative research respectively. Capital letters indicate more dominant or increased weight. The plus sign (+) represents a concurrent sequence, while the arrow (→) represents a sequential sequence (Dörnyei, 2007; Johnson & Christensen, 2012).

The present study adopted a QUAN + qual design (Dörnyei, 2007; Johnson & Christensen, 2012), which has also been labelled as *concurrent triangulation strategy* (Creswell, 2003) or *convergent parallel design* (Creswell & Plano Clark, 2011) by different researchers. In this design, the quantitative and qualitative data collection took place in one phase of the research, with quantitative part having a dominant status as shown in Table 8. In the present research, the quantitative part consisted of offline vocabulary tests and online eye-tracking data, while the qualitative part included stimulated recall data. The same participants took part in both quantitative and qualitative parts. As illustrated in Figure 9, data analyses and discussion were completed separately for each part, and the results were integrated at the late stage to corroborate each other (Dörnyei, 2007).

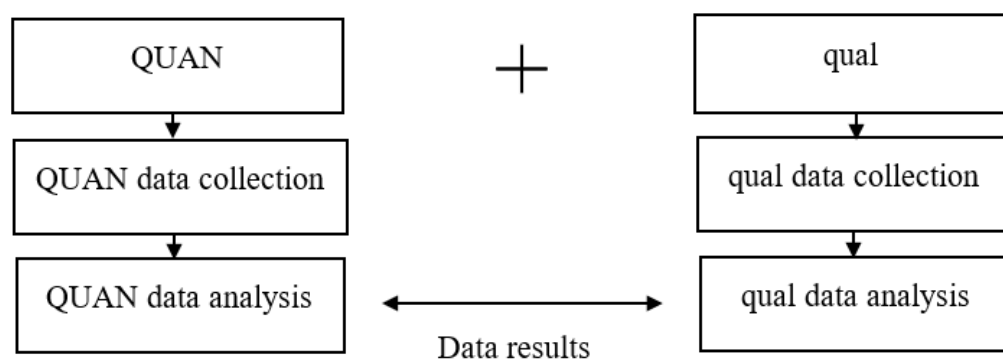


Figure 9. QUAN + qual Research Design (Creswell, 2003, p. 214)

My rationale for using this quantitative dominant mixed approach was twofold. First, the quantitative data provided statistical results to show learners' learning gains, their attention allocation during viewing and the potential relationship between gains and attention, but it only provided a limited view of learners' cognitive engagement. Adding the qualitative data helped to provide in-depth exploration of learners' underlying cognitive engagement, which helped to gain a better understanding of learners' processing of the unknown words during subtitled viewing and offset the limitations of the eye-tracking method. Moreover, triangulation of the stimulated recall findings and the quantitative results could also illuminate the quantitative findings and facilitate the understanding of the relationship between learners' cognitive engagement and learning gains by adding additional evidence.

The research procedures of the present study are illustrated in Figure 10. The whole procedure of the main study included two sessions. In the first session, participants completed a series of vocabulary pretests and a vocabulary size test. In the second session, the viewing activity was completed with participants' eye movements recorded, followed by a comprehension test and vocabulary posttests. Stimulated recall was administered after posttest tests in order to avoid contaminating learners' posttest scores. These were followed by a 3K vocabulary levels test (Schmitt, Schmitt, &

Clapham, 2001) and a background questionnaire. Specific data collection and analysis procedures used in the present study are elaborated in section 3.4.7.

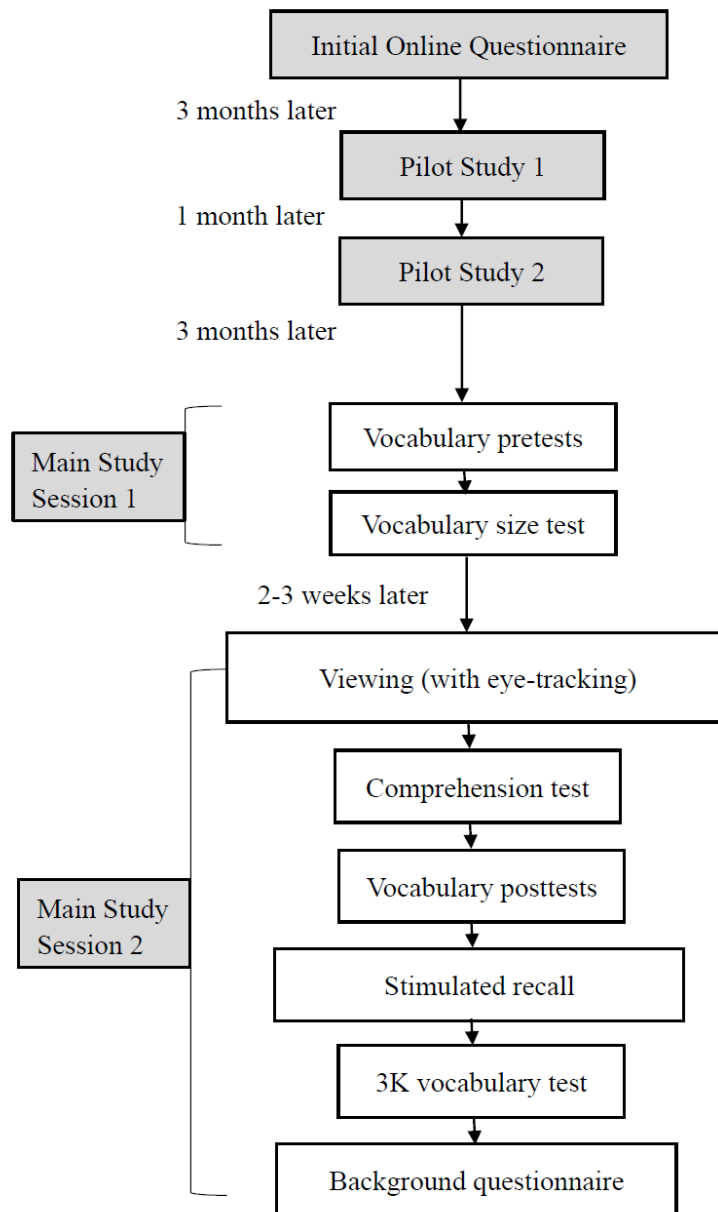


Figure 10. Visual Diagram of the Research Procedures in the Present Study

### 3.3. Pilot Studies

#### 3.3.1. Pilot Study One

The first pilot study was conducted with the aim to check the suitability of the selected video material and the offline test design. Six female PhD students whose L1 was Chinese and who were studying at the UCL Institute of Education participated in the pilot. Their prior knowledge of the TWs was examined using a checklist test one week before the viewing. During the viewing session, participants were asked to watch the stimulus under one of the four subtitling conditions (i.e., captions, L1, bilingual, and no subtitles) in a quiet room. They were asked to watch on their own laptops without eye-tracking. Vocabulary posttest (in the VKS format, see Figure 2 for an example) and comprehension test (true/false and multiple-choice formats) were completed after the viewing. No stimulated recall was conducted. Informal group interview was held with all participants to receive feedback about the procedures and measurements. Results of the first pilot demonstrated that participants did learn some of the target vocabulary while viewing, and they had no problems in understanding the video. All participants confirmed that the video was interesting, and that the captions and subtitles were accurate and easy to follow. While the use of VKS test captured some learning, it was found not precise enough to capture knowledge at the recognition level, leading to small learning gains. Based on the results, a multiple-choice meaning recognition vocabulary test was added to capture more subtle vocabulary gains.

### **3.3.2. Pilot Study Two**

A more thorough pilot study was conducted one month after the first pilot study with seven female Chinese learners of English doing their master courses at the UCL Institute of Education. Participants' average age was 26.43 ( $SD = 2.44$ , 95% CI [24.17, 28.69]) and their average IELTS score was 7.29 ( $SD = 0.64$ , 95% CI [6.90, 7.87]). All of them reported the use of captions/subtitles when watching English videos, and all of

them had heard about bilingual subtitles. Participants were asked to follow the whole experimental procedure as presented in section 3.4.7. They attended two experiment sessions in the eye-tracking lab individually. In the first session, they completed vocabulary pretests (including a VKS test and an added multiple-choice meaning recognition test) and a vocabulary size test. In the second session, they were randomly assigned to one of the four subtitling groups (i.e., captions, L1, bilingual, and no subtitles) to watch the stimulus with their eye movements recorded, and completed a series of tests and the stimulated recall interview after the viewing. The test results were analysed using descriptive statistics. No inferential statistical analyses were run due to the limited sample size.

Seven modifications regarding the test design and stimulated recall questions were made for the main study, based on the results of the second pilot study and the comments received from my upgrade examiners: 1) the time gap between pre- and posttest was extended to two weeks in order to diminish the pretest effects; 2) the VKS test used in both pre- and posttests was changed to a combination of form recognition and meaning recall test in the main study, to make the results more comparable to previous studies and to make it easier to draw conclusions about learners' vocabulary development; 3) five TWs that were known to the participants or appeared more than five times in the stimulus were removed from the list of target items, and seven phrasal verbs were removed from the vocabulary tests, resulting in a final 24 single TWs; 4) for the comprehension test, the true/false comprehension questions were deleted or changed into multiple-choice format in order to reduce guessing, resulting in 34 multiple-choice questions; 5) a few options of the comprehension test were modified based on participants' feedback; 6) the speed of participants' eye-movement recording stimuli, used as prompt in the stimulated recall, was reduced to 50% of the original videos, due

to participants' reported difficulty in following their rapid eye movements to recall their thoughts; 7) lastly, stimulated recall questions were fine-tuned to make sure it was clear to the participants that the focus was on their thoughts during their initial viewing, instead of describing their eye movements.

### **3.4. Main Study**

#### **3.4.1. Participants**

One-hundred-and-twelve Chinese learners of English participated in the study. All participants reported normal or corrected-to-normal vision. They were all Chinese learners of English studying at the UCL Institute of Education with various academic backgrounds. Data from three participants who did not complete the experiment were discarded from the analyses. This resulted in a total of 95 female and 14 males. Their ages ranged between 18 and 34 years ( $M = 23.42$ ,  $SD = 2.47$ , 95% CI [22.93, 23.87]), and their length of living in the UK varied from 1 to 24 months ( $M = 2.24$ ,  $SD = 2.79$ , 95% CI [1.70, 2.78]). Their overall IELTS scores varied from 5.5 to 8 ( $M = 6.84$ ,  $SD = 0.61$ , 95% CI [6.67, 6.90]), which approximately corresponds to B2 to C1 levels in CEFR, according to the IELTS official guidelines (<https://www.ielts.org/about-ielts/ielts-in-cefr-scale>). Their mean vocabulary size was 6274.31 word families ( $SD = 1704.65$ , 95% CI [5950.67, 6597.95]). Differences in vocabulary size were accounted for in the analysis.

As reported in the background questionnaire, most of the participants liked watching English audio-visual materials as an entertainment (79.25%) or in the English classroom (83.96%). Most of the participants used captions/subtitles during viewing, with the most frequently used subtitling type being bilingual subtitles ( $M = 4.44$ ,  $Max = 6$ ), followed by captions ( $M = 3.14$ ), L1 subtitles ( $M = 3.03$ ), and no subtitles ( $M = 2.18$ ).

Similarly, bilingual subtitles were also reported to be the most preferred subtitling type ( $M = 5.09$ ,  $\text{Max} = 6$ ), followed by captions ( $M = 4.25$ ), L1 subtitles ( $M = 3.48$ ), and no subtitles ( $M = 2.31$ ). All participants had heard about bilingual subtitles.

The 3K level of the Vocabulary Levels Tests (Schmitt et al., 2001) was administered to ensure the comprehensibility of the selected viewing material for participants. The rationale for this selection procedure is based on findings from Webb and Rodger's (2009) corpus analysis of television programs, which suggested that incidental vocabulary learning might occur if learners knew the 3,000 most frequent word families. Sixteen participants scored below 24 out of 30 on the 3K test, failing to meet the mastery threshold, as suggested by Xing and Fulcher (2007). Among the 16 participants, two of them stated difficulty of understanding the video content after the viewing and one did not complete the whole viewing session. Their data were discarded from analysis. The statistical analyses were conducted with and without the remaining 13 participants and results remained the same. Therefore, their data were kept for further analysis. Consequently, 106 participants (93 female and 13 male) were included in the offline data analysis. Due to issues with the online data quality (see section 4.1.2), data from 6 participants were further removed from the online data analysis, resulting in a total of 100 participants. Stimulated recall interview was only administered with the participants in the captions, L1, and bilingual subtitles groups with a total number of 82 participants (see section 3.4.6).

Table 9 provides descriptive statistics for the participants' reported IELTS test scores and their vocabulary size test results. The comparability of the four groups was examined using One-way ANOVA analyses. The results showed that there were no group differences in terms of participants' vocabulary size,  $F(3, 102) = 0.01$ ,  $p = .10$ , and their overall IELTS scores,  $F(3, 102) = 0.51$ ,  $p = .68$ .

Table 9. Descriptive Statistics for IELTS Scores and Vocabulary Size by Group

| Group           | Captions<br>( <i>n</i> = 27) |                               | L1 subtitles<br>( <i>n</i> = 24) |                               | Bilingual<br>subtitles<br>( <i>n</i> = 30) |                               | No subtitles<br>( <i>n</i> = 25) |                               |
|-----------------|------------------------------|-------------------------------|----------------------------------|-------------------------------|--------------------------------------------|-------------------------------|----------------------------------|-------------------------------|
|                 | <i>M</i> ( <i>SD</i> )       | 95% CI                        | <i>M</i> ( <i>SD</i> )           | 95% CI                        | <i>M</i> ( <i>SD</i> )                     | 95% CI                        | <i>M</i> ( <i>SD</i> )           | 95% CI                        |
| IELTS overall   | 6.80<br>(0.62)               | [6.55,<br>7.04]               | 6.84<br>(0.62)                   | [6.58,<br>7.10]               | 6.68<br>(0.66)                             | [6.44,<br>6.93]               | 6.83<br>(0.55)                   | [6.61,<br>7.05]               |
| IELTS Listening | 7.11<br>(0.98)               | [6.72,<br>7.50]               | 7.38<br>(1.03)                   | [6.95,<br>7.81]               | 6.90<br>(1.01)                             | [6.52,<br>7.28]               | 7.17<br>(0.87)                   | [6.82,<br>7.51]               |
| IELTS Reading   | 7.56<br>(0.86)               | [7.22,<br>7.90]               | 7.52<br>(0.92)                   | [7.14,<br>7.90]               | 7.33<br>(0.99)                             | [6.97,<br>7.70]               | 7.57<br>(0.79)                   | [7.26,<br>7.89]               |
| IELTS Writing   | 6.11<br>(0.47)               | [5.93,<br>6.30]               | 6.08<br>(0.49)                   | [5.88,<br>6.28]               | 6.15<br>(0.42)                             | [5.99,<br>6.31]               | 6.20<br>(0.42)                   | [6.04,<br>6.37]               |
| IELTS Speaking  | 6.15<br>(0.77)               | [5.84,<br>6.45]               | 6.26<br>(0.71)                   | [5.97,<br>6.55]               | 6.13<br>(0.51)                             | [5.94,<br>6.32]               | 6.19<br>(0.42)                   | [6.02,<br>6.35]               |
| Vocab size      | 6329.6<br>3<br>(1713.9<br>3) | [5651.6<br>2,<br>7007.6<br>4] | 6266.6<br>7<br>(1675.0<br>5)     | [5559.3<br>5,<br>6973.9<br>8] | 6316.6<br>7<br>(1713.4<br>5)               | [5676.8<br>5,<br>6956.4<br>8] | 6356.0<br>0<br>(1797.0<br>1)     | [5614.2<br>3,<br>7097.7<br>7] |

### 3.4.2. Research Ethics

Research ethics approval was obtained from the Research Ethics Committee at the UCL Institute of Education (Data Protection Registration Number: Z6364106/2018/11/09). The approved ethics form, information sheet and consent form are included in Appendix S2 to S4. Participation was voluntary and participants received a £10 compensation for their participation in the main study. At the beginning of the first session of the study, a general introduction together with an information sheet was provided to the participants. Since the main aim of this research was to examine incidental vocabulary learning, it was important for participants not to be

aware of the vocabulary posttests before the viewing. Thus, participants were told that the purpose of the study was to examine the effects of viewing for L2 learning and comprehension, without specifying that it was about vocabulary learning. Each participant was asked to read the information sheet which explained the purpose of the study, the procedure and duration of the study, and the measures taken to maintain their privacy and confidentiality. Participants were allowed to take a break after the viewing session if necessary. Participants were also informed of their right to withdraw from the study at any time. Each participant's agreement to participate in the study was obtained by signing a consent form (see Appendix S4). At the end of the study, clear explanations of the real purpose of this research were given.

### **3.4.3. Materials**

#### **3.4.3.1. Viewing Material**

After inspecting several possible videos, the BBC documentary *Animal Odd Couples* was chosen as the viewing material for the study. Four authentic video excerpts from this documentary (in total 23 minutes, 3488 words) were extracted and put together using the video editing software Corel VideoStudio Pro 2018 (Corel, 2018). This documentary consists of several journeys taken by a wildlife biologist Liz Bonnin, with the mission to find out why animals of different species build up unusual close bonds with each other. The chosen clips included four pairs of animal couples: brotherhood between the bear, the lion, and the tiger; relationship between the rhinos and their friends; friendship between the dog and the deer; and relationship between the cat and the ducks. These four excerpts were initially chosen because of the amount of potentially unknown words by participants.

Documentary is considered to be an appropriate type of viewing material for its rich imagery support (Rodgers, 2018), clear oral presentation, and the potential of directing learners' noticing of the word forms, and it was considered more appropriate than fiction genres (Gilabert et al., 2018). Different from the typical documentary characterised by a single narrator and slow-moving pace, this documentary also includes interactive interviews between different speakers, which is considered more engaging for participants. After the viewing activity, I orally asked participants if they liked the video, and 89% of them said that they had enjoyed the content of the selected video. The length of the video can also be considered appropriate according to Rodgers and Webb (2017), who recommend viewing materials ranging from 22 to 42 min to obtain sufficient L2 aural input.

The video scripts were analysed using the Range software (Nation & Heatley, 2002), with the British National Corpus (BNC) as the reference corpus. The results showed that after deleting proper nouns and the exclamation words, the most frequent 3K words provided a coverage of 95.47%, and the most frequent 6K words provided a lexical coverage of 98.00%. The chosen video was considered appropriate for the participants since it was expected that international students starting their postgraduate studies in the UK should know the first 3K words, reaching a 95.47% lexical coverage that was sufficient for adequate comprehension (Webb & Rodgers, 2009). As reported in section 3.4.1, participants' knowledge of the first 3K words was later confirmed with the vocabulary levels test.

#### **3.4.3.2. Captions and Subtitles**

The original video script in English was downloaded online and then manually translated into Chinese. The translation process did not follow a word-for-word strategy

but followed the main principles of communicative and semantic translation (Newmark, 1981). I ensured that all the words in the scripts were accurately translated, with particular attention to the TWs. Cambridge Online Dictionary (English-Chinese version) (<https://dictionary.cambridge.org/dictionary/>) was used to check the translation accuracy of TWs. The translation was compared to the online amateur translation (retrieved from <https://www.bilibili.com/video/av21620515/>), then checked by three Chinese L1 speakers fluent in English, and piloted twice with 13 advanced Chinese learners of English to ensure its accuracy. Minor modifications were taken at the syntax level to achieve more natural and smooth translation style.

The captions and L1 subtitles were added to the video using SrtEdit (PortableSoft, 2012) and Corel VideoStudio Pro 2018 (Corel, 2018) software. The production of captions and subtitles was in accordance with the BBC Subtitle Guidelines Version 1.1.7 (<http://bbc.github.io/subtitle-guidelines/>). Font size was designed around 4.8% of active video height, which is in accordance with the recommended presentation font size for desktop computers. The subtitling was displayed at the bottom of the screen during the interlocutors' speaking or narration. To meet requirements of the eye-tracking experiment, all the L1 subtitles and captions were kept within one line with the max line length being 68% of the width of the screen for each frame. In the bilingual subtitles condition, L1 and L2 lines were presented simultaneously with the L1 above L2 lines, which is in accordance with the common presentation of bilingual subtitles in China. Line breaks were taken at a sign of punctuation like a full stop, comma or dash. Long sentences were broken into several lines at natural and logical points. The line breaks were kept the same across conditions. English was presented in Calibri font, and Chinese was presented in Songti (宋体) font, both in size 35. The average duration of subtitle presentation was 2,168 ms ( $SD = 4454$ , 95% CI = [1790, 2546]). Four versions

of the video were created, one for each of the subtitling conditions (i.e., captions, L1 subtitles, bilingual subtitles, no subtitles). Example screenshots of the four subtitling conditions are presented in Figure 11.

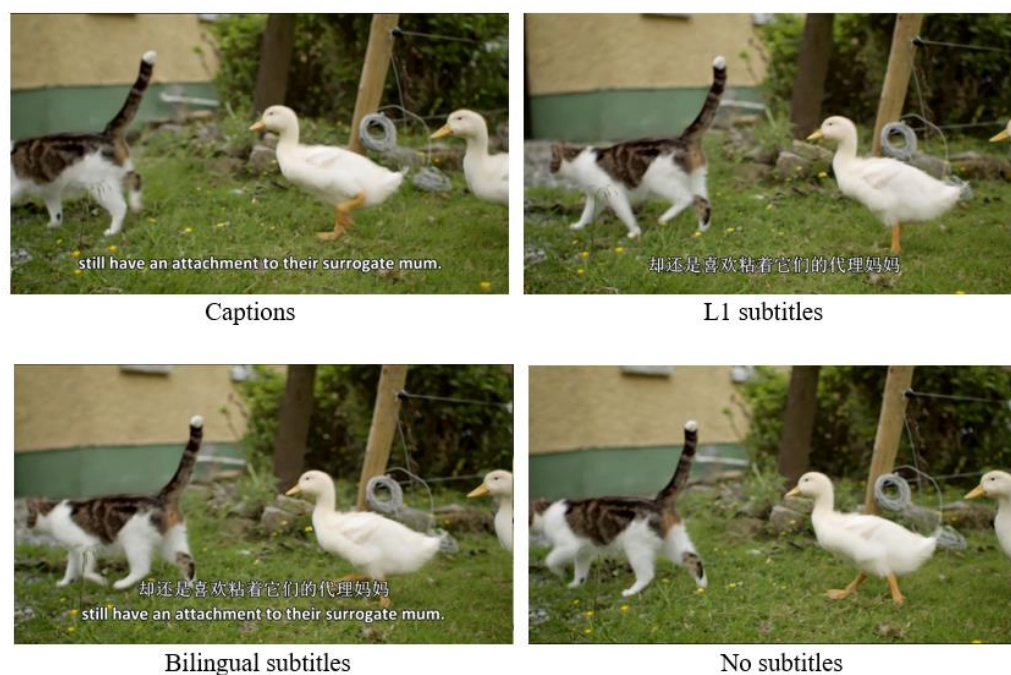


Figure 11. Screenshots of the Four Subtitling Types in the Present Research

### 3.4.3.3. Target Words and Distractor Items

In order to maintain the ecological validity of the study, the original content and audio of the video were used without any manipulations. Four steps were taken to choose the TWs which had to be unknown for participants. First, the script was initially inspected and a list of 52 potentially unknown words above the 5K level was created. Secondly, this list was consulted with seven experienced Chinese IELTS teachers who were asked to select the words that they thought would be unknown to the target participants with an overall IELTS score of around 7. This led to a revised list with 28 words. Thirdly, eight Chinese students with similar characteristics to the participants in the study were asked to indicate their knowledge of these words. Lastly, the words and

materials were piloted twice with 13 Chinese learners of English (see section 3.3), and words that were indicated as known by all participants were deleted. This resulted in a final list of 24 TWs, including 10 nouns, 10 verbs, and 4 adjectives that varied in frequency of occurrence. They were evenly distributed across each clip, with five TWs appearing in the first clip, six in the second, seven in the third, and six in the last clip. Details of the characteristics of the TWs are summarised in Table 10.

Table 10. Characteristics of the 24 Target Words (in alphabetical order)

| Target word | Word length | Part of speech | Frequency of occurrence | Clip No. | Time of occurrence                          | Presentation duration (in ms) | Absolute AOI size for TW (in pixels) | Relative AOI size for TW | Chinese translation | Number of Chinese characters | Absolute AOI size for translation (in pixels) | Relative AOI size for translation |
|-------------|-------------|----------------|-------------------------|----------|---------------------------------------------|-------------------------------|--------------------------------------|--------------------------|---------------------|------------------------------|-----------------------------------------------|-----------------------------------|
| barneys     | 7           | n.             | 1                       | 1        | 03'42''                                     | 2806                          | 24800                                | 1.20%                    | 争执                  | 2                            | 14700                                         | 0.71%                             |
| bizarre     | 7           | adj.           | 1                       | 4        | 20'30''                                     | 2249                          | 22000                                | 1.06%                    | 古怪的                 | 3                            | 20900                                         | 1.01%                             |
| buffering   | 9           | v.             | 1                       | 1        | 05'04''                                     | 1571                          | 28700                                | 1.38%                    | 缓解                  | 2                            | 14400                                         | 0.69%                             |
| bunting     | 7           | v.             | 4                       | 3        | 13'02'';<br>13'02'';<br>13'03'';<br>13'09'' | 4238                          | 27300                                | 1.32%                    | 撞                   | 1                            | 21400                                         | 1.03%                             |
| confiscated | 11          | v.             | 1                       | 1        | 02'13''                                     | 3040                          | 34100                                | 1.64%                    | 没收                  | 2                            | 14900                                         | 0.72%                             |
| cortisol    | 8           | n.             | 3                       | 3        | 08'06'';<br>08'13'';<br>08'15''             | 5986                          | 24200                                | 1.17%                    | 皮质醇                 | 3                            | 21400                                         | 1.03%                             |
| dinky       | 5           | adj.           | 1                       | 2        | 10'03''                                     | 1684                          | 17700                                | 0.85%                    | 小巧的                 | 3                            | 20300                                         | 0.98%                             |
| endearing   | 9           | adj.           | 1                       | 3        | 10'30''                                     | 1924                          | 30600                                | 1.48%                    | 可爱的                 | 3                            | 20900                                         | 1.01%                             |
| fawn        | 4           | n.             | 3                       | 3        | 12'04'';                                    | 7276                          | 17600                                | 0.85%                    | 小鹿                  | 2                            | 15000                                         | 0.72%                             |

| Target word | Word length | Part of speech | Frequency of occurrence | Clip No. | Time of occurrence                          | Presentation duration (in ms) | Absolute AOI size for TW (in pixels) | Relative AOI size for TW | Chinese translation | Number of Chinese characters | Absolute AOI size for translation (in pixels) | Relative AOI size for translation |
|-------------|-------------|----------------|-------------------------|----------|---------------------------------------------|-------------------------------|--------------------------------------|--------------------------|---------------------|------------------------------|-----------------------------------------------|-----------------------------------|
|             |             |                |                         |          | 15'39''                                     |                               |                                      |                          |                     |                              |                                               |                                   |
|             |             |                |                         |          | 15'48''                                     |                               |                                      |                          |                     |                              |                                               |                                   |
| foal        | 4           | n.             | 1                       | 2        | 09'04''                                     | 3796                          | 13500                                | 0.65%                    | 小马驹                 | 3                            | 21100                                         | 1.02%                             |
| foraging    | 8           | v.             | 1                       | 4        | 20'58''                                     | 2453                          | 25600                                | 1.23%                    | 觅食                  | 2                            | 14200                                         | 0.68%                             |
| gland       | 5           | n.             | 1                       | 2        | 08'13''                                     | 1943                          | 18200                                | 0.88%                    | 腺体                  | 2                            | 14300                                         | 0.69%                             |
| hump        | 4           | v.             | 1                       | 3        | 13'16''                                     | 3596                          | 18000                                | 0.87%                    | 弓起                  | 2                            | 14200                                         | 0.68%                             |
| midwife     | 7           | n.             | 1                       | 4        | 21'20''                                     | 925                           | 25100                                | 1.21%                    | 助产士                 | 3                            | 22400                                         | 1.08%                             |
| nuzzle      | 6           | v.             | 1                       | 3        | 16'24''                                     | 2031                          | 20400                                | 0.98%                    | 用鼻子蹭                | 4                            | 34900                                         | 1.68%                             |
| poaching    | 8           | v.             | 2                       | 2        | 06'51''                                     | 4315                          | 28300                                | 1.36%                    | 偷猎                  | 2                            | 13400                                         | 0.65%                             |
| purring     | 7           | v.             | 1                       | 4        | 19'22''                                     | 1064                          | 23800                                | 1.15%                    | 发出咕噜声               | 5                            | 34600                                         | 1.67%                             |
| sanctuary   | 9           | n.             | 4                       | 1        | 00'31'';<br>02'02'';<br>02'16'';<br>04'15'' | 8671                          | 30025                                | 1.45%                    | 保护区                 | 3                            | 21700                                         | 1.05%                             |

| Target word | Word length | Part of speech | Frequency of occurrence | Clip No. | Time of occurrence | Presentation duration (in ms) | Absolute AOI size for TW (in pixels) | Relative AOI size for TW | Chinese translation | Number of Chinese characters | Absolute AOI size for translation (in pixels) | Relative AOI size for translation |
|-------------|-------------|----------------|-------------------------|----------|--------------------|-------------------------------|--------------------------------------|--------------------------|---------------------|------------------------------|-----------------------------------------------|-----------------------------------|
| sedated     | 7           | v.             | 1                       | 1        | 03'46''            | 3249                          | 24900                                | 1.20%                    | 被注射镇静剂              | 6                            | 40100                                         | 1.93%                             |
| surrogate   | 9           | n.             | 1                       | 4        | 15'46''            | 3982                          | 29300                                | 1.41%                    | 代理                  | 2                            | 14600                                         | 0.70%                             |
| traumatised | 11          | adj.           | 1                       | 2        | 06'35''            | 4882                          | 36600                                | 1.77%                    | 受创伤的                | 4                            | 28300                                         | 1.36%                             |
| twirls      | 6           | n.             | 1                       | 3        | 15'22''            | 3268                          | 17500                                | 0.84%                    | 缠绕                  | 2                            | 12000                                         | 0.58%                             |
| ulcers      | 6           | n.             | 1                       | 2        | 08'20''            | 1875                          | 18700                                | 0.90%                    | 溃疡                  | 2                            | 14900                                         | 0.72%                             |
| waddled     | 7           | v.             | 1                       | 4        | 19'00''            | 2243                          | 27200                                | 1.31%                    | 摇摇摆摆地走              | 6                            | 35700                                         | 1.72%                             |

*Note.* Averaged presentation time across all exposures was calculated for the target words that appeared more than once in the video

In order to diminish the potential test effects from the pretests, 33 distractors were included in pre- and post- vocabulary tests. The distractors included 10 lower frequency words (ranked above 5K) and 23 higher frequency words (ranked within 3K) from the same documentary series to make the test less challenging and reduce guessing. The distractors shared the same semantic domain and part of speech as the TWs to reduce their salience, as shown in Table 11. Six out of the 10 lower frequency distractors which were included in the tests but did not appear in the video were used to control for potential test effects in the analysis (see section 4.2.1 for results).

Table 11. 33 Distractors Used in the Vocabulary Tests (in alphabetical order)

| Distractors | Part of speech | Chinese translation |
|-------------|----------------|---------------------|
| affection   | n.             | 喜爱                  |
| appear      | v.             | 出现                  |
| attacks     | v.             | 攻击                  |
| boggling    | v.             | （使）犹豫不决             |
| bonkers     | adj.           | 愚蠢的                 |
| calf        | n.             | 小牛                  |
| captive     | adj.           | 豢养的                 |
| channel     | n.             | 管道                  |
| combining   | v.             | （使）结合               |
| confident   | adj.           | 自信的                 |
| crisis      | n.             | 危机                  |
| discovered  | v.             | 发现                  |
| enamoured   | adj.           | 迷恋的                 |
| fecund      | adj.           | 多产的                 |
| frisson     | n.             | 兴奋感                 |
| fulfil      | v.             | 实现                  |
| individual  | n.             | 个体                  |

| Distractors | Part of speech | Chinese translation |
|-------------|----------------|---------------------|
| interfere   | v.             | 干涉                  |
| knowledge   | n.             | 知识                  |
| loyalty     | n.             | 忠诚                  |
| marvellous  | adj.           | 绝妙的                 |
| mature      | adj.           | 成熟的                 |
| obvious     | adj.           | 明显的                 |
| realise     | v.             | 意识到                 |
| rhino       | n.             | 犀牛                  |
| ridiculous  | adj.           | 荒唐的                 |
| same        | adj.           | 相同的                 |
| separate    | v.             | 分开                  |
| serious     | adj.           | 严重的                 |
| suckle      | v.             | 吃奶                  |
| uncertain   | adj.           | 不确定的                |
| understand  | v.             | 理解                  |
| wildebeest  | n.             | 角马                  |

#### **3.4.4. Instruments**

##### **3.4.4.1. Vocabulary Size Test**

Since learners' vocabulary size can affect their incidental vocabulary learning gains (Puimège & Peters, 2019), participants' vocabulary size was measured via the online Mandarin Chinese version of Nation's and Beglar's (2007) Vocabulary Size Test (<http://my.vocabularysize.com/>). It is a "discrete, selective, relatively context-independent" vocabulary test measuring learners' written receptive vocabulary knowledge for reading, presented in multiple-choice format (Nation, 2012, p. 1). It provides a rough estimate of a learners' vocabulary size by sampling 10 items from each

of the first 14 frequency bands of 1,000 word families, resulting in 140 test items in total (Peters, 2019). One of the common and accepted ways to explore the validity of the vocabulary size is to test the correlation between Vocabulary Size Test scores and scores in a proficiency test (Schmitt et al., 2001). Thus, Pearson correlation was conducted between participants' overall IELTS scores and their vocabulary size scores. The results revealed a significant and large correlation ( $r = .61, p = .002$ ), providing convergent evidence for the test validity.

#### **3.4.4.2. Vocabulary Tests**

As suggested by Nation (2001) and Nation and Webb (2011), it is important to assess vocabulary knowledge using various tests examining different aspects of knowing a word. In addition, form and meaning recognition tests are believed to be useful to probe the initial stages of vocabulary acquisition (Schmitt, 2010). Therefore, in line with previous studies (e.g., Mohamed, 2018; Montero Perez et al., 2015, 2018), form recognition, meaning recall, and meaning recognition tests in pencil-and-paper format were used in both pre- and posttests with randomised item order. An occurrence decision test was added in posttests as a way to control for potential test effects. The 57 single words (24 TWs and 33 distractors) were included in each test and presented in the identical form as shown in the video.

##### ***3.4.4.2.1. Form Recognition & Meaning Recall Test***

In order to reduce number of exposures to the TWs and minimise potential test effects (see also Peters, 2019; Peters & Webb, 2018), the form recognition test was combined with the meaning recall test. Following Montero Perez (2019, 2020a) and

Peters (2019), in order not to favour any group, each word was presented in both spoken and written form using PowerPoint on an iPad, with the audio recording of each word played twice. The audio files were recorded by a female L1 English speaker from London in a quiet room. After the presentation of each word, participants were asked to indicate whether they had seen/heard the word before by ticking Yes or No in the answer sheet. They were then asked to provide a translation/synonym/explanation for the words they had ticked (see Table 12 for a sample item).

Table 12. Example of the Form Recognition and Meaning Recall Pretest Answer Sheet

| Items | 你见/听过这个词吗? Have you ever seen/heard the word before? |    | 我见过或听过这个词, 它的意思是 (请写出中文翻译/解释) I have seen or heard this word before and it means...<br>(translation, English synonym, definition) |
|-------|------------------------------------------------------|----|-----------------------------------------------------------------------------------------------------------------------------------|
| 1.    | Yes                                                  | No |                                                                                                                                   |
| 2.    | Yes                                                  | No |                                                                                                                                   |

#### 3.4.4.2.2. *Meaning Recognition Test*

The meaning recognition test was a written multiple-choice test, with each item accompanied by four options in Chinese: the key, three distractors, and an “I don’t know” option to minimise guessing (Peters & Webb, 2018). All the distractors shared same parts of speech with the test items and were relevant to the theme of the video. Following the procedures suggested and used by Nation and Webb (2011) and Rodgers (2013), the first distractor option was the translation of another TW. The second and third distractors were randomly chosen from a distractor pool, which consisted of the Chinese translations of 34 synonyms of the 24 TWs together with the translations of ten low frequency non-target distractors.

Following Rodgers (2013), several steps were taken in creating this distractor pool. First, the Thesaurus website (<https://www.thesaurus.com/>) was used to find the synonyms for each TW. All synonyms for each TW were then checked for their lexical frequencies in the BNC corpus. The synonym that was from the same or a lower frequency band in the BNC word list as its corresponding TW was chosen and translated into Chinese using online Cambridge Dictionary (<https://dictionary.cambridge.org/zhs/>). Then, to select the suitable Chinese translation to be included in the distractor pool, for each synonym, the translation that did not correspond to the actual translation of the TW was chosen. For example, the synonym from the same frequency band of the TW *bizarre* was *peculiar*, which can be translated to 奇怪的(*odd*), 古怪的(*weird*) or 特有的(*special*) in Chinese. In this particular case, the translation 特有的(*special*) was chosen in the distractor pool. Since the other two translations were same as the Chinese translation of the TW *bizarre* (古怪的), thus they were not qualified to serve as distractors. If all the Chinese translations of the synonyms were same as the translations of the TW, one of the synonyms was used as a key word to find further synonyms. For example, the TW *barneys* had a limited number of synonyms, and all the Chinese translations of its synonyms were the same as the Chinese translation of *barneys*. Thus, *argument*, one of the synonyms of *barneys*, was used as a key word to find synonyms for *barneys*. This could guarantee that all distractors shared similar word domain and similar frequency level with their corresponding TWs. The synonym distractor of each TW was not included as an option for its corresponding TW but served as an option for other TWs.

Consequently, each TW had four options: a key, a TW distractor (the translation of another TW), two distractors from the distractor pool (translations of the synonyms of other TWs), and an “I don’t know” option (see Appendix S5 for the complete test). All

translations of the TWs and non-target distractors would therefore appear twice in multiple-choice test options to reduce the salience of the correct answer. The distractors were numbered and the *Random Sequence Generator* function in a randomisation programme ([www.random.org](http://www.random.org)) was used to select distractors for each test item. The *Integer Generator* function was used to choose the key's position. *Sequence Generator* function was used to randomise the order of the items in vocabulary tests. These three vocabulary tests were used both in pre- and posttests, but the order of the test items was randomised.

#### **3.4.4.2.3. Occurrence Decision Test**

In the posttests, apart from the three tests mentioned above, an occurrence decision test was added after the form recognition test. In the posttests, if participants indicated on the form recognition test that they had seen/heard the word before, they then needed to indicate whether the word appeared in the video or not (see Table 13 for a sample item). It was presented jointly with the form recognition test to tap into learners' knowledge of word form. Since participants' form recognition knowledge could have been obtained from the pretests or outside the viewing, the addition of this test allowed me to account for cases in which participants thought they had seen the TW before but did not remember having seen it in the video. This test only included TWs and lower frequency distractors (above 11K frequency level). High frequency distractors were not included in this test due to the difficulty for participants to decide whether the familiar words, which they might pay less attention to, appeared in the video or not.

Table 13. Example of the Form Recognition, Occurrence Decision and Meaning Recall Posttest Answer Sheet

| Items | 你见/听过这个词吗？（包括在视频中）<br>Have you ever seen/heard the word before? (including in the video) |    | 它出现在了视频中<br>It has appeared in the video | 我见过或听过这个词，它的意思是（请写出中文翻译/解释）<br>I have seen or heard this word before and it means...<br>(translation, English synonym, definition) |
|-------|------------------------------------------------------------------------------------------|----|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
|       | Yes                                                                                      | No |                                          |                                                                                                                                    |
|       | Yes                                                                                      | No |                                          |                                                                                                                                    |

### 3.4.4.3. Comprehension Test

A comprehension test was conducted to investigate whether using different types of subtitles could facilitate viewers' comprehension of the video. Thus, information that was presented through the non-verbal input but not presented in the verbal input was not assessed. In line with Montero Perez et al. (2014) and Rodgers (2013), the development of the comprehension test was based on Buck's (2001) "competency-based" default listening construct (p. 114), which includes the abilities to process the general information, understand the detailed content, and make inferences. The inferencing ability was not tested in this study due to the restriction of the viewing material, in which the type of certain and factual information provided left no room to infer information (see also Montero Perez et al., 2014). Consequently, global questions, which dealt with the general understanding of the content, and local questions, which targeted the detailed content, were included in the test.

Concerning the format of the comprehension test. After the pilot studies, a four-option multiple-choice test was considered the most appropriate for the present study for the following reasons: 1) it was one of the most common test formats to assess

listening comprehension in many major international tests (Brindley, 1998), and in most empirical studies investigating comprehension from viewing (e.g., Rodgers, 2013; Winke et al., 2010); 2) compared to open-ended format, multiple-choice was considered easier for L2 listening test (In'Nami & Koizumi, 2009), which might reduce participants' stress during the experiment; 3) it required a minimal amount of time to complete while covering sufficient information and leading to high internal consistency reliability (Thompson, 1995); 4) the scoring was easy, efficient, and unambiguous (Brindley, 1998); 5) compared to true/false format, it reduced the possibility of guessing by including four options.

In order to design the multiple-choice items, the transcript of the video clip was firstly parsed into *idea units* (Pellicer-Sánchez et al., 2020; Rodgers, 2013), which were defined by (Rodgers, 2013, p. 33) as “distinct events, actions, or dialogue spoken in the course of the program” (p. 33). Then the idea units which only described the images, showed common sense, or required memory on trivial information (for example, remembering the numbers, dates, names, and locations etc.) were deleted. The remaining idea units were then used to create multiple-choice questions. For each test item, a question was written to serve as the stem, the idea unit was rewritten as the key. Reasonable distractors were written for each stem. Care was taken to ensure: 1) distractors matched the key in length and grammatical structure; 2) distractors were plausible and reasonable but could not be considered correct based on the video content. Three most plausible distractors were finally chosen for each stem. All stems and options were written in Chinese, to ensure that test scores were not influenced by other intervening variables (Buck, 2001). The *Integer Generator* function ([www.random.org](http://www.random.org)) was used to choose the key's position. Questions were arranged in the same order in

which they occurred in the video. Forty text-based multiple-choice comprehension items were first produced.

In order to ensure that the test cannot be answered correctly without understanding the video (Buck, 2001), six Chinese learners of English were invited to complete the test without watching the video. Items that had been answered correctly by all were discarded. Modifications were made based on test results and feedback received. Then, the modified 34 multiple-choice items were tested online for their reliability prior to the experiment with a group of learners of similar characteristics ( $N = 38$ ). The bilingual subtitle version of the video was used for this pilot test. The Cronbach's Alpha coefficient was .67. Based on the online pilot for the comprehension test, 30 out of the 34 questions were kept unchanged and four were further modified. The final version of the comprehension test included a total of 34 items (see Appendix S6).

#### **3.4.4.4. 3K Vocabulary Levels Test**

As mentioned earlier, knowledge of the first 3K words provided a lexical coverage of 95.47% in the selected viewing material. Therefore, it was important to ensure that participants had knowledge of the 3K words to understand the content of the selected video. The participants' 3K vocabulary knowledge was examined using the Vocabulary Levels Test (VLT) at the 3000-word level. The VLT was originally developed by Nation (1990) as a diagnostic vocabulary test to provide an estimate learners' vocabulary size. Schmitt et al. (2001) provided a revised and validated version of the original test. This revised and expanded version was used in the present study. The 3K level of the test was a matching test which consisted of ten sections. Each section comprised six 3K words and three short English definitions. Participants were required to match each of the three definitions with one of the six words (see Appendix S7). A

score of 24 out of 30 was used as the threshold for demonstrating mastery of the 3K words, as suggested by Schmitt through a personal communication with Xing and Fulcher (2007) in 2003.

#### **3.4.4.5. Background Questionnaire**

A background questionnaire was included at the end of the experiment using a Chinese online questionnaire builder *WJX.cn* (<https://www.wjx.cn/>). The online questionnaire was administered to obtain information about participants' language background information such as age, IELTS scores, time living in the UK, frequency of their use of English, and their viewing habits (e.g., habits of watching English videos, preferred types of subtitles, etc.; see Appendix S8).

#### **3.4.5. Experiment Design and Apparatus**

The eye-tracking experiment was designed with Experiment Builder (SR Research, 2011), a graphical programming environment for creating computer-based psychology and neuroscience experiments. The selected video clip was chosen as the stimulus and uploaded to the software. The soundtrack was uploaded separately from the video frames and were triggered simultaneously with the play of the video. To design the experiment, an introduction page was first displayed, followed by a nine-point calibration. A practice session with a one-minute video on a related topic to the experimental video was performed before the experiment to help participants adapt to the experimental environment. The nine-point calibration was conducted again before the experimental video started. A thank you page was presented after the end of the video, indicating the end of the viewing session. Four versions of the experiment were

created with the same videos for four subtitling conditions (i.e., captions, L1, bilingual, and no subtitles).

Participants' eye movements were recorded with EyeLink 1000 plus eye-tracker (SR Research, 2016), in desk-mounted mode. It uses the infrared light to illuminate the eye and records participants' pupil and corneal reflection to track the participants' eye movements. The system has a data sampling rate of 1,000 Hz (taking 1,000 snapshots of the eye per second). It parses the participant's eye movements into saccades, fixations, and blinks automatically. Recording was monocular (right eye). The eye-tracker was placed below the monitor, installed 60 cm in front of the participants. The system was interfaced with a display DELL computer and a 19-inch DELL monitor with a 1920 × 1080 screen resolution. An adjustable head and chin rest was installed 60 cm in front of the monitor to minimize participants' head movements. A laptop host PC was connected to the display PC and used by the researcher to perform calibration and monitor real-time eye movements during viewing sessions.

#### ***3.4.6. Stimulated Recall***

The aim of the stimulated recall was to further explore participants' engagement with each unknown TW during the subtitled viewing. Thus, stimulated recall was only administered with participants in the captions ( $n = 27$ ), L1 subtitles ( $n = 25$ ), and bilingual subtitles ( $n = 30$ ) groups individually. As explained in section 2.5.5, two aspects of cognitive engagement were examined in stimulated recall interviews: awareness and vocabulary processing strategies. Before the stimulated recall interview, oral instructions adapted from Gass and Mackey (2017) were given (see Appendix S9). Participants were encouraged to ask questions about the procedures. For each participant, recordings of their eye movements during the presentation of each

occurrence of the 24 TWs (35 occurrences in total) were played to elicit participants' memory recall. The 35 recorded stimuli were played at a 50% speed using the EyeLink DataViewer software (SR Research, 2018).

The researcher first pronounced each English TW, and then played the corresponding stimulated recall stimulus. Participants were then asked if they had noticed the word at that time, and if they had, what they had been thinking about the TW at that time. Participants were asked to inform the researcher if the TW had already been familiar to them before the viewing session. The procedure was repeated for the 35 occurrences of the 24 TWs, and participants were asked to verbalise their thoughts on each occurrence of the TWs. Participants could ask to replay the stimulus to support their recall. They were asked to report the thoughts that they had during viewing rather than their thoughts at the time when the stimulated recall was conducted. Following Gass and Mackey (2017), no concrete responses were given to the participants' responses, except repeating their responses, or providing "back-channelling cues or nonresponses" such as "Oh, mhm, great, good, I see, uh-huh, ok" (p. 55). If there was an indication that the participant was talking about his/her current thoughts of the TW, the researcher brought the participant back on track by asking, "Is this what you were thinking at that time during viewing or your current thoughts?". No further questions were asked if participants were unable to recall their thoughts relating to the TW or they did not notice the TW. All stimulated recall interviews were held in participants' L1 Mandarin Chinese.

#### **3.4.7. Data Collection Procedure**

The overall data collection for the main study lasted approximately five months in two sessions with a time gap of two to three weeks in between. The first session was

conducted individually or in pairs at the UCL study room or in the eye-tracking lab. The second session was administered individually in the eye-tracking lab. The length and procedure for each session are presented in Table 14.

Table 14. Data Collection Procedures and Time Duration of the Present Main Study

|                             | Contents                                                                                             | Approximate Time |
|-----------------------------|------------------------------------------------------------------------------------------------------|------------------|
| Session one<br>(45 mins)    | Information sheet & Consent form                                                                     | 3 mins           |
|                             | Vocabulary pretests<br>(Form recognition, meaning recall, meaning recognition)                       | 26 mins          |
|                             | Online Vocabulary Size Test                                                                          | 16 mins          |
| 2–3 Weeks Later             |                                                                                                      |                  |
| Session two<br>(70–90 mins) | Watching video                                                                                       | 26 mins          |
|                             | Comprehension test                                                                                   | 8 mins           |
|                             | Vocabulary posttests<br>(Form recognition, occurrence decision, meaning recall, meaning recognition) | 27 mins          |
|                             | Stimulated recall                                                                                    | 0–20 mins        |
|                             | 3K vocabulary levels test                                                                            | 5 mins           |
|                             | Online background questionnaire                                                                      | 5 mins           |

In the first session, a general introduction to the research was first provided before obtaining the consent to participate. Then, three vocabulary pretests were conducted in paper-and-pencil format, followed by an online Vocabulary Size Test (Nation & Beglar, 2007) (<http://my.vocabularysize.com/>). No time limit was set for any tests. For participants who accomplished this session in pairs, they were not allowed to

communicate during the whole session. The first session lasted on average 45 minutes. It should be noted that in the first session, as noted in the Ethics application (Appendix S2), participants were not informed of specific goals of the study but were told that the purpose of the tests was to assess their English proficiencies, and the goal of this study was to explore the use of video subtitles for comprehension. As noted in the information sheet, two to three weeks after the first session, each participant was contacted by the research individually to schedule a time for the second session.

In the second session, the participants were invited to the eye-tracking lab and randomly assigned to one of the four conditions: captions ( $n = 27$ ), L1 subtitles ( $n = 25$ ), bilingual subtitles ( $n = 30$ ), and no subtitles ( $n = 28$ ). Participants were told to complete the viewing session while their eye movements were recorded. They were seated in a comfortable chair about 60 cm away from the computer screen attached to the EyeLink 1000 plus eye-tracking camera. Participants were asked to wear the headphones during the viewing session and put their chin on a chin rest, with the forehead leaning against a forehead rest. After the instructions, a short eye-tracking calibration was first completed, where participants were asked to gaze on nine points presented at random on the computer screen. Calibration was repeated until getting an adequate level and a good validation. Then, a one-minute practice video clip was played as the warm-up practice, after which, participants were allowed to adjust their sitting position and the volume of the headphones, or ask questions if needed. After a second nine-point calibration, the 23-minute viewing stimulus was played.

After the viewing session, the participants were asked to complete the pencil-and-paper, multiple-choice comprehension questions and four vocabulary posttests. For the participants in three subtitled groups (i.e., captions, L1, and bilingual subtitles;  $N = 82$ ), stimulated recall interviews were individually held immediately after

vocabulary posttests. After instructions, each participant was asked to describe their thoughts while viewing, prompted by the replay of their eye-movement recordings during the presentations of each TW. All the stimulated recall was audio recorded with a portable recorder. A 3K vocabulary levels test and an online background questionnaire were conducted at the last stage. At the end of the experiment, participants were debriefed the real purpose of this study. They were then given a £10 Amazon Voucher for their participation. The second session lasted approximately from 70 minutes to 90 minutes.

Chapter 3 has reported the research design adopted in the present investigation. The analyses and results of the quantitative and qualitative data are reported in the next two chapters separately.

## **Chapter 4. QUAN Analyses and Results**

This chapter reports procedures followed to analyse the quantitative data, followed by a report of results and an interim discussion. The quantitative analyses reported in this chapter have the aim of exploring the effectiveness of bilingual subtitles for the acquisition of vocabulary and comprehension, as well as examining learners' processing of the on-screen text and TWs and their relationship with vocabulary gains. The following research questions are addressed by the analysis presented in this chapter:

RQ1: To what extent does the use of bilingual subtitles increase learners' vocabulary knowledge (as measured by form recognition, meaning recall, and meaning recognition tests), compared to captions, L1, and no subtitles?

RQ2: To what extent does the use of bilingual subtitles facilitate learners' viewing comprehension, compared to captions, L1, and no subtitles?

RQ3: How do learners allocate their attention to different areas (i.e., images, subtitling areas, unknown TWs and/or their L1 translations) during bilingual subtitled viewing, compared to captions, L1, and no subtitles, as revealed by eye-tracking data?

RQ4: Do learners' online processing of unknown TWs and their corresponding L1 translations predict their vocabulary gains in different subtitling conditions (i.e., captions, bilingual, and L1 subtitles)?

In this chapter, I first explain scoring procedures for offline tests which include vocabulary tests, the 3K vocabulary levels test, and the comprehension test. Then, the analyses of eye-tracking data at three levels are illustrated, after which, a description of quantitative statistical analyses used in the present study to answer each research

question is presented. Results from the analyses and an interim discussion are then presented.

## **4.1. QUAN Analyses**

### ***4.1.1. Scoring of Offline Tests***

The form recognition, meaning recall, and meaning recognition test each included 57 items (24 TWs and 33 distractors), and the occurrence decision included 34 items (24 TWs and 10 low-frequency distractors). Only responses to TWs were scored dichotomously with 0 for an incorrect response and 1 for a correct response. The comprehension test included 34 questions and the 3K vocabulary levels test included 30 items. One point was given for each correct response. For vocabulary tests, only TWs were taken into account for data analyses, resulting in a maximum of 24 points in each of the vocabulary test, 34 points in the comprehension test, and 30 in the 3K levels test.

For the scoring of the meaning recall test, only answers that clearly demonstrated the knowledge of the words' meaning were given 1 point. No half score for partial knowledge was given (see Appendix S10 for detailed scoring scheme). For example, for the TW *waddled*, answers that included the feature of walking with short steps and moving body from one side to the other were considered correct. Those answers which only included the feature of walking were considered incorrect and no scores were given. Pre- and post- meaning recall tests were scored by the researcher and a second rater. The second rater, who was doing a PhD degree in Applied Linguistics, scored 20% of both tests. A detailed scoring scheme was provided to the second rater. Inter-coder reliability was determined using Cohen's kappa. Interrater reliability for both the pretest and posttest was very high: Cohen's kappa ( $\kappa$ ) = .98,  $p$  = .002 for the

pretest, and Cohen's kappa ( $\kappa$ ) = .99,  $p$  = .001 for the posttest, indicating a high level of agreement between the two raters beyond chance.

#### ***4.1.2. Eye-Movement Data Analysis***

The eye-movement data were first inspected using DataViewer software (SR Research, 2018). Data from six participants were removed from the analysis of online data due to poor data quality, resulting in 100 participants included in the analysis of online data. Poor data contained sizable track loss or problematic drift as demonstrated in the temporal graph and spatial overlay view by plotting the raw data in EyeLink DataViewer software (SR Research, 2018). Then the eye-movement data were cleaned following suggestions by Conklin et al. (2018) and Godfroid (2020a). Fixations shorter than 50ms were merged if they were within 1° of visual angle (0.34% of the data), and those that were still below 50ms were removed from the dataset (8.35% of the data).

In order to provide a comprehensive examination of learners' processing and viewing behaviour across subtitling groups, the analysis of the eye-movement data was performed at three levels: the overall subtitling area, L1/L2 subtitle lines, and individual TWs. Different AOIs were first created for each level:

Level 1: The overall subtitling area in four viewing conditions:



Figure 12. Illustrations of Level 1 Areas of Interest for Eye-Movement Data Analysis in Four Groups

The aim of Level 1 analysis was to explore potential differences in processing the overall subtitling area across four subtitling conditions. In order to ensure the comparability between groups, the bilingual subtitles group (with the largest subtitling area size) was set as the baseline group in choosing the size of the AOI. The overall subtitling area (the green area shown in Figure 12) covered  $1920 \times 270$  pixels, including the whole width of the screen and the height between the top of on-screen text and the bottom of the screen. The rest of the screen ( $1920 \times 810$  pixels) was taken as the image AOI. The same AOI was also created for the other three groups.

Five-hundred-and-thirty-five interest periods (IPs) were generated manually according to the presentation time of on-screen text. Following Montero Perez (2019), only the eye-movement data that occurred within the AOIs and during the 535 IPs were included in the analysis.

The relative attention being paid to the subtitling area was examined using four eye-tracking measures (see Table 15). Instead of using the total reading time and fixation count, the proportions of the processing time and fixation counts on subtitling areas were analysed for Level 1, as this is the common procedure when examining eye movements to multimodal materials (Pellicer-Sánchez et al., 2020). Since I interested in participants' relative attention distribution on the subtitling area, the use of proportion could well capture participants' processing of different areas and lead to more comparable results to previous studies using different stimuli. Run count and skip rate were also calculated since they have been used in previous viewing studies (e.g., d'Ydewalle & De Bruycker, 2007; Muñoz, 2017) to inform us about the frequency of an AOI being referred to or skipped, which could help to reveal learners' different frequency of using subtitling and image areas during subtitled viewing.

Table 15. The Definitions of Four Eye-Tracking Measures at Level 1 and Level 2 Analyses

| Measurement          | Definition                                                                                           |
|----------------------|------------------------------------------------------------------------------------------------------|
| Total reading time % | Percentage of all summed time spent on an AOI within the defined IP                                  |
| Fixation %           | Percentage of the total number of fixations on an AOI within the defined IP                          |
| Run count            | Total number of times the AOI is entered and left within the defined IP                              |
| Skip rate            | An AOI is considered skipped (i.e., SKIP=1) if no fixation occurred in the AOI within the defined IP |

Level 2: L1 and L2 line areas in three groups with on-screen text:



Figure 13. Illustrations of Level 2 Areas of Interest for Eye-Movement Data Analysis in Three Groups with On-Screen Text

The aim of the Level 2 analysis was to investigate the processing of different lines in bilingual subtitles and compare them with the other two monolingual subtitling conditions. For Level 2, two AOIs with a same size of  $1920 \times 100$  pixels were created for L1 and L2 lines separately and were applied across three subtitling conditions. Two AOIs were used for the bilingual subtitles group, with one covering the L1 line and the other covering the L2 line (as the blue and yellow areas shown in Figure 13 respectively). One AOI with the same size was used for the captions and L1 subtitles groups (as shown in Figure 13). Following the analysis of Level 1, the eye-movement data were only calculated during the activation of the 535 IPs, and same eye-tracking measures were used for Level 2.

Level 3: Target word and L1 equivalent areas in three groups with on-screen text:



Figure 14. Illustrations of Level 3 Areas of Interest for Eye-Movement Data Analysis in Three Groups with On-Screen Text

The aim of Level 3 analysis was to compare learners' processing of unknown TWs and/or their corresponding L1 translations in the three subtitling conditions. As shown in Figure 14, for each TW and its corresponding L1 Chinese translation, AOIs were created. These AOIs were only activated during the time that TWs were presented on the screen. The average time duration of each TW was 3,294 ms ( $SD = 1911$ , 95% CI = [2488, 4101]). The height of AOIs was fixed at 100 pixels (the same as at Level 2). The width of an AOI varied according to the length of TWs. The size of an AOI for a given TW was kept the same across conditions. The average size of the AOI of an English TW was 24,338.54 pixels ( $SD = 5826.55$ , 95% CI = [21878.21, 26798.88]), and 20,845.83 pixels ( $SD = 8144.99$ , 95% CI = [17406.51, 24285.16]) for a Chinese translation of the TW (see Table 10 for AOI size details). In line with previous eye-tracking research (e.g., Mohamed, 2018; Montero Perez et al., 2015; Williams & Morris, 2004), seven eye-tracking measures were used at Level 3 to capture participants' eye movements at both early and late stages of word processing (see Table

16). Since learners' familiarity of a word would affect their processing behaviour (e.g., Godfroid et al., 2013; Pellicer-Sánchez, 2016; Williams & Morris, 2004), only the unknown words for each participant (scored as 0 in the form recognition pretest) were included in the analysis. Form recognition was chosen as the criteria for item selection since its scores are usually higher than recall scores (Laufer & Goldstein, 2004), which can help to control for minimum or partial knowledge before the experiment.

Table 16. The Definitions of Seven Eye-Tracking Measures at Level 3 Analysis

| Measurement              | Definition                                                                                                                                    |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| First-pass reading time  | The sum of the duration of all fixations on an AOI before exiting the AOI within the defined IP                                               |
| First fixation duration  | The duration of the first fixation event on an AOI within the defined IP                                                                      |
| Total reading time       | The sum of the duration across all fixations on an AOI within the defined IP                                                                  |
| Fixation count           | The total number of fixations on an AOI within the defined IP                                                                                 |
| Second-pass reading time | The sum of the duration across all fixations that are made within an AOI when the eyes visit the area for a second time within the defined IP |
| Second fixation duration | The duration of the second fixation on an AOI within the defined IP                                                                           |
| Skip rate                | An AOI is considered skipped (i.e., SKIP=1) if no fixation occurred in the AOI within the defined IP                                          |

#### **4.1.3. Quantitative Statistical Analyses**

IBM SPSS Statistics (Version 25) predictive analytics software was used to first inspect the data for potential outliers, and compute descriptive as well as correlational statistics for the data. The reliability of the tests and the inter-coder reliability of scoring

were also examined using the same software. To be specific, the interrelationships between the four vocabulary test scores were determined using pairwise point-biserial correlation tests, and Pearson's coefficient was used for different eye-tracking measures. The internal consistency reliability of the different tests and the inter-coder reliability were determined using Cronbach's alpha and Cohen's kappa, respectively. The significant threshold value for alpha was set at  $p < .05$  for this study.

The statistical software package R (v 3.6.1; R Development Core Team, 2019) was used to construct mixed-effects models to analyse the differences among groups in terms of vocabulary test scores and eye-movement data. The effects of the independent variables (i.e., eye-movement data) on the dependent variables (i.e., different vocabulary test scores) were also examined using mixed-effects models. Mixed-effects regression analyses have the advantages of accommodating nested data and including various fixed effects, covariates, and random effects in the analyses (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Cunnings, 2012). Different from means-based parametric statistical techniques, for example t-tests and ANOVA, mixed-effects models could take into account differences across subjects and items, enabling to generalize study findings to a larger learner population and different linguistic materials (Baayen et al., 2008; Linck & Cunnings, 2015). In the present study, participants' answer to each sampled TW and their eye-tracking data were collected nested within each participant. Therefore, by adding Participant and Item as random effects, the research findings can be better generalised. Moreover, mixed-effects models are also robust against violation of homoscedasticity and missing data (Cunnings, 2012).

Linear, logistic, and Poisson mixed-effects models were constructed for continuous, binary, and count dependent variables accordingly using the *lmer* or *glmer* function in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Regarding assumption

tests, the collinearity, normal distribution of residuals, and homoscedasticity (constant variance of residuals) assumptions were checked for all linear mixed-effects models using sjPlot package (v 2.8.4; Lüdtke, 2020), while glmmTMB package (v 1.0.1; Brooks et al., 2017) was used for generalized linear mixed models. VIF, tolerance, and average VIF were calculated as measures of collinearity assumption check for models using the car package (v 3.0-8; Fox & Weisberg, 2019) when necessary. Tukey post-hoc tests were ran using the multcomp package (v 1.4-13; Hothorn, Bretz, & Westfall, 2008) for pairwise comparisons. For linear mixed-effects models, LmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) was used to obtain *p*-values. Both marginal and conditional  $R^2$  values were calculated as measures of variance explained by fixed and random effects using the *r.squared GLMM* function in the MuMIn package (Barton, 2020). Cohen's *d* was used to calculate the effect size for linear regressions using the *cohensD* function in the lsr package (Navarro, 2015). *D* values of .40, .70 and 1.00 were considered small, medium and large effect sizes, respectively (Plonsky & Oswald, 2014). Outliers were identified using “model criticism” (Godfroid, 2020a, p. 267) method after fitting the best models using the *romr.fnc()* function in the LMERConvenienceFunctions package (Tremblay & Ransijn, 2020). Data points with absolute standardized residuals exceeding 2.5 *SD* ( $2.5 < |z|$ ) were treated as outliers. This method for dealing with outliers was recommended by Baayen and Milin (2010) since it did fewer manipulations of the data. Sensitivity analyses with and without the outliers were run to reveal potential differences. Outliers were removed from analyses when they changed the statistical significance of the fixed effects in models. For logistic mixed-effects models, Hosmer-Lemeshow test was run using the *logitgof* function in the generalhoslem package (Jay, 2019) to examine the goodness of fit. *Plot* function was used to plot the Pearson residual for generalized linear mixed effect models to detect

outliers. Odds ratio (OR) was used as an alternative for logistic regression to measure the effect size (Field, Miles, & Field, 2012). An OR larger than one indicates positive relationship and an OR less than one indicates a negative relationship. ORs greater than 3 or less than 0.33 are considered to be strong (Haddock, Rindskopf, & Shadish, 1998). Details of statistical analyses for each RQ are presented in the following section.

#### **4.1.4. Main Statistical Analyses**

To avoid repeating information, the common steps taken for all models are first presented before moving to the analyses for each research question.

The modelling started by constructing a null model only with the random intercepts of Participant and Item (e.g., `FR <- glmer(posttest ~ (1|Participant) + (1|Item), data = FRdata)`). The best models were constructed using forward selection method and reported based on likelihood ratio tests with the *anova()* function and on Akaike information criterion (AIC) scores. Maximal random-effects structure was adopted since it has been recommended for confirmatory hypothesis testing research to strengthen the generalisation of findings (Linck & Cunnings, 2015). Each of the fixed effects or covariates was entered into the null model step-wisely (e.g., `FR1 <- glmer(posttest ~ Group + (1|Participant) + (1|Item), data = FRdata)`), and they were kept in the model only when the inclusion of fixed effects or covariates significantly improved the model fit according to AIC scores. Participant-level variable (i.e., vocabulary size) and item-level variables (i.e., word class, frequency of occurrence, word length for offline measures/AOI size of each TW for online measures, and presentation time of each TW for online measures) were also entered into the regression models as covariates because previous studies have shown that these factors may influence vocabulary learning through viewing (e.g., Peters, 2019; Peters & Webb, 2018; Puimège & Peters, 2019). It

should be stressed that the aim of the study was not to investigate the roles of these word-related characteristics, but rather to take these parameters into account in analyses. All continuous variables were log-transformed or rescaled before being added to models to address the skewness problem (Godfroid, 2020a). The interactions between fixed effects and covariates were also checked. Participant and Item were always added as random intercepts. When analysing mean differences among groups, Group was also included as the random slope by Item if it improved the model fit. If the maximal random structure models failed to converge, the optimizing function using the control argument, *control = glmerControl(optimizer = "bobyqa")* was added to refit the model, as suggested by Linck and Cunnings (2015). Random effect parameters resulting in the least variance were removed one by one when models failed to converge, until convergence was achieved. Random slope and covariates were only kept in models only when they improved the goodness of model fit.

RQ1 aimed to compare vocabulary gains in the bilingual subtitles group to the other three subtitling conditions. After checking the comparability of four groups at the outset of the study, participants' pre- and posttests scores and absolute learning gains of TWs at item level (see also Peters & Webb, 2018) were calculated by group. Due to the binary nature of vocabulary scores as dependent variable, logistic mixed-effects models were fitted with Group as fixed effect. Apart from the aforementioned covariates, participants' pretest scores for each item were also added as a covariate in the analysis to control for their prior knowledge of TWs. The potential test effects were also examined for form recognition, meaning recall, and meaning recognition tests, by checking whether Item Type (i.e., TWs or distractors) could predict participants' posttest scores of the 24 TWs and the 6 low-frequency distractors (included in tests but did not appear in the viewing material), while controlling for their pretest scores.

RQ2 investigated participants' comprehension of bilingual subtitled viewing compared to the other three subtitling conditions. Since one overall comprehension score was assigned to each participant, linear regression was used to explore the effect of Group on participants' comprehension scores without including any random effects. Participants' vocabulary size scores were log-transformed and added as a covariate. The interaction between Group and vocabulary size was also checked.

To address RQ3, which compared participants' online processing in four subtitling groups, a series of mixed-effects models were constructed based on the type of dependent variables. Linear mixed-effects models were built for continuous dependent variables (i.e., total reading time %, fixation %, total reading time, first-pass reading time, first fixation duration, second-pass reading time, second fixation duration); logistic mixed-effects models were constructed for binary dependent variables (i.e., skip rate); Poisson mixed-effects analyses were conducted for count dependent variable (i.e., fixation count, run count). For all models, each eye-tracking measure was taken as dependent variable and Group as the fixed effect. For Level 1 (i.e., whole subtitling area) and Level 2 (i.e., each subtitle line) analyses, where no TWs were concerned, Participant and different IPs were added as random intercepts. Group was also checked as random slope by IPs. For Level 3 analyses at TW level, only TWs that were unknown (scored 0 on the form recognition pretest) to each participant were analysed.

RQ4 explored the relationship between the online and offline measures, to examine whether participants' eye movements on each unknown TW increased their vocabulary learning gains. The effects of three eye-tracking measures (i.e., total reading time, first-pass reading time, second-pass reading time) on three vocabulary posttests (i.e., form recognition, meaning recall, meaning recognition) were explored. The eye-movement data for this RQ in particular were first transformed from milliseconds to

seconds for a clearer interpretation of findings. Logistic mixed-effects models with vocabulary posttest scores as dependent variable and eye-movement data as independent variable were constructed for each subtitling condition separately in order to rule out the multicollinearity issue (Montero Perez et al., 2015). For the bilingual subtitles group, two sets of models were conducted for L1 translations of unknown TWs and L2 unknown TWs separately.

#### ***4.1.5. Preliminary Analyses***

Before conducting the main analyses, preliminary analyses were run to ensure the reliability of tests and the validity of results. Results showed that Cronbach's alpha coefficients for all tests were above .80 (pre-and post- form recognition:  $\alpha = .83$  and  $.89$ ; occurrence decision test:  $\alpha = .83$ ; pre- and post- meaning recall:  $\alpha = .83$  and  $.84$ ; pre- and post- meaning recognition:  $\alpha = .85$ ; comprehension test:  $\alpha = .83$ ), indicating good reliability. Pairwise point-biserial correlation tests among four vocabulary tests indicated that all correlations reached significance level ( $p < .001$ ), with a large correlation between form recognition and occurrence decision tests ( $r_{pb} = 0.74$ ). Medium correlations were reported between form recognition and meaning recall ( $r_{pb} = 0.38$ ), form recognition and meaning recognition ( $r_{pb} = 0.30$ ), and between meaning recall and meaning recognition ( $r_{pb} = 0.44$ ).

## **4.2. QUAN Results**

### ***4.2.1. RQ1 Effects of Subtitles on Vocabulary Learning***

RQ1 aimed to explore relative effects of bilingual subtitles on incidental vocabulary learning by comparing to other subtitling conditions. The comparability of four groups at the outset of the study was first checked. Descriptive statistics for the performance of each group on pretests and posttests are summarised in Table 17. The comparability was examined using three sets of logistic mixed-effects models with participants' pretest scores as dependent variables. The results showed that by comparing with null models, models with Group as fixed effect did not significantly improve the model fit for the three vocabulary tests: form recognition ( $\chi^2(3) = 0.72, p = .87, R^2 < .001$ ), meaning recall ( $\chi^2(3) = 0.78, p = .85, R^2 < .001$ ), and meaning recognition ( $\chi^2(3) = 1.79, p = .62, R^2 < .001$ ), indicating no significant group differences at the onset of the study.

Table 17. Descriptive Statistics for the Vocabulary Pretests and Posttests by Group

| Group               | Captions<br>( <i>n</i> = 27) |                 |                |                 | L1 subtitles<br>( <i>n</i> = 24) |                 |                |                 | Bilingual subtitles<br>( <i>n</i> = 30) |                 |                |                 | No subtitles<br>( <i>n</i> = 25) |                 |                |                 |
|---------------------|------------------------------|-----------------|----------------|-----------------|----------------------------------|-----------------|----------------|-----------------|-----------------------------------------|-----------------|----------------|-----------------|----------------------------------|-----------------|----------------|-----------------|
|                     | Pretests                     |                 | Posttests      |                 | Pretests                         |                 | Posttests      |                 | Pretests                                |                 | Posttests      |                 | Pretests                         |                 | Posttests      |                 |
|                     | <i>M</i>                     | 95%             | <i>M</i>       | 95%             | <i>M</i>                         | 95%             | <i>M</i>       | 95%             | <i>M</i>                                | 95%             | <i>M</i>       | 95%             | <i>M</i>                         | 95%             | <i>M</i>       | 95%             |
|                     | ( <i>SD</i> )                | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )                    | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )                           | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )                    | CI              | ( <i>SD</i> )  | CI              |
| Form recognition    | 0.26<br>(0.44)               | [0.23,<br>0.30] | 0.66<br>(0.47) | [0.63,<br>0.70] | 0.25<br>(0.44)                   | [0.22,<br>0.29] | 0.52<br>(0.50) | [0.48,<br>0.56] | 0.24<br>(0.43)                          | [0.21,<br>0.27] | 0.53<br>(0.50) | [0.49,<br>0.56] | 0.28<br>(0.45)                   | [0.24,<br>0.31] | 0.49<br>(0.50) | [0.45,<br>0.53] |
| Occurrence Decision | \                            | \               | 0.58<br>(0.49) | [0.54,<br>0.62] | \                                | \               | 0.34<br>(0.48) | [0.30,<br>0.38] | \                                       | \               | 0.43<br>(0.50) | [0.40,<br>0.47] | \                                | \               | 0.33<br>(0.47) | [0.30,<br>0.37] |
| Meaning recall      | 0.06<br>(0.25)               | [0.05,<br>0.08] | 0.17<br>(0.38) | [0.14,<br>0.20] | 0.06<br>(0.24)                   | [0.04,<br>0.08] | 0.14<br>(0.35) | [0.11,<br>0.17] | 0.05<br>(0.23)                          | [0.04,<br>0.07] | 0.20<br>(0.40) | [0.17,<br>0.23] | 0.08<br>(0.28)                   | [0.06,<br>0.11] | 0.13<br>(0.34) | [0.10,<br>0.16] |
| Meaning recognition | 0.25<br>(0.43)               | [0.22,<br>0.29] | 0.42<br>(0.49) | [0.38,<br>0.45] | 0.20<br>(0.40)                   | [0.17,<br>0.23] | 0.43<br>(0.50) | [0.39,<br>0.47] | 0.22<br>(0.41)                          | [0.19,<br>0.25] | 0.53<br>(0.50) | [0.49,<br>0.56] | 0.23<br>(0.42)                   | [0.20,<br>0.27] | 0.36<br>(0.48) | [0.32,<br>0.40] |

*Note.* max = 1 in all cases

Descriptive statistics in Table 17 show that for the four groups, posttest scores were in general higher than those in pretests. Following Peters and Webb (2018), the absolute learning gains for the 24 TWs were calculated at item level and summarised in Table 18. The captions group obtained the highest absolute gains on both form recognition ( $M = 0.42$ ,  $SD = 0.49$ ) and occurrence decision test ( $M = 0.58$ ,  $SD = 0.49$ ), while the highest learning gains on meaning recall ( $M = 0.15$ ,  $SD = 0.35$ ) and meaning recognition ( $M = 0.33$ ,  $SD = 0.47$ ) were observed in the bilingual subtitles group.

Table 18. Absolute Vocabulary Learning Gains of 24 Target Words on Four Offline Vocabulary Tests by Group

| Group               | Form recognition |              | Occurrence decision |              | Meaning recall |              | Meaning recognition |              |
|---------------------|------------------|--------------|---------------------|--------------|----------------|--------------|---------------------|--------------|
|                     | $M$              | 95%          | $M$ ( $SD$ )        | 95%          | $M$ ( $SD$ )   | 95%          | $M$ ( $SD$ )        | 95%          |
|                     | ( $SD$ )         | CI           |                     | CI           |                | CI           |                     | CI           |
| Captions            | 0.42<br>(0.49)   | [0.38, 0.46] | 0.58<br>(0.49)      | [0.54, 0.62] | 0.11<br>(0.32) | [0.09, 0.14] | 0.19<br>(0.39)      | [0.16, 0.22] |
| L1 subtitles        | 0.31<br>(0.46)   | [0.27, 0.35] | 0.34<br>(0.48)      | [0.30, 0.38] | 0.08<br>(0.27) | [0.06, 0.10] | 0.25<br>(0.43)      | [0.22, 0.29] |
| Bilingual subtitles | 0.32<br>(0.47)   | [0.28, 0.35] | 0.43<br>(0.50)      | [0.40, 0.47] | 0.15<br>(0.35) | [0.12, 0.17] | 0.33<br>(0.47)      | [0.29, 0.36] |
| No subtitles        | 0.26<br>(0.44)   | [0.23, 0.30] | 0.33<br>(0.47)      | [0.30, 0.37] | 0.06<br>(0.23) | [0.04, 0.08] | 0.17<br>(0.37)      | [0.14, 0.20] |

*Note.* max = 1 in all cases

To further investigate the statistical significance concerning the learning of the 24 TWs between groups, four separate sets of logistic mixed-effects models for each type of vocabulary test were fitted (see Table 19). The bilingual subtitles group was set as the baseline group for all analyses. As revealed in Table 19, both participants' vocabulary size and their prior knowledge of the TWs (except for occurrence decision

test) showed significant positive effects on four posttest scores. The interaction between Group and participants' vocabulary size was also checked but it did not significantly improve the model fit. Word length and frequency of occurrence had significant positive effects on meaning recall and meaning recognition, respectively, indicating that the meaning of a longer TW had a better chance of being recalled correctly, and that the meaning of a word was more likely to be recognised if it appeared more frequently in the video. Post hoc analyses were performed to investigate pairwise group differences for the four vocabulary tests separately.

Table 19. Results of the Logistic Mixed-Effects Models on Four Offline Vocabulary Tests Between Four Subtitling Groups

| Form Recognition                                                                 |          |                |           |          |          | Occurrence Decision                                                  |                |           |          |          |
|----------------------------------------------------------------------------------|----------|----------------|-----------|----------|----------|----------------------------------------------------------------------|----------------|-----------|----------|----------|
| Fixed effects                                                                    | <i>b</i> | 95% CI         | <i>SE</i> | <i>z</i> | <i>p</i> | <i>b</i>                                                             | 95% CI         | <i>SE</i> | <i>z</i> | <i>p</i> |
| Intercept                                                                        | -1.37    | [-1.92, -0.82] | 0.28      | -4.90    | <.001    | -1.38                                                                | [-1.93, -0.83] | 0.28      | -4.94    | <.001    |
| Captions                                                                         | 0.77     | [0.26, 1.28]   | 0.26      | 2.94     | .003     | 0.73                                                                 | [0.22, 1.24]   | 0.26      | 2.81     | .004     |
| L1 subtitles                                                                     | -0.07    | [-0.60, 0.46]  | 0.27      | -0.25    | .80      | -0.52                                                                | [-1.05, 0.01]  | 0.27      | -1.90    | .06      |
| No subtitles                                                                     | -0.30    | [-0.83, 0.23]  | 0.27      | -1.11    | .27      | -0.59                                                                | [-1.12, -0.06] | 0.27      | -2.17    | .03      |
| PreScores                                                                        | 1.92     | [1.63, 2.21]   | 0.15      | 12.97    | <.001    | \                                                                    | \              | \         | \        | \        |
| res.Vsize                                                                        | 2.62     | [1.80, 3.44]   | 0.42      | 6.22     | <.001    | 2.44                                                                 | [1.62, 3.26]   | 0.42      | 5.80     | <.001    |
| Random effects                                                                   | Variance |                | <i>SD</i> |          |          | Variance                                                             |                | <i>SD</i> |          |          |
| Participant (Intercept)                                                          | 0.71     |                | 0.84      |          |          | 0.75                                                                 |                | 0.87      |          |          |
| Item (Intercept)                                                                 | 0.33     |                | 0.57      |          |          | 0.31                                                                 |                | 0.56      |          |          |
| Best model: frV ~ Group + PreScores + res.Vsize + (1   Participant) + (1   Item) |          |                |           |          |          | Best model: odV ~ Group + res.Vsize + (1   Participant) + (1   Item) |                |           |          |          |
| Hosmer and Lemeshow’s $R^2 = .09$                                                |          |                |           |          |          | Hosmer and Lemeshow’s $R^2 = .13$                                    |                |           |          |          |

| Meaning Recall                                                                                      |          |                |           |          |          | Meaning Recognition                                                                            |                |           |          |          |
|-----------------------------------------------------------------------------------------------------|----------|----------------|-----------|----------|----------|------------------------------------------------------------------------------------------------|----------------|-----------|----------|----------|
| Fixed effects                                                                                       | <i>b</i> | 95% CI         | <i>SE</i> | <i>z</i> | <i>p</i> | <i>b</i>                                                                                       | 95% CI         | <i>SE</i> | <i>z</i> | <i>p</i> |
| Intercept                                                                                           | -4.16    | [-5.06, -3.26] | 0.46      | -9.09    | <.001    | -1.56                                                                                          | [-2.05, -1.07] | 0.25      | -6.16    | <.001    |
| Captions                                                                                            | -0.32    | [-0.81, 0.17]  | 0.25      | -1.28    | .20      | -0.78                                                                                          | [-1.17, -0.39] | 0.20      | -3.99    | <.001    |
| L1 subtitles                                                                                        | -0.77    | [-1.30, -0.24] | 0.27      | -2.80    | .01      | -0.48                                                                                          | [-0.87, -0.09] | 0.20      | -2.42    | .02      |
| No subtitles                                                                                        | -1.23    | [-1.80, -0.66] | 0.29      | -4.24    | <.001    | -1.08                                                                                          | [-1.47, -0.69] | 0.20      | -5.36    | <.001    |
| PreScores                                                                                           | 4.42     | [3.73, 5.11]   | 0.35      | 12.47    | <.001    | 2.68                                                                                           | [2.37, 3.00]   | 0.16      | 16.32    | <.001    |
| res.Vsize                                                                                           | 3.20     | [2.34, 4.06]   | 0.44      | 7.26     | <.001    | 2.58                                                                                           | [1.95, 3.21]   | 0.32      | 7.95     | <.001    |
| res.ItemLength                                                                                      | 1.51     | [0.18, 2.84]   | 0.68      | 2.21     | .03      | \                                                                                              | \              | \         | \        | \        |
| res.FoO                                                                                             | \        | \              | \         | \        | \        | 1.13                                                                                           | [0.17, 2.09]   | 0.49      | 2.32     | .02      |
| Random effects                                                                                      | Variance |                | <i>SD</i> |          |          | Variance                                                                                       |                | <i>SD</i> |          |          |
| Participant (Intercept)                                                                             | 0.41     |                | 0.64      |          |          | 0.27                                                                                           |                | 0.52      |          |          |
| Item (Intercept)                                                                                    | 0.68     |                | 0.82      |          |          | 0.52                                                                                           |                | 0.72      |          |          |
| Best model: mrecallVL ~ Group + PreScores + res.Vsize + res.ItemLength + (1 Participant) + (1 Item) |          |                |           |          |          | Best model: mrecoVF ~ Group + PreScores + res.Vsize + res.ItemFoO + (1 Participant) + (1 Item) |                |           |          |          |
| Hosmer and Lemeshow's $R^2 = .20$                                                                   |          |                |           |          |          | Hosmer and Lemeshow's $R^2 = .15$                                                              |                |           |          |          |

Controlling for the relevant covariates (i.e., pretest scores, vocabulary size, word length, frequency of occurrence) for each vocabulary test, the post hoc results, shown in Table 20, indicate that for form recognition, the captions group significantly outperformed the no subtitles, L1, and bilingual subtitles groups, with the odds of a correct answer in the posttest being 2.92, 2.32, and 2.17 times higher than in these three groups, respectively. Same patterns were observed in the occurrence decision test, with the odds of a correct answer in the captions group being 3.74, 3.49, and 2.08 times higher than those in the no subtitles, L1, and bilingual subtitles groups, respectively. No group differences were revealed among the L1, bilingual, and no subtitles groups in terms of form aspect.

Table 20 also shows that, in terms of meaning recall, the bilingual subtitles group significantly outperformed the L1 and no subtitles groups, with the odds of a correct answer being 3.42 and 2.16 times higher than those in these two groups separately. Moreover, the captions group also significantly outperformed the no subtitles group, with 2.48 times higher odds of a correct answer. No significant difference was revealed between the bilingual subtitles and captions groups in meaning recall. For meaning recognition, the bilingual subtitles significantly outperformed the no subtitles and captions groups, with 2.94 and 2.18 times higher odds of a correct answer than these two groups, respectively. The significant difference between the L1 and bilingual subtitles was also approaching significance level ( $p = .07$ ). Additionally, the L1 subtitles significantly outperformed the no subtitles group on meaning recognition, with 1.82 times higher odds of a correct answer. No significant difference was revealed between captions and L1 subtitles on meaning-related vocabulary tests.

Table 20. Results of Post Hoc Contrasts on Four Offline Vocabulary Tests Between Four Subtitling Groups

| Group                      | <i>b</i> | 95% CI        | <i>SE</i> | <i>z</i> | OR   | OR 95% CI    | <i>p</i>        |
|----------------------------|----------|---------------|-----------|----------|------|--------------|-----------------|
| <b>Form Recognition</b>    |          |               |           |          |      |              |                 |
| Captions > Bilingual       | 0.77     | [0.26, 1.28]  | 0.26      | 2.94     | 2.17 | [1.29, 3.62] | <b>.02</b>      |
| Bilingual – L1             | 0.07     | [-0.46, 0.60] | 0.27      | 0.25     | 1.07 | [0.63, 1.82] | .99             |
| Bilingual – No             | 0.30     | [-0.23, 0.83] | 0.27      | 1.11     | 1.35 | [0.80, 2.29] | .68             |
| Captions > L1              | 0.84     | [0.29, 1.39]  | 0.28      | 3.04     | 2.32 | [1.35, 3.99] | <b>.01</b>      |
| Captions > No              | 1.07     | [0.52, 1.62]  | 0.28      | 3.88     | 2.92 | [1.70, 5.00] | <b>&lt;.001</b> |
| L1 – No                    | 0.23     | [-0.32, 0.78] | 0.28      | 0.81     | 1.26 | [0.73, 2.18] | .85             |
| <b>Occurrence Decision</b> |          |               |           |          |      |              |                 |
| Captions > Bilingual       | 0.73     | [0.22, 1.24]  | 0.26      | 2.81     | 2.08 | [1.25, 3.48] | <b>.03</b>      |
| Bilingual – L1             | 0.52     | [-0.01, 1.05] | 0.27      | -1.90    | 0.59 | [0.35, 1.01] | .23             |
| Bilingual – No             | 0.59     | [-0.06, 1.12] | 0.27      | -2.17    | 0.55 | [0.33, 0.94] | .13             |
| Captions > L1              | 1.25     | [0.70, 1.80]  | 0.28      | 4.47     | 3.49 | [2.02, 6.06] | <b>&lt;.001</b> |
| Captions > No              | 1.32     | [0.77, 1.87]  | 0.28      | 4.76     | 3.74 | [2.17, 6.45] | <b>&lt;.001</b> |
| L1 – No                    | 0.07     | [-0.50, 0.64] | 0.29      | -0.23    | 0.93 | [0.53, 1.64] | .10             |
| <b>Meaning Recall</b>      |          |               |           |          |      |              |                 |
| Bilingual – Captions       | 0.32     | [-0.17, 0.81] | 0.25      | 1.28     | 1.38 | [0.84, 2.25] | .57             |
| Bilingual > L1             | 0.77     | [0.24, 1.30]  | 0.27      | 2.80     | 2.16 | [1.26, 3.69] | <b>.03</b>      |

| Group                           | <i>b</i> | 95% CI        | <i>SE</i> | <i>z</i> | OR   | OR 95% CI    | <i>p</i>        |
|---------------------------------|----------|---------------|-----------|----------|------|--------------|-----------------|
| Bilingual > No<br>Captions – L1 | 1.23     | [0.66, 1.80]  | 0.29      | 4.24     | 3.42 | [1.93, 6.01] | <b>&lt;.001</b> |
| Captions > No<br>L1 – No        | 0.45     | [-0.12, 1.02] | 0.29      | 1.57     | 1.57 | [0.89, 2.77] | .40             |
|                                 | 0.91     | [0.32, 1.50]  | 0.30      | 3.03     | 2.48 | [1.38, 4.45] | <b>.01</b>      |
|                                 | 0.46     | [-0.17, 1.09] | 0.32      | 1.45     | 1.58 | [0.85, 2.97] | .47             |
| <b>Meaning Recognition</b>      |          |               |           |          |      |              |                 |
| Bilingual > Captions            | 0.78     | [0.39, 1.17]  | 0.20      | 3.99     | 2.18 | [1.49, 3.21] | <b>&lt;.001</b> |
| Bilingual – L1                  | 0.48     | [0.09, 0.87]  | 0.20      | 2.42     | 1.62 | [1.09, 2.39] | .07             |
| Bilingual > No                  | 1.08     | [0.69, 1.47]  | 0.20      | 5.36     | 2.94 | [1.99, 4.29] | <b>&lt;.001</b> |
| L1 – Captions                   | 0.30     | [-0.11, 0.71] | 0.21      | 1.45     | 1.35 | [0.66, 2.04] | .47             |
| Captions – No                   | 0.30     | [-0.11, 0.71] | 0.21      | 1.44     | 1.35 | [0.89, 2.04] | .47             |
| L1 > No                         | 0.60     | [0.19, 1.01]  | 0.21      | 2.81     | 1.82 | [1.20, 2.77] | <b>.03</b>      |

To examine potential test effects, participants' posttest scores of 24 TWs and 6 distractors were compared in each vocabulary test, while controlling for their pretest scores. Model summaries are presented in Table 21 to Table 23. The logistic mixed-effects models showed that Item Type (TW or distractor) significantly predicted posttest scores in form recognition ( $\chi^2(2) = 256.79, p < .001, R^2 = .16$ ), meaning recall ( $\chi^2(2) = 355.73, p < .001, R^2 = .23$ ), and meaning recognition ( $\chi^2(2) = 440.73, p < .001, R^2 = .27$ ), indicating that the odds of a correct answer in the posttest were 3.55 times higher for TWs than for distractors in form recognition, 25.12 times higher in meaning recall, and 7.48 times higher in meaning recognition. This suggests that there were significant gains from the treatments beyond the possible test effects.

Table 21. Results of the Logistic Mixed-Effects Models Comparing Form Recognition Posttest Scores for 24 Target Words and 6 Distractors

| Form Recognition                                                           |          |           |           |      |              |          |
|----------------------------------------------------------------------------|----------|-----------|-----------|------|--------------|----------|
| Fixed effects                                                              | <i>b</i> | <i>SE</i> | <i>z</i>  | OR   | OR 95% CI    | <i>p</i> |
| Intercept                                                                  | -1.42    | 0.27      | -5.22     | 0.24 | [0.14, 0.41] | <.001    |
| FRPre                                                                      | 1.96     | 0.14      | 14.30     | 7.10 | [5.42, 9.30] | <.001    |
| TW                                                                         | 1.27     | 0.28      | 4.56      | 3.55 | [2.05, 6.17] | <.001    |
| Random effects                                                             | Variance |           | <i>SD</i> |      |              |          |
| Participant (Intercept)                                                    | 1.24     |           | 1.11      |      |              |          |
| Item (Intercept)                                                           | 0.31     |           | 0.55      |      |              |          |
| Best model: FR2 <- glmer(FRPost ~ FRPre + TW + (1 Participant) + (1 Item)) |          |           |           |      |              |          |
| Hosmer and Lemeshow's $R^2 = .16$                                          |          |           |           |      |              |          |

Table 22. Results of the Logistic Mixed-Effects Models Comparing Meaning Recall Posttest Scores for 24 Target Words and 6 Distractors

| Meaning Recall |          |           |          |    |           |          |
|----------------|----------|-----------|----------|----|-----------|----------|
| Fixed effects  | <i>b</i> | <i>SE</i> | <i>z</i> | OR | OR 95% CI | <i>p</i> |

|                                                                                           |          |      |       |        |                 |       |
|-------------------------------------------------------------------------------------------|----------|------|-------|--------|-----------------|-------|
| Intercept                                                                                 | -5.89    | 0.65 | -9.08 | 0.003  | [0.001, 0.01]   | <.001 |
| MRecallPre                                                                                | 4.79     | 0.37 | 13.13 | 120.74 | [57.97, 249.64] | <.001 |
| TW                                                                                        | 3.22     | 0.66 | 4.89  | 25.12  | [6.89, 91.84]   | <.001 |
| Random effects                                                                            | Variance |      | SD    |        |                 |       |
| Participant (Intercept)                                                                   | 1.10     |      | 1.05  |        |                 |       |
| Item (Intercept)                                                                          | 0.84     |      | 0.92  |        |                 |       |
| Best model: MRecall2 <- glmer(MRecallPost ~ MRecallPre + TW + (1 Participant) + (1 Item)) |          |      |       |        |                 |       |
| Hosmer and Lemeshow's $R^2 = .23$                                                         |          |      |       |        |                 |       |

Table 23. Results of the Logistic Mixed-Effects Models Comparing Meaning Recognition Posttest Scores for 24 Target Words and 6 Distractors

| Meaning Recognition                                                                 |          |           |           |       |                |          |
|-------------------------------------------------------------------------------------|----------|-----------|-----------|-------|----------------|----------|
| Fixed effects                                                                       | <i>b</i> | <i>SE</i> | <i>z</i>  | OR    | OR 95% CI      | <i>p</i> |
| Intercept                                                                           | -2.89    | 0.38      | -7.66     | 0.06  | [0.03, 0.12]   | <.001    |
| MrecoPre                                                                            | 2.80     | 0.16      | 18.00     | 16.44 | [12.06, 22.42] | <.001    |
| TW                                                                                  | 2.01     | 0.40      | 4.98      | 7.48  | [3.42, 16.28]  | <.001    |
| Random effects                                                                      | Variance |           | <i>SD</i> |       |                |          |
| Participant (Intercept)                                                             | 0.77     |           | 0.88      |       |                |          |
| Item (Intercept)                                                                    | 0.65     |           | 0.81      |       |                |          |
| Best model: MReco2 <- glmer(MrecoPost ~ MrecoPre + TW + (1 Participant) + (1 Item)) |          |           |           |       |                |          |
| Hosmer and Lemeshow's $R^2 = .27$                                                   |          |           |           |       |                |          |

In sum, in terms of form recognition, the bilingual subtitles seemed to be less effective than captions, as revealed in both form recognition and occurrence decision tests. However, bilingual subtitles tended to be more helpful for facilitating meaning knowledge. They significantly outperformed the L1 and no subtitles groups on meaning recall, and the captions and no subtitles groups on meaning recognition. The bilingual

subtitles group also outperformed the L1 subtitles at an approaching significance level on meaning recognition. Therefore, the advantages of bilingual subtitles seemed to be more salient for learning word meanings. The next research question compared the comprehension scores across four groups.

#### 4.2.2. RQ2 Effects of Subtitles on Comprehension

RQ2 investigated the relative effects of bilingual subtitles on comprehension. As shown in the descriptive statistics presented in Table 24, both the bilingual and L1 subtitles groups achieved the highest mean scores, with an average correct rate around 80%. Linear regression models were run to explore group differences. As shown in Table 25, there were significant group differences between the bilingual subtitles and other groups, and participants with a larger vocabulary size achieved higher comprehension scores.

Table 24. Descriptive Statistics for Comprehension Scores by Group

| Group                                | Comprehension scores |                |
|--------------------------------------|----------------------|----------------|
|                                      | <i>M (SD)</i>        | 95% CI         |
| Captions ( <i>n</i> = 27)            | 23.48 (4.15)         | [21.84, 25.12] |
| L1 subtitles ( <i>n</i> = 24)        | 27.38 (4.71)         | [25.39, 29.36] |
| Bilingual subtitles ( <i>n</i> = 30) | 27.93 (4.23)         | [26.35, 29.51] |
| No subtitles ( <i>n</i> = 25)        | 20.48 (4.81)         | [18.50, 22.46] |

*Note.* max = 34 in all cases

Table 25. Results of the Linear Regression Model Examining Comprehension Performance

|           | <i>b</i> | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> |
|-----------|----------|----------------|-----------|----------|----------|
| Intercept | 23.68    | [20.23, 27.13] | 1.76      | 13.48    | <.001    |
| Captions  | -4.45    | [-6.70, -2.20] | 1.15      | -3.88    | <.001    |

|              | <i>b</i> | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i>        |
|--------------|----------|----------------|-----------|----------|-----------------|
| L1 subtitles | -0.53    | [-2.86, 1.80]  | 1.19      | -0.45    | .66             |
| No subtitles | -7.48    | [-9.78, -5.19] | 1.17      | -6.37    | <b>&lt;.001</b> |
| log.Vsize    | 2.18     | [1.08, 3.28]   | 0.56      | 2.69     | <b>.01</b>      |

Best model: Comprehension\_scores ~ Group + log.Vsize  
 $R^2 = .37$

Post hoc pairwise comparisons were run to further investigate group differences. Table 26 shows that both the bilingual and L1 subtitles groups performed significantly better than the no subtitles group with large effect sizes. Moreover, the bilingual and L1 subtitles groups also significantly outperformed the captions group, with medium and small effect sizes, respectively. It should be noted that the difference between the captions ( $M = 69\%$ ) and no subtitles groups ( $M = 60\%$ ) was only approaching significance ( $p = .06$ ).

Table 26. Results of Post Hoc Contrasts on Comprehension Performance

| Group                | <i>b</i> | 95% CI        | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|----------------------|----------|---------------|-----------|----------|-----------------|----------|
| Bilingual > Captions | 4.46     | [2.21, 6.71]  | 1.15      | 3.88     | <b>&lt;.001</b> | 0.89     |
| Bilingual – L1       | 0.52     | [-1.81, 2.85] | 1.19      | 0.44     | .97             | 0.26     |
| Bilingual > No       | 7.48     | [5.19, 9.77]  | 1.17      | 6.38     | <b>&lt;.001</b> | 1.47     |
| L1 > Captions        | 3.94     | [1.55, 6.33]  | 1.22      | 3.24     | <b>.01</b>      | 0.56     |
| Captions – No        | 3.02     | [0.67, 5.37]  | 1.20      | 2.51     | .06             | 0.69     |
| L1 > No              | 6.96     | [4.53, 9.39]  | 1.24      | 5.62     | <b>&lt;.001</b> | 1.13     |

### 4.2.3. RQ3 Eye Movements During Subtitled Viewing

#### 4.2.3.1. Level 1: Overall Subtitling Area

To answer RQ3, the eye-tracking data were analysed based on three levels (see section 4.1.2). Level 1 compared participants' attention allocation to the overall subtitling area across four groups. Descriptive statistics for the attention distribution to subtitling and image areas in the four groups are provided in Table 27. Four types of eye-tracking measures (i.e., total reading time %, fixation %, run count, skip rate) were reported for this level.

As shown in Table 27, participants in both captions and bilingual subtitles groups spent about 60% of their time on the subtitling area while viewing, and about 35% of the time was allocated to the image area. Moreover, these two groups had less chance of skipping the on-screen text than other groups. On the contrary, participants using L1 subtitles tended to spend slightly more time on images (55%) than the subtitling area (44%). Reasonably, the no subtitles group, without any presented subtitles, spent most of the time on images (94%), and recorded the highest skip rate (0.88) and lowest run count (0.13) on the subtitling area. To compare participants' processing time on the subtitling area in four subtitling groups, four sets of mixed-effects models were constructed for the four eye-tracking measures separately.

Table 27. Descriptive Statistics for Eye-Movement Data at Level 1 for Overall Subtitling and Image Area by Group

| Group                                      | Total reading time % |                 |                |                 | Fixation %     |                 |                |                 | Run count      |                 |                |                 | Skip rate      |                 |                |                 |
|--------------------------------------------|----------------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|
|                                            | Subtitle             |                 | Image          |                 | Subtitle       |                 | Image          |                 | Subtitle       |                 | Image          |                 | Subtitle       |                 | Image          |                 |
|                                            | <i>M</i>             | 95%             | <i>M</i>       | 95%             | <i>M</i>       | 95%             | <i>M</i>       | 95%             | <i>M</i>       | 95%             | <i>M</i>       | 95%             | <i>M</i>       | 95%             | <i>M</i>       | 95%             |
|                                            | ( <i>SD</i> )        | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )  | CI              | ( <i>SD</i> )  | CI              |
| Captions<br>( <i>n</i> = 25)               | 0.64<br>(0.31)       | [0.63,<br>0.64] | 0.35<br>(0.31) | [0.35,<br>0.36] | 0.68<br>(0.28) | [0.68,<br>0.69] | 0.31<br>(0.28) | [0.31,<br>0.32] | 1.15<br>(0.56) | [1.18,<br>1.22] | 1.07<br>(0.78) | [1.05,<br>1.08] | 0.07<br>(0.26) | [0.07,<br>0.07] | 0.24<br>(0.42) | [0.23,<br>0.24] |
| L1<br>subtitles<br>( <i>n</i> = 23)        | 0.44<br>(0.29)       | [0.43,<br>0.44] | 0.55<br>(0.29) | [0.54,<br>0.55] | 0.50<br>(0.27) | [0.49,<br>0.50] | 0.48<br>(0.27) | [0.48,<br>0.49] | 1.21<br>(0.65) | [1.20,<br>1.22] | 1.36<br>(0.70) | [1.34,<br>1.37] | 0.10<br>(0.30) | [0.09,<br>0.10] | 0.08<br>(0.28) | [0.08,<br>0.09] |
| Bilingual<br>subtitles<br>( <i>n</i> = 28) | 0.60<br>(0.32)       | [0.59,<br>0.60] | 0.39<br>(0.32) | [0.38,<br>0.39] | 0.63<br>(0.30) | [0.62,<br>0.63] | 0.36<br>(0.29) | [0.35,<br>0.36] | 1.22<br>(0.63) | [1.21,<br>1.23] | 1.14<br>(0.78) | [1.12,<br>1.15] | 0.08<br>(0.27) | [0.07,<br>0.08] | 0.21<br>(0.41) | [0.20,<br>0.21] |
| No<br>subtitles<br>( <i>n</i> = 24)        | 0.04<br>(0.13)       | [0.03,<br>0.04] | 0.94<br>(0.20) | [0.93,<br>0.94] | 0.04<br>(0.12) | [0.03,<br>0.04] | 0.94<br>(0.20) | [0.93,<br>0.94] | 0.13<br>(0.40) | [0.13,<br>0.14] | 1.04<br>(0.32) | [1.03,<br>1.04] | 0.88<br>(0.32) | [0.88,<br>0.89] | 0.03<br>(0.17) | [0.03,<br>0.03] |

The results of mixed-effects models revealed that Group was a significant predictor for participants' processing of the overall subtitling area, indicating significant group differences. In addition, participants' vocabulary size also showed a significant negative effect on participants' attention allocation to the subtitling area, suggesting that learners with a larger vocabulary size tended to spend less time on on-screen text, and were more likely to skip the on-screen text. The potential interaction between Group and vocabulary size was further analysed at Level 2. Model summaries for four eye-tracking measures (i.e., total reading time %, fixation %, run count, skip rate) are presented in Appendix S11.

To further investigate group differences, Table 28 and Table 29 summarise post hoc pairwise comparisons with Tukey correction for the four eye-tracking measures. Results presented in Table 28 revealed that, as expected, the no subtitles group spent significantly less time on the subtitling area than the other three groups, as reported by both total reading time % and fixation %, with large effect sizes. Among the three groups with on-screen text, the bilingual subtitles and captions groups spent similar amount of time on the subtitling area, which were both significantly longer than that in the L1 subtitles group, with small effect sizes. In terms of the run count and skip rate, as shown in Table 29, three groups with on-screen text had significant higher odds of switching between the image and subtitling areas and lower odds of skipping the subtitling area compared to the no subtitles group. However, there were no statistical differences among the three groups with on-screen text in terms of run count or skip rate. Only the L1 subtitles group showed a higher skip rate than the captions group, with an approaching significance group difference ( $p = .07$ ).

Table 28. Results of Post Hoc Contrasts for Total Reading Time % and Fixation % at Level 1 Overall Subtitling Area

| Group                       | <i>b</i> | 95% CI        | <i>SE</i> | <i>z</i> | <i>p</i>        | <i>d</i> |
|-----------------------------|----------|---------------|-----------|----------|-----------------|----------|
| <b>Total reading time %</b> |          |               |           |          |                 |          |
| Captions – Bilingual        | 0.03     | [-0.01, 0.07] | 0.02      | 1.11     | .69             | 0.13     |
| Bilingual > L1              | 0.11     | [0.07, 0.15]  | 0.02      | 4.45     | <b>&lt;.001</b> | 0.53     |
| Bilingual > No              | 0.42     | [0.38, 0.46]  | 0.02      | 17.61    | <b>&lt;.001</b> | 2.25     |
| Captions > L1               | 0.13     | [0.09, 0.17]  | 0.02      | 5.39     | <b>&lt;.001</b> | 0.68     |
| Captions > No               | 0.44     | [0.40, 0.48]  | 0.02      | 18.20    | <b>&lt;.001</b> | 2.53     |
| L1 > No                     | 0.31     | [0.27, 0.35]  | 0.02      | 12.49    | <b>&lt;.001</b> | 1.81     |
| <b>Fixation %</b>           |          |               |           |          |                 |          |
| Captions – Bilingual        | 0.03     | [-0.01, 0.07] | 0.02      | 1.52     | .42             | 0.18     |
| Bilingual > L1              | 0.08     | [0.05, 0.12]  | 0.02      | 3.74     | <b>.001</b>     | 0.46     |
| Bilingual > No              | 0.44     | [0.40, 0.48]  | 0.02      | 19.92    | <b>&lt;.001</b> | 2.53     |
| Captions > L1               | 0.12     | [0.08, 0.16]  | 0.02      | 5.09     | <b>&lt;.001</b> | 0.66     |
| Captions > No               | 0.47     | [0.43, 0.51]  | 0.02      | 20.85    | <b>&lt;.001</b> | 2.93     |
| L1 > No                     | 0.36     | [0.32, 0.40]  | 0.02      | 15.38    | <b>&lt;.001</b> | 2.21     |

Table 29. Results of Post Hoc Contrasts for Run Count and Skip Rate at Level 1 Overall Subtitling Area

| Group                | <i>b</i> | <i>SE</i> | <i>z</i> | OR    | OR 95% CI      | <i>p</i>        |
|----------------------|----------|-----------|----------|-------|----------------|-----------------|
| <b>Run count</b>     |          |           |          |       |                |                 |
| Captions – Bilingual | -0.06    | 0.08      | -0.73    | 0.94  | [0.80, 1.10]   | .89             |
| L1 – Bilingual       | -0.01    | 0.08      | -0.10    | 0.99  | [0.84, 1.17]   | 1.00            |
| Bilingual > No       | 2.30     | 0.09      | 26.83    | 9.97  | [8.43, 11.80]  | <b>&lt;.001</b> |
| L1 – Captions        | 0.05     | 0.09      | 0.60     | 1.05  | [0.89, 1.24]   | .93             |
| Captions > No        | 2.24     | 0.09      | 25.49    | 9.40  | [7.91, 11.17]  | <b>&lt;.001</b> |
| L1 > No              | 2.29     | 0.09      | 25.57    | 9.89  | [8.30, 11.79]  | <b>&lt;.001</b> |
| <b>Skip rate</b>     |          |           |          |       |                |                 |
| Captions – Bilingual | -0.57    | 0.35      | -1.62    | 0.57  | [0.28, 1.13]   | .37             |
| L1 – Bilingual       | 0.04     | 0.36      | 0.95     | 1.40  | [0.70, 2.81]   | .78             |
| Bilingual > No       | -6.26    | 0.35      | -17.82   | 0.002 | [0.001, 0.004] | <b>&lt;.001</b> |
| L1 – Captions        | 0.90     | 0.37      | 2.45     | 2.47  | [1.20, 5.10]   | .07             |

| Group         | <i>b</i> | <i>SE</i> | <i>z</i> | OR    | OR 95% CI      | <i>p</i> |
|---------------|----------|-----------|----------|-------|----------------|----------|
| Captions > No | -6.83    | 0.37      | -18.65   | 0.001 | [0.001, 0.002] | <.001    |
| L1 > No       | -5.92    | 0.37      | -16.04   | 0.003 | [0.001, 0.006] | <.001    |

In sum, based on Level 1 analyses, it can be concluded that, as expected, the no subtitles group was significantly different from other groups as significantly less time was spent on the subtitling area. This indicated that the time spent on the subtitling area in the three groups with on-screen text indeed reflected participants' reading of the on-screen text rather than the processing of covered images. Regarding the groups with on-screen text, the bilingual subtitles group spent similar amount of time processing the subtitling area as the captions group, and they both spent significantly more time than the L1 subtitles group, as revealed by total reading time % and fixation %. However, no differences were reported in terms of run count or skip rate among the three groups with on-screen text. For the analyses of Level 2, participants' attention distribution on two subtitling lines when using bilingual subtitles was further explored.

#### 4.2.3.2. Level 2: Subtitling Lines

The aim of the Level 2 analysis was twofold: 1) to explore the reading of L1 and L2 lines within the bilingual subtitles group; and 2) to compare the reading of L1 and L2 lines in the bilingual subtitles group to the monolingual groups (i.e., captions and L1 subtitles groups) separately.

Table 30 shows the descriptive statistics for the reading of L1 and L2 lines within the bilingual subtitles group. The findings indicated that when using bilingual subtitles, participants tended to spend less time reading L2 lines than L1 lines. This difference was further confirmed by total reading time % ( $b = -0.16$ ,  $t(534) = -33.45$ ,  $p < .001$ ,  $d =$

0.82) and fixation % ( $b = -0.17$ ,  $t(534) = -34.47$ ,  $p < .001$ ,  $d = 0.87$ ) using linear mixed-effects models (see Appendix S12 for model summaries), with small effect sizes. In addition, when using bilingual subtitles, the frequency of entering in L2 lines was significantly less than in L1 lines as revealed by run count ( $OR = 0.59$ , 95% CI = [0.58, 0.61],  $p < .001$ ), and L2 lines were skipped more often than L1 lines ( $OR = 9.69$ , 95% CI = [8.73, 10.75],  $p < .001$ ), as reported in generalised mixed-effects models (see Appendix S12 for model summaries). Participants' vocabulary size did not significantly contribute to the difference in reading L1 and L2 lines when using bilingual subtitles.

Table 30. Descriptive Statistics for Eye-Movement Data at Level 2 for L1 and L2 Line Area within the Bilingual Subtitles Group

| Bilingual group | Total time %              |              | Fixation %                              |              | Run count                               |              | Skip rate                               |              |
|-----------------|---------------------------|--------------|-----------------------------------------|--------------|-----------------------------------------|--------------|-----------------------------------------|--------------|
|                 | <i>M</i><br>( <i>SD</i> ) | 95%<br>CI    | <i>M</i> ( <i>SD</i> )<br>( <i>SD</i> ) | 95%<br>CI    | <i>M</i> ( <i>SD</i> )<br>( <i>SD</i> ) | 95%<br>CI    | <i>M</i> ( <i>SD</i> )<br>( <i>SD</i> ) | 95%<br>CI    |
| L1 lines        | 0.42<br>(0.28)            | [0.41, 0.42] | 0.44<br>(0.26)                          | [0.43, 0.44] | 1.25<br>(0.70)                          | [1.24, 1.26] | 0.11<br>(0.31)                          | [0.10, 0.11] |
| L2 lines        | 0.20<br>(0.24)            | [0.20, 0.21] | 0.22<br>(0.24)                          | [0.21, 0.22] | 0.75<br>(0.77)                          | [0.74, 0.76] | 0.43<br>(0.50)                          | [0.42, 0.43] |

However, the above results should be treated with caution because we are comparing the reading behaviour of two different languages. More appropriate comparisons should be made for different lines of the bilingual subtitles compared to their corresponding lines in monolingual subtitle groups.

To compare the bilingual subtitles with monolingual subtitles, the reading of L2 (English) lines in the bilingual subtitles and captions group was first compared. As observed in Table 31, the descriptive data showed that participants using bilingual subtitles tended to spend less time on reading L2 lines than participants using captions, and they had higher chance of skipping L2 lines than the captions group. This group

difference was further confirmed by four sets of mixed-effects models (see Appendix S13 for model summaries). For the captions group, the total reading time % ( $b = 0.39$ ,  $t(49) = 6.98$ ,  $p < .001$ ,  $d = 1.24$ ) and fixation % ( $b = 0.39$ ,  $t(49) = 7.15$ ,  $p < .001$ ,  $d = 1.36$ ) on L2 lines was significantly higher than that in the bilingual subtitles group. Moreover, participants' vocabulary size negatively predicted their processing of L2 lines but was only significant for the captions group, as reported by both measures. Similarly, the odds of entering L2 lines in the captions group was 1.65 times higher than those in the bilingual subtitles group ( $OR = 1.65$ , 95% CI = [1.38, 1.96],  $p < .001$ ), and the odds of skipping L2 lines in the captions group decreased significantly compared to the bilingual subtitles group ( $OR = 0.01$ , 95% CI = [0.003, 0.021],  $p < .001$ ). Participants' vocabulary size revealed a significant positive effect on skipping rate in the captions group but no effects on run count (see Appendix S13 for model summaries).

Table 31. Descriptive Statistics for Eye-Movement Data at Level 2 for L2 Line Area in the Bilingual Subtitles and Captions Groups

| Group              | Total time %   |              | Fixation %     |              | Run count      |              | Skip rate      |              |
|--------------------|----------------|--------------|----------------|--------------|----------------|--------------|----------------|--------------|
|                    | <i>M</i>       | 95%          | <i>M</i>       | 95%          | <i>M</i>       | 95%          | <i>M</i>       | 95%          |
|                    | ( <i>SD</i> )  | CI           | ( <i>SD</i> )  | CI           | ( <i>SD</i> )  | CI           | ( <i>SD</i> )  | CI           |
| Bilingual L2 lines | 0.20<br>(0.24) | [0.20, 0.21] | 0.22<br>(0.24) | [0.21, 0.22] | 0.75<br>(0.77) | [0.74, 0.76] | 0.43<br>(0.50) | [0.42, 0.43] |
| Captions           | 0.56<br>(0.32) | [0.55, 0.56] | 0.59<br>(0.31) | [0.58, 0.59] | 1.15<br>(0.64) | [1.14, 1.16] | 0.11<br>(0.31) | [0.10, 0.11] |

Then, participants' eye movements on L1 (Chinese) lines in the bilingual subtitles group were compared to the L1 subtitles group. The descriptive data reported in Table 32 seems to show that the bilingual subtitles group spent more time on L1 lines and skipped less than the L1 subtitles group. The group differences were confirmed by mixed-effects models (see Appendix S14 for model summaries). Bilingual subtitle users

spent significantly more time on L1 lines than the L1 subtitles group, as revealed by total reading time % ( $b = 0.06$ ,  $t(50) = 2.65$ ,  $p < .001$ ,  $d = 0.28$ ), with a very small effect size. However, no group difference was revealed in terms of fixation % ( $\chi^2(1) = 2.48$ ,  $p = .12$ ,  $R^2 < .001$ ), showing that number of fixations made on L1 lines was similar between the bilingual and L1 subtitles groups. In terms of run count, participants using bilingual subtitles entered L1 lines area 1.14 times more frequently than the L1 subtitles group ( $OR = 1.14$ , 95% CI = [1.03, 1.26],  $p = .01$ ), and their odds of skipping the L1 lines was 0.49 times less compared to the L1 subtitles group ( $OR = 0.49$ , 95% CI = [0.26, 0.84],  $p = .01$ ). Participants' vocabulary size did not have significant effects on participants' use of L1 lines.

Table 32. Descriptive Statistics for Eye-Movement Data at Level 2 for L1 Line Area in the Bilingual Subtitles and L1 Subtitles Groups

| Group              | Total time %   |              | Fixation %     |              | Run count      |              | Skip rate      |              |
|--------------------|----------------|--------------|----------------|--------------|----------------|--------------|----------------|--------------|
|                    | <i>M</i>       | 95%          | <i>M</i>       | 95%          | <i>M</i>       | 95%          | <i>M</i>       | 95%          |
|                    | ( <i>SD</i> )  | CI           | ( <i>SD</i> )  | CI           | ( <i>SD</i> )  | CI           | ( <i>SD</i> )  | CI           |
| Bilingual L1 lines | 0.42<br>(0.28) | [0.41, 0.42] | 0.44<br>(0.26) | [0.43, 0.44] | 1.25<br>(0.70) | [1.24, 1.26] | 0.11<br>(0.31) | [0.10, 0.11] |
| L1 subtitles       | 0.34<br>(0.27) | [0.34, 0.35] | 0.39<br>(0.27) | [0.39, 0.40] | 1.10<br>(0.71) | [1.09, 1.11] | 0.18<br>(0.38) | [0.17, 0.18] |

Based on the findings at Level 2, it can be concluded that in general, participants using bilingual subtitles seemed to spend more time reading L1 (Chinese) lines than L2 (English) lines, as revealed by within group comparisons with four eye-tracking measures. This was further confirmed by conducting between group comparisons. Participants using bilingual subtitles spent significantly less time reading L2 lines than the caption users as shown by all eye-tracking measures. On the contrary, participants using bilingual subtitles seemed to spend more time on L1 lines than participants using

L1 subtitles. However, it should be noted that this difference failed to reach a significant level in fixation %. In addition, participants with a larger vocabulary size tended to spend less time processing captions, while vocabulary size did not significantly affect the processing of on-screen text when using the bilingual or L1 subtitles. In the following analyses, participants' processing of individual unknown TWs and their corresponding L1 translations were examined.

#### 4.2.3.3. Level 3: Target Word Area

Level 3 focused on participants' processing of unknown TWs, therefore, for each participant, only the unknown TWs (scored 0 on the form recognition pretest) were included in following analyses. Seven eye-tracking measures (i.e., first-pass reading time, first fixation duration, total reading time, fixation count, second-pass reading time, second fixation duration, skip rate) were used to investigate participants' processing of unknown TWs and their corresponding L1 translations in different subtitling conditions. The correlations among the seven measures were also examined with Pearson correlation, as suggested by Godfroid and Hui (2020). As can be observed from Table 33, apart from skip rate, different measures all had strong and positive correlations with each other (all  $ps < .01$ ).

Table 33. Correlation Matrix of Seven Eye-Tracking Measures

|                       | Total<br>reading | First-pass<br>reading | First<br>fixation<br>duration | Second-pass<br>reading | Second<br>fixation<br>duration | Fixation<br>count | Skip<br>rate |
|-----------------------|------------------|-----------------------|-------------------------------|------------------------|--------------------------------|-------------------|--------------|
| Total<br>reading      | <b>1</b>         | <b>.83</b>            | <b>.73</b>                    | <b>.66</b>             | <b>.78</b>                     | <b>.91</b>        | <b>-.55</b>  |
| First-pass<br>reading |                  | <b>1</b>              | <b>.83</b>                    | <b>.21</b>             | <b>.65</b>                     | <b>.73</b>        | <b>-.52</b>  |

|                                | Total<br>reading | First-pass<br>reading | First<br>fixation<br>duration | Second-pass<br>reading | Second<br>fixation<br>duration | Fixation<br>count | Skip<br>rate |
|--------------------------------|------------------|-----------------------|-------------------------------|------------------------|--------------------------------|-------------------|--------------|
| First<br>fixation<br>duration  |                  |                       | <b>1</b>                      | <b>.25</b>             | <b>.45</b>                     | <b>.61</b>        | <b>-.58</b>  |
| Second-pass<br>reading         |                  |                       |                               | <b>1</b>               | <b>.59</b>                     | <b>.63</b>        | <b>.30</b>   |
| Second<br>fixation<br>duration |                  |                       |                               |                        | <b>1</b>                       | <b>.70</b>        | <b>-.40</b>  |
| Fixation<br>count              |                  |                       |                               |                        |                                | <b>1</b>          | <b>-.60</b>  |
| Skip rate                      |                  |                       |                               |                        |                                |                   | <b>1</b>     |

According to the descriptive statistics for seven eye-tracking measures by group, as summarised in Table 34, when using bilingual subtitles, participants seemed to spend more time processing L1 translations of unknown TWs than their L2 forms. By comparing the bilingual subtitles with monolingual groups, less time was spent on L2 unknown TWs in bilingual subtitles than captions, but more time was spent on L1 translations in bilingual subtitles than in L1 subtitles. Twenty-one sets of mixed-effects models were constructed to obtain statistical evidence for these differences. Similar to the analyses at Level 2, comparisons were first made within the bilingual subtitles group, followed by between group comparisons.

Table 34. Descriptive Statistics for the Seven Eye-Movement Measures on Unknown Target Words and Corresponding L1 Translations by Group

|                               | Bilingual L1  |              | Bilingual L2  |              | Captions      |               | L1 Subtitles  |              |
|-------------------------------|---------------|--------------|---------------|--------------|---------------|---------------|---------------|--------------|
|                               | <i>M (SD)</i> | 95% CI       | <i>M (SD)</i> | 95% CI       | <i>M (SD)</i> | 95% CI        | <i>M (SD)</i> | 95% CI       |
| Total reading time (ms)       | 355 (358)     | [324, 387]   | 236 (347)     | [206, 266]   | 546 (480)     | [502, 591]    | 317 (398)     | [278, 356]   |
| First-pass reading time (ms)  | 251 (247)     | [230, 273]   | 179 (273)     | [155, 202]   | 352 (368)     | [318, 387]    | 235 (294)     | [206, 263]   |
| First fixation duration (ms)  | 181 (168)     | [167, 196]   | 122 (167)     | [108, 137]   | 235 (219)     | [215, 255.81] | 177 (205)     | [157, 197]   |
| Second-pass reading time (ms) | 86 (164)      | [72, 100]    | 48 (146)      | [36, 61]     | 153 (233)     | [131, 175]    | 59 (141)      | [46, 73]     |
| Second fixation duration (ms) | 103 (145)     | [91, 116]    | 667 (148)     | [54, 80]     | 154 (176)     | [137, 170]    | 81 (155)      | [66, 97]     |
| Fixation count                | 1.94 (1.79)   | [1.79, 2.10] | 1.17 (1.60)   | [1.03, 1.31] | 2.65 (2.15)   | [2.45, 2.85]  | 1.56 (1.72)   | [1.39, 1.73] |
| Skip rate                     | 0.19 (0.40)   | [0.16, 0.23] | 0.48 (0.50)   | [0.44, 0.53] | 0.15 (0.36)   | [0.12, 0.18]  | 0.29 (0.45)   | [0.24, 0.33] |

First, the processing of L2 TWs that were unknown in pretests and their corresponding L1 translations in the bilingual subtitles group was examined. Seven sets of mixed-effects models were constructed for seven eye-movement measures controlling for the presentation time and size of each AOI (see Appendix S15 for model summaries). Within the bilingual subtitles group, the time spent on the L2 unknown TWs was significantly shorter compared to that on their L1 translations, as revealed by total reading time ( $b = -1.84, t(24) = -5.84, p < .001, d = 0.34$ ), first-pass reading time ( $b = -1.69, t(24) = -5.76, p < .001, d = 0.28$ ), first fixation duration ( $b = -1.40, t(23) = -4.91, p < .001, d = 0.35$ ), second-pass reading time ( $b = -1.01, t(26) = -3.81, p < .001, d = 0.24$ ), second fixation duration ( $b = -1.22, t(26) = -3.93, p < .001, d = 0.25$ ), and fixation count ( $OR = 0.46, 95\% CI = [0.33, 0.65], p < .001$ ). Also, the skip rate of L2 unknown TWs was significantly higher than that of L1 translations ( $OR = 6.17, 95\% CI = [3.23, 11.79], p < .001$ ). Moreover, the frequency of occurrence of TWs had a significant positive effect on participants' reading of TWs as well as their corresponding L1 translations in the bilingual subtitles group, as revealed by most eye tracking measures except for second-pass reading time, second fixation duration, and skip rate. This indicated that unknown words that appeared more frequently in the video had a better chance of attracting participants' general and early attention.

The reading of unknown TWs in the captions and bilingual subtitles groups was then compared. Results of mixed-effects models showed that the captions group spent significantly longer time on the unknown L2 TWs than the bilingual subtitles group, as revealed by total reading time ( $b = 2.18, t(55) = 6.50, p < .001, d = 0.75$ ), first-pass reading time ( $b = 1.87, t(53) = 5.92, p < .001, d = 0.54$ ), first fixation duration ( $b = 1.75, t(53) = 5.95, p < .001, d = 0.59$ ), second-pass reading time ( $b = 1.71, t(51) = 6.27, p < .001, d = 0.55$ ), second fixation duration ( $b = 1.81, t(51) = 5.75, p < .001, d = 0.54$ ),

and fixation count ( $OR = 2.70$ , 95% CI = [1.98, 3.82],  $p < .001$ ). The skip rate in the captions group was significantly lower than that in the bilingual subtitles group ( $OR = 0.08$ , 95% CI = [0.04, 0.17],  $p < .001$ ). Participants' vocabulary size was found to have a significant negative effect on the second-pass reading time on unknown TWs, indicating that participants with a smaller vocabulary size tended to spend more time rereading the L2 unknown TWs in both groups. As expected, the presentation time of TWs was also found to be a significant predictor of learners' processing in all eye-tracking measures, indicating that an unknown word was processed longer with longer presentation time in the video. Longer words were also more likely to be reread when using bilingual subtitles and captions (see Appendix S16 for model summaries).

The processing of L1 translations of unknown TWs in bilingual and L1 subtitles groups was then compared. Mixed-effects models showed that participants using bilingual subtitles spent significantly more time on L1 translations of unknown TWs than the L1 subtitles group, as revealed by total reading time ( $b = 0.63$ ,  $t(910) = 2.10$ ,  $p = .04$ ,  $d = 0.12$ ), first-pass reading time ( $b = 0.56$ ,  $t(910) = 1.99$ ,  $p = .05$ ,  $d = 0.07$ ), second-pass reading time ( $b = 0.40$ ,  $t(910) = 2.01$ ,  $p = .05$ ,  $d = 0.17$ ), second fixation duration ( $b = 0.58$ ,  $t(910) = 2.50$ ,  $p = .02$ ,  $d = 0.15$ ), and fixation count ( $OR = 0.81$ , 95% CI = [0.65, 1.01],  $p = .05$ ). The skip rate in the bilingual subtitles group was also significantly lower than that in the L1 subtitles group ( $OR = 0.51$ , 95% CI = [1.03, 3.77],  $p = .04$ ). However, no significant difference was reported in term of first fixation duration ( $\chi^2(1) = 3.33$ ,  $p = .07$ ,  $R^2 = .001$ ). The presentation time and the length of L1 translations were found to be significant predictors of learners' processing time in most eye-tracking measures. However, participants' vocabulary size did not affect their processing of the L1 translations (see Appendix S17 for model summaries).

In sum, Level 3 analyses showed that, when unknown words appeared in the video, participants using bilingual subtitles tended to spend more time on corresponding L1 translations of TWs than L2 TWs. Between-group comparisons also revealed that the bilingual subtitles group spent significantly less time on the L2 unknown TWs than the captions group, whereas they spent significantly more time to process the L1 translations of unknown words than the L1 subtitles group.

#### ***4.2.4. RQ4 Relationship Between Eye Movements and Vocabulary Tests***

To address RQ4, 36 sets of logistic mixed-effects models were conducted by group (captions, L1 subtitles, and bilingual subtitles), with L1 and L2 AOIs separately, to explore the potential relationship between the reading of unknown TWs (as measured by total reading time, first-pass reading time, and second-pass reading time, following Montero Perez et al., 2015) and vocabulary gains (as measured by form recognition, meaning recall, and meaning recognition posttests scores). Following the same procedure as in RQ3, only the unknown words (as indicated by the form recognition pretest) for each participant were included in analyses.

The main findings of 36 logistic mixed-effects models are summarised in Table 35. Detailed model summaries are summarised in Appendix S18 to S20. Results in Table 35 show that, for the bilingual subtitles group, total reading time and first-pass reading time on L2 unknown TWs significantly predicted form recognition gains with a large effect size. This indicated that a one-second increase in total time and first-pass reading time spent on an unknown L2 TW increased the odds of form recognition success by 3.01 and 5.45 times. Similarly, meaning recall scores were significantly predicted by total reading time and first-pass reading time on L2 TWs, with a one-second increase in reading led to 3.09 and 3.38 times higher odds of gains, respectively. However, none of

the measures predicted the meaning recognition gains. Second-pass reading time on the unknown L2 TWs was not a significant predictor of vocabulary gains in the bilingual subtitles group.

Table 35. Results of the Logistic Mixed-Effects Models Examining Relationship Between Processing of Unknown Target Words and Vocabulary Gains by Group

|                               | Bilingual – L2 unknown TWs |           |      |                 |                 | Bilingual – L1 translations |           |      |                 |          | Captions – L2 unknown TWs |           |      |              |            | L1 Subtitles – L1 translations |           |      |                 |          |
|-------------------------------|----------------------------|-----------|------|-----------------|-----------------|-----------------------------|-----------|------|-----------------|----------|---------------------------|-----------|------|--------------|------------|--------------------------------|-----------|------|-----------------|----------|
|                               | <i>b</i>                   | <i>SE</i> | OR   | OR<br>95%<br>CI | <i>p</i>        | <i>b</i>                    | <i>SE</i> | OR   | OR<br>95%<br>CI | <i>p</i> | <i>b</i>                  | <i>SE</i> | OR   | 95%<br>CI    | <i>p</i>   | <i>b</i>                       | <i>SE</i> | OR   | OR<br>95%<br>CI | <i>p</i> |
| <b>Form Recognition</b>       |                            |           |      |                 |                 |                             |           |      |                 |          |                           |           |      |              |            |                                |           |      |                 |          |
| 1 <sup>st</sup> -pass reading | 1.70                       | 0.43      | 5.45 | [2.42, 13.14]   | <b>&lt;.001</b> | 0.16                        | 0.45      | 1.17 | [0.48, 2.82]    | .72      | 0.90                      | 0.39      | 2.45 | [1.14, 5.47] | <b>.02</b> | -0.002                         | 0.43      | 1.00 | [0.42, 2.36]    | 1.00     |
| 2 <sup>nd</sup> -pass reading | 0.39                       | 0.70      | 1.48 | [0.36, 5.85]    | .57             | -0.79                       | 0.65      | 0.46 | [0.12, 1.62]    | .23      | 0.08                      | 0.55      | 1.08 | [0.36, 3.19] | .88        | -1.82                          | 0.96      | 0.16 | [0.02, 0.98]    | .06      |
| Total time                    | 1.10                       | 0.32      | 3.01 | [1.63, 5.76]    | <b>&lt;.001</b> | -0.29                       | 0.32      | 0.75 | [0.39, 1.39]    | .36      | 0.49                      | 0.31      | 1.63 | [0.88, 3.04] | .11        | -0.44                          | 0.34      | 0.65 | [0.32, 1.25]    | .20      |
| <b>Meaning Recall</b>         |                            |           |      |                 |                 |                             |           |      |                 |          |                           |           |      |              |            |                                |           |      |                 |          |
| 1 <sup>st</sup> -pass reading | 1.22                       | 0.49      | 3.38 | [1.29, 9.26]    | <b>.01</b>      | 0.64                        | 0.54      | 1.90 | [0.81, 2.13]    | .23      | 0.54                      | 0.59      | 1.72 | [0.50, 5.35] | .36        | 0.45                           | 0.69      | 1.57 | [0.36, 5.77]    | .51      |
| 2 <sup>nd</sup> -pass reading | 0.99                       | 0.89      | 2.68 | [0.39, 13.84]   | .27             | -0.76                       | 1.02      | 0.47 | [0.05, 2.91]    | .46      | 0.40                      | 1.02      | 1.50 | [0.17, 9.93] | .69        | -0.32                          | 1.72      | 0.72 | [0.01, 13.27]   | .85      |
| Total time                    | 1.13                       | 0.39      | 3.09 | [1.43, 6.89]    | <b>.004</b>     | 0.16                        | 0.39      | 1.18 | [0.49, 2.48]    | .68      | 0.42                      | 0.50      | 1.51 | [0.52, 3.93] | .41        | 0.18                           | 0.56      | 1.19 | [0.36, 3.34]    | .75      |
| <b>Meaning Recognition</b>    |                            |           |      |                 |                 |                             |           |      |                 |          |                           |           |      |              |            |                                |           |      |                 |          |
| 1 <sup>st</sup> -pass reading | 0.65                       | 0.45      | 1.91 | [0.80, 4.82]    | .15             | 0.70                        | 0.54      | 2.01 | [0.70, 5.84]    | .19      | 0.79                      | 0.41      | 2.21 | [0.99, 5.01] | <b>.05</b> | 0.02                           | 0.46      | 1.02 | [0.41, 2.53]    | .96      |
| 2 <sup>nd</sup> -pass reading | 0.59                       | 0.77      | 1.81 | [0.37, 8.20]    | .44             | 0.46                        | 0.71      | 1.58 | [0.39, 6.60]    | .52      | 0.42                      | 0.61      | 1.53 | [0.44, 5.08] | .49        | 0.89                           | 0.94      | 2.43 | [0.39, 16.52]   | .34      |
| Total time                    | 0.50                       | 0.35      | 1.65 | [0.83, 3.37]    | .16             | 0.55                        | 0.38      | 1.74 | [0.82, 3.74]    | .15      | 0.68                      | 0.33      | 1.97 | [1.03, 3.83] | <b>.04</b> | 0.17                           | 0.35      | 1.18 | [0.59, 2.37]    | .64      |

As Table 35 shows, for the captions group, form recognition scores were significantly predicted by the first-pass reading time on L2 TWs. This indicates that with a one-second increase of first-pass reading time, the odds of correctly recognising the form of each unknown TW increased 2.45 times. The time spent on L2 TWs did not significantly relate to meaning recall scores. Meaning recognition results pointed to a positive effect of total reading time and first-pass reading time on vocabulary scores, suggesting 1.97 and 2.21 times higher odds of meaning recognition success with one-second increase in reading. Similar to the bilingual subtitles group, second-pass reading time was not a significant predictor of any vocabulary scores.

Regarding the time spent on L1 translations of TWs, Table 35 indicates that none of the eye-tracking measures on L1 translations showed significant effects on any type of vocabulary test in both bilingual and L1 subtitles groups. This indicates that in general, the processing time spent on the L1 translation of an unknown TWs did not increase the chance to learn vocabulary irrespective of the subtitling type. However, when using L1 subtitles, the second-pass reading time on L1 translations of TWs demonstrated an approaching significance negative effect on form recognition gains ( $p = .06$ ). This would suggest that the more rereading of L1 translations during L1 subtitled viewing, the less likely it was that participants noticed the form of unknown TWs.

In sum, for the reading of L2 unknown TWs, both total reading time and first-pass reading time were to some extent predictive of word learning, whereas second-pass reading time was not significantly related to vocabulary gains. The reading of L1 translations of unknown TWs did not predict vocabulary gains in any of the groups and measurements.

### **4.3. QUAN Interim Discussion**

#### **4.3.1. RQ1 – Vocabulary Tests**

The first research question concerns the relative effects of bilingual subtitles on incidental vocabulary learning compared to captions, L1 subtitles, and no subtitles. Three components of lexical mastery were examined, i.e., form recognition (as measured by both form recognition and occurrence decision tests), meaning recall, and meaning recognition. Overall, results showed that participants in all subtitling conditions learned vocabulary, further supporting the effectiveness of viewing for vocabulary learning. In line with previous research, form recognition was the easiest component to acquire, followed by word meaning (e.g., Mohamed, 2017; Pellicer-Sánchez, 2016; Peters et al., 2016). Moreover, meaning recall gains were moderate in all subtitling conditions (e.g., Li, 2016; Montero Perez et al., 2015; Peters, 2019; Peters et al., 2016; Pujadas & Muñoz, 2019), and learning gains were higher in meaning recognition than in meaning recall, supporting earlier research findings (e.g., Peters et al., 2016; Peters & Webb, 2018).

The present research further supports the claim that captions and bilingual subtitles are beneficial for intermediate and advanced L2 learners' incidental vocabulary learning (Danan, 2004; Li, 2016; Montero Perez et al., 2013). However, L1 subtitles did not show significant advantages over captions or bilingual subtitles, which is consistent with previous findings (e.g., Dizon & Thanyawatpokin, 2021; Li, 2016; Peters et al., 2016; Pujadas & Muñoz, 2019). This finding suggests that the benefits of L1 subtitles might be limited among higher-level L2 learners compared to young and less skilled learners (Danan, 2004; Dizon & Thanyawatpokin, 2021; Lwo & Lin, 2012).

Regarding form recognition, the results from both form recognition and occurrence

decision tests demonstrated a general advantage of captions over L1 and no subtitles conditions, in line with previous studies (e.g., Peters, 2019; Peters et al., 2016). The benefits of captions for form recognition could be attributed to their positive role in segmenting speech, making unfamiliar words more salient and noticeable, and helping to establish a link between the auditory and written forms of words (Bisson et al., 2014; Peters, 2019; Winke et al., 2010). This study is one of the first to investigate the relative effectiveness of bilingual subtitles for facilitating formal knowledge. Results showed a disadvantage of bilingual subtitles compared to captions in learning word forms. This indicates that having the L2 written form of unknown vocabulary supported learners' learning of the L2 form of novel words, but the presentation of L1 might compromise the acquisition of word forms. The present findings seem to support the claim that the use of L1 subtitles can increase learners' dependence on L1, distracting learners' attention from the L2 audio input, and limiting learners' learning of L2 forms (Danan, 1992; Kuppens, 2010; Peters, 2019).

Regarding the acquisition of meaning, bilingual subtitles showed an advantage over no subtitles in both meaning tests, supporting the findings of Li's (2016) study. However, this finding disagrees with other studies that either found no significant difference between bilingual subtitles and no subtitles (e.g., Hao et al., 2021; Lwo & Lin, 2012) or group differences were only significant for some participants but not for others (e.g., Y. Wang, 2019). As mentioned in section 2.3.3.2, some of the findings from previous bilingual subtitles research should be treated with caution. The lack of group differences can be caused by the unsuitability of viewing materials (e.g., Hao et al., 2021), an unnatural viewing process (e.g., Lwo & Lin, 2012), or participants' unequal prior knowledge of TWs in different subtitling conditions (e.g., Y. Wang, 2019). These limitations could have influenced participants' viewing process and affected learning

outcomes. In the present research, bilingual subtitles were also significantly more beneficial than captions for meaning recognition but not for meaning recall. This finding might be due to the fact that different test constructs reflect different dimensions of word knowledge (Laufer & Aviad-Levitzky, 2017; Read, 2000). Meaning recognition tests examine the initial stages of vocabulary learning (Schmitt, 2010), while meaning recall tests reflect deeper vocabulary knowledge. Meaning recall tests do not take into account partial knowledge and require better memory traces than recognition (Laufer & Goldstein, 2004). Bilingual subtitles can help in establishing the initial form-meaning link by providing L2 forms and L1 translations, which can be detected via recognition tests. However, according to the Depth of Processing Theory ( Craik & Lockhart, 1972), the given translations may have diminished learners' cognitive analysis and their attempts to infer the meanings of unknown words, leading to the formation of shallower memory traces which were not enough to develop the ability to recall the meanings of newly learned words that was superior to the captions group.

Partially in line with Li's (2016) findings, bilingual subtitles also outperformed L1 subtitles in meaning recall. This benefit could be attributed to the presentation of L2 TWs, which could draw the learners' attention to unknown word forms (Winke et al., 2010), reduce the chance of bypassing the spoken form of unknown words, leading to a clearer opportunity to establish a form-meaning connection (Li, 2016). However, the superiority of bilingual subtitles over L1 subtitles in meaning recognition was only approaching significance, which is different from the significant result reported by Lazareva and Loerts (2017). These inconsistent results could be explained by the different research designs. Participants in the study by Lazareva and Loerts (2017) had no prior knowledge of the L2, therefore it might have been more challenging for them to match auditory L2 forms with L1 translations during L1 subtitled viewing without the

help of L2 written forms. Moreover, the target items measured in their test included single words and short phrases which might even have increased this difficulty in matching.

In terms of the factors influencing vocabulary learning gains, the present findings support Nation's (2013) claim that L2 learners with a larger vocabulary size can achieve better comprehension and process more L2 input, which can further enlarge their vocabulary size. In this study, participants with a larger vocabulary size achieved greater learning gains as revealed in all tests, echoing previous viewing studies (e.g., Peters, 2019; Peters et al., 2016). In addition, word length was also found to positively predict participants' meaning recall scores, with the meanings of longer unknown words having a better chance of being recalled successfully. This finding contradicts the reading research by Godfroid et al. (2018), which revealed a negative predictive role for word length in incidental vocabulary learning. This inconsistency can be explained by the real-time nature of subtitled viewing, where longer words in the video might be more salient and more likely to attract more attention (Montero Perez et al., 2015; Puimège & Peters, 2019). Moreover, the meanings of words that occurred more frequently in the video were more likely to be recognised, supporting previous findings (e.g., Li, 2016; Peters, 2019; Peters et al., 2016; Peters & Webb, 2018; Uchiyama et al., 2019).

In sum, bilingual subtitles tended to be more effective for facilitating meaning knowledge, but not as effective as captions for learning word forms. Although the presence of L1 in bilingual subtitles seems to compromise learners' learning of L2 word forms, by presenting both L2 and L1, the form-meaning link is more likely to be established. Contrary to the redundancy principle (Sweller, 2005b), the seemingly

redundant information in bilingual subtitles did not have a detrimental effect on learning but rather appeared to support vocabulary learning.

#### **4.3.2. RQ2 – Comprehension**

The second research question examined the relative effects of bilingual subtitles on comprehension, compared to captions, L1, and no subtitles. The results showed that bilingual subtitles were as helpful as L1 subtitles for facilitating comprehension, and both of them were significantly better than captions and no subtitles, which concurs with previous findings (e.g., Bianchi & Ciabattini, 2008; Birulés-Muntané & Soto-Faraco, 2016; Dizon & Thanyawatpokin, 2021; Hao et al., 2021; Lwo & Lin, 2012; Markham & Peter, 2003; Markham et al., 2001; Pujadas & Muñoz, 2019, 2020; Y. Wang, 2019). This seems logical since the use of L1 could facilitate understanding. Additionally, in line with the study by Pujadas and Muñoz (2020), participants with a larger vocabulary size also had a better understanding of the video, regardless of the subtitling condition.

This study added empirical evidence to show that the presentation of L1 (i.e., using bilingual subtitles and L1 subtitles) could facilitate L2 learners' comprehension, as measured by multiple-choice comprehension questions conducted in the participants' L1. Moreover, the presentation of both L1 and L2 did not hinder learners' comprehension of the video but was in fact beneficial. The findings also support Paivio's (1986) bilingual version of the Dual Coding Theory by showing that activation of the imagery system and two verbal systems could augment learners' memory. It is arguable, however, whether the benefits of bilingual subtitles were only due to the presentation of L1. In other words, is it the case that participants who used bilingual

subtitles only made use of L1 lines, which contributed to their comprehension in a similar way to the use of L1 subtitles only? The following section discusses participants' attention allocation during their subtitled viewing.

#### **4.3.3. RQ3 – Eye-Tracking data**

The third research question examined the processing patterns of the subtitling area and TWs when using bilingual subtitles, as compared to the other subtitling conditions. Participants' eye-tracking data are discussed at three levels of analysis: the overall subtitling area (Level 1), each subtitle line (Level 2), and unknown TWs (Level 3).

##### **4.3.3.1. Level 1: Overall Subtitling Area**

The eye-movement data demonstrated that the results for the three groups with on-screen text were significantly different from the no subtitles group, as reflected in all four eye-tracking measures (i.e., total reading time %, fixation %, run count, and skip rate). This indicates that the eye movements scanning the subtitling area were indeed caused by participants' reading of the subtitling rather than the actions or images displayed in the subtitling area. This finding also supports previous research showing that viewers process on-screen text regardless of the subtitling type and learners' knowledge of the languages used in subtitles (e.g., Bisson et al., 2014; d'Ydewalle et al., 1991; Kruger et al., 2014; Liao et al., 2020). The total reading time % on the subtitling area in the captions (64%) and L1 subtitles (44%) groups were also in line with previous studies conducted on experienced L2 learners (e.g., d'Ydewalle & De Bruycker, 2007; Gass et al., 2019; Winke et al., 2013).

In addition, the descriptive data support the claim that the reading of subtitles does not prohibit viewers' processing of images (Bisson et al., 2014; Perego et al., 2010). Concerning the distribution of attention between images and subtitling areas, the captions group spent more time on the subtitling area than the images, which is congruent with some previous research (e.g., Gass et al., 2019; Winke et al., 2013) but not with other studies (e.g., Bisson et al., 2014; Kruger et al., 2014). However, it should be noted that participants in the study by Bisson et al. (2014) did not have prior knowledge of the L2, which might have resulted in less time being spent on reading captions (in L2). Participants in the study by Kruger et al. (2014), however, were used to attending lectures in English (as their L2), which might explain their lower reliance on captions when watching a video lecture in English. In the present research, the L1 subtitles group spent slightly more time on images than the subtitling area, which also concurs with preceding studies conducted on experienced L2 learners (e.g., d'Ydewalle & De Bruycker, 2007; Kruger et al., 2014), but not studies conducted on participants without prior knowledge of the L2s (e.g., Bisson et al., 2014; Perego et al., 2010). It is reasonable that viewers who did not understand the audio soundtrack would rely more on L1 subtitles to aid their comprehension.

Comparing the three groups with on-screen text, the results were in line with Liao et al.'s (2020) findings showing that the amount of time spent on the subtitling area in the bilingual subtitles group was similar to the captions group, but significantly more than in the L1 subtitles group. However, the present findings reveal that the bilingual subtitles group spent about 60% of total reading time on the overall subtitling area while 39% of their time was spent on images, whereas participants in Liao et al.'s (2020) study spent only 34% of total reading time on the subtitling area but more time on images (64%). This discrepancy could be attributed to the use of different viewing

materials, as all the groups in Liao et al.'s (2020) study spent more time on images (varying from 64% to 73%) than the subtitling area (varying from 22% to 34%). Notably, Liao et al.'s (2020) findings should be interpreted with caution due to the limited sample size ( $N = 16$ ) for their eye-tracking data. Despite the discrepancy concerning specific processing time, the present study also echoes Liao et al. (2020) by showing that when using bilingual subtitles, the presentation of both L1 and L2 lines did not trigger more processing time than using captions. This suggests that participants in the bilingual subtitles did not process all the information provided or they processed the information in a faster way. This was indeed addressed in the Level 2 analysis. Comparing the monolingual subtitles groups, the captions group spent more time than the L1 subtitles group on the subtitling area, which corroborates previous research showing that learners' processing time of L1 subtitles is shorter and faster than in their L2s (e.g., Kruger et al., 2014; Muñoz, 2017).

It should be noted that no significant differences were revealed among the three subtitling conditions in terms of run count and skip rate, suggesting that participants' frequency of referring to and skipping the subtitling area were similar, regardless of the subtitling type. However, the present findings demonstrate a tendency for L1 subtitles more likely being skipped than captions, which is in line with Muñoz' (2017) findings, where L1 subtitles were skipped more than captions by adult learners and higher level L2 learners. These findings suggest that, in general, participants using different subtitles all made use of the on-screen text, but the length of time allocated to the processing of the subtitling area was different across groups.

Participants' vocabulary size was a significant predictor of the processing of subtitles, suggesting that learners with a larger vocabulary size tended to skip more and thus spend less time on the subtitling area. This is in line with previous findings

suggesting that subtitles and captions act like a crutch to facilitate learners' understanding (Danan, 2004), and higher proficiency level learners tend to rely less on on-screen text than lower level learners (Muñoz, 2017). This relationship was explored in the analysis of Level 2 by different subtitling lines.

#### **4.3.3.2. Level 2: Subtitling Line Area**

Level 2 explored participants' processing of L1 and L2 lines when using bilingual subtitles. Comparisons were first made within the bilingual subtitles group, followed by between group comparisons. Although within group comparisons were made between two different language systems, as summarised in section 2.4.1, the reading of Chinese and English share more similarities than differences, and eye movements during reading were more related to linguistic content rather than the form of languages (Rayner et al., 2005; Rayner et al., 2007).

When using bilingual subtitles, participants spent 42% of their total reading time on the L1 lines during the presentation of on-screen text, while this dropped significantly to 20% of their reading on L2 lines. Significant differences were also revealed in other measures. These results provide direct eye-movement evidence to support previous interview findings showing that learners tend to spend more time on L1 and skip more L2 lines during bilingual subtitled viewing (e.g., Li, 2016). The higher skip rate recorded for the L2 lines (0.43) indicates that instead of being read faster, the L2 lines actually tended to be skipped. This could be explained by the fact that participants were asked to watch the video for comprehension and the use of L1 is efficient in aiding learners' understanding. In addition, the L1 lines were always presented above the L2 lines, which might have attracted more attention. Participants'

reliance on L1 lines when using bilingual subtitles also explains their similar performance on the comprehension test to the L1 subtitles group. This finding is different from Liao et al.'s (2020) research where no different processing time was reported within bilingual subtitles. This divergence of research findings could be related to the within-subject design adopted by Liao et al. (2020), in which participants in three out of four groups watched a bilingual video after watching a captioned video. Thus, it could be the case that some participants' use of bilingual subtitles was affected by their prior use of captions, making them more inclined to read the L2 during bilingual subtitled viewing. Also, the video clips in their study were only 5 minutes long for each subtitling type, which might be too short to accurately portray participants' subtitled viewing process.

Between-group comparisons of the processing of L2 lines showed that the bilingual subtitles group spent significantly less time on the L2 lines than the captions group, which corroborates Liao et al.'s (2020) findings. This is reasonable since participants using captions could only rely on L2 lines to assist their listening comprehension, whereas participants using bilingual subtitles had the opportunity to turn to their L1 for comprehension. Notably, despite the participants' reliance on L1 lines when using bilingual subtitles, they still processed the L2 lines. This can be explained by the fact that the dynamic nature of readable subtitles triggered automatic reading behaviour (Bisson et al., 2014; d'Ydewalle et al., 1991), allowing a degree of attention to be paid to all the dynamic information presented. In addition, as suggested by Li (2016), L2 learners might turn to L2 lines to learn English vocabulary and expressions as a learning strategy or use L2 lines as a way to confirm L1 translations. This is also supported by the study by Liao et al. (2020), in which half of the participants were found to use L2 lines as dominant verbal-visual support, indicating

that at least for intermediate to advanced level L2 learners, the use of L2 can also support their comprehension.

In terms of processing L1 lines, the current findings show that learners in the bilingual subtitles group spent significantly more total reading time on L1 lines than the L1 subtitles group, which is different from the findings of Liao et al. (2020) where no significant difference was observed. However, both groups had a similar fixation % on the L1 lines, indicating that although participants using bilingual subtitles were less likely to skip L1 lines and spent more time in the L1 lines area, they had a similar number of fixations on the L1 lines area as the L1 subtitles group. There are two potential explanations for these findings: 1) since the L1 lines were presented above the L2 lines in the bilingual subtitles condition, participants were very likely to pass through the L1 lines to read the L2 lines, which might contribute to more reading time and a higher run count and a lower skip rate recorded for the L1 lines area; 2) participants using bilingual subtitles made indeed more use of L1 lines as a reference for L2 lines. Previous research has shown that the L1 lines are especially useful for learners to check the meaning of unknown words when using bilingual subtitles (Li, 2016). Therefore, it is possible that the longer time spent on L1 lines in bilingual subtitles was used to check the L1 translations of specific expressions or unfamiliar words, instead of merely for comprehension. This possible explanation can be checked by looking at the analysis of Level 3 for the processing of unknown TWs (see following section).

Participants with a larger vocabulary size tended to spend less time on captions and were more likely to skip them. However, participants' vocabulary size did not significantly affect their processing of bilingual or L1 subtitles, indicating that when L1 translations were included, participants' use of on-screen text was not significantly influenced by their vocabulary size.

#### **4.3.3.3. Level 3: Target Word Area**

Level 3 analysis sought to investigate participants' processing of unknown TWs and their corresponding L1 translations during bilingual subtitled viewing. Seven early and late measures were used to paint a comprehensive picture of the reading of each unknown TW. Since the different measures demonstrated similar results, only total reading time, first-pass reading time, second-pass reading time, fixation count, and skip rate are discussed in line with relevant studies (e.g., Godfroid et al., 2018; Mohamed, 2018; Montero Perez, 2019; Montero Perez et al., 2015; Pellicer-Sánchez, 2016).

Similar to the findings in Level 2, within the bilingual subtitles group, participants processed both L2 words and their translations but spent significantly more time on L1 translations. This might reflect participants' reliance on L1 for better comprehension, since understanding the content was the aim of the activity. This finding also supports Lwo and Lin's (2012) claim that learners who use bilingual subtitles might turn to their L1 as a shortcut to facilitate comprehension, resulting in less attention being paid to the L2 forms.

When comparing bilingual subtitles with the captions and L1 subtitles groups, participants seeing bilingual subtitles spent significantly less time on L2 TWs than the captions group, as revealed by all the measures examined. This finding also concurs with the form recognition results, showing significant advantages of captions over other subtitling conditions. The longer processing time for TWs in the captions group might indicate learners' attempts to encode new forms into memory or guess the meanings of unknown words. Reading research does indeed suggest that a longer initial reading time as well as cumulative time may reflect readers' attempts to infer word meanings (e.g.,

Godfroid et al., 2013, 2018). However, when using bilingual subtitles, viewers can directly refer to L1 translations to understand unknown TWs, which may account for the shorter reading time for L2 forms.

Interestingly, participants in the bilingual subtitles group also spent significantly more time reading L1 translations than the L1 subtitles group. This reading of L1 translations could signal participants' attempts to build form-meaning connections, indicating the benefits of using L1 in establishing an initial form-meaning link. This finding echoes Paivio's (1986, 2014) bilingual version of the Dual Coding Theory, showing that the use of L1 and L2 together with images can help in building a stronger connection for an individual's information process, complementing additive effects. This finding also supports Li's (2016) finding that bilingual subtitles have a "building connection" function, as mentioned by one of Li's participants: "It is easier to combine the two languages and build a connection between them by bilingual subtitles" (p. 198). Importantly, this pattern of eye movements helps to explain the advantage of bilingual subtitles for learning the meaning of unknown words, as shown in the results of the vocabulary tests.

In line with the findings of Montero Perez et al. (2015), the frequency of occurrence of unknown TWs showed a positive effect on learners' cumulative processing time of unknown words when using bilingual subtitles. Moreover, unknown words with longer presentation times in the video were also more likely to be attended to when using bilingual subtitles and captions. Additionally, in line with previous reading findings (Rayner, 2009), words containing more letters tended to be processed for longer, as reflected in the refixation measures during bilingual subtitled and captioned viewing.

To sum up, participants using bilingual subtitles spent similar amounts of time on the subtitling area as the captions group. However, instead of distributing their attention evenly between L1 and L2 subtitle lines, participants spent more time on L1 lines than L2 lines when using bilingual subtitles. The time spent on L1 lines was even longer than the L1 subtitles group, whereas the time spent on L2 lines was shorter than the captions group. Similar patterns were revealed for the reading of unknown TWs. Especially, bilingual subtitle users spent more time reading the L1 translations of unknown words than L1 subtitle users, reflecting their potential effort to establish a form-meaning link for unknown vocabulary when using bilingual subtitles.

#### ***4.3.4. RQ4 – Relationship Between Offline and Online Measures***

The fourth research question explores the potential relationship between participants' attention allocation to unknown TWs (or their translations) and their vocabulary gains. The predictive role of eye-fixation times differed by vocabulary component and by subtitling condition.

In terms of form recognition, a longer reading time on L2 unknown TWs significantly predicted learning gains for both the bilingual subtitles and captions groups. This is in line with previous findings showing that a longer time spent on unknown words was related to successful form recognition in a posttest, for both reading (e.g., Godfroid et al., 2013; Mohamed, 2017) and viewing (e.g., Montero Perez et al., 2015). In line with results of Montero Perez et al. (2015), for the captions group, the first-pass reading time for L2 words significantly predicted form recognition gains, while total reading time failed to reach a significant level. For the bilingual subtitles group, both measures significantly predicted form recognition gains, indicating that the first-pass

reading time for unknown words may involve cognitive processes that are essential for word learning. However, second-pass reading time was not a useful predictor for any vocabulary gains. In intentional learning settings, second-pass reading time could reflect learners' increased efforts to memorise unknown words, resulting in greater learning gains (Montero Perez et al., 2015). However, in incidental learning settings, as Montero Perez et al., (2015) argue, a longer second-pass reading time on a novel word might indicate some processing difficulty and unsuccessful integration of the word, rather than successful learning. This is similar to L1 reading where second-pass reading reflects the reanalysis of a word when encountering an initial processing difficulty (Godfroid, 2020a).

Moving on to the meaning aspects of vocabulary learning, the processing time for L2 words was also a significant predictor of meaning recall gains in the bilingual subtitles condition. This relationship has also been reported in L2 reading studies (e.g., Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, 2016). For the captions group, a longer time spent on L2 TWs led to higher meaning recognition scores, which also supports the findings of L2 reading studies (e.g., Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez et al., 2021). The time spent on unknown words only predicted meaning recall gains when using bilingual subtitles, and it only predicted meaning recognition gains for the captions group. This discrepancy may relate to learners' different cognitive processes triggered by different subtitling types. The presence of L1 translations in bilingual subtitles allowed the participants to map L1 translations to L2 forms more easily. Therefore, a longer processing time for L2 forms might indicate participants' intention to commit novel form-meaning links to memory, which may result in greater meaning recall gains. However, when using captions, a longer processing time for L2 forms might reflect participants' greater effort to infer the

meanings of unknown words. However, since no L1 translations were available, participants' inferring of outcomes might only be strong enough to be manifested in meaning recognition but not reflected in a more demanding meaning recall test.

Overall, the results of the present study provide further evidence to support the role of eye-movement measures with L2 unknown vocabulary to predict vocabulary gains, in line with previous studies (e.g., Godfroid et al., 2013, 2018; Mohamed, 2017; Montero Perez et al., 2015; Pellicer-Sánchez, 2016). Especially, the predictive role of eye movement measures was stronger in the bilingual subtitles group than in the captions group, as revealed by the larger effect sizes reported for the former group. Interestingly, the processing of L1 translations failed to predict any of the vocabulary scores in either the L1 or bilingual subtitles groups. This points to an interesting contradiction. Participants in the bilingual subtitles group spent more time on L1 and that extra time is interpreted as a reflection of attempts to build form-meaning connections, which is then reflected as an advantage of bilingual subtitles over captions and L1 subtitles in meaning recognition and meaning recall scores, respectively. However, the time spent processing L1 was not a significant predictor of vocabulary gains. This suggests that it is not only the amount of attention allocated to translations that is important, as measured by fixation durations, but what participants actually do when processing those translations and the underlying cognitive processes involved.

To sum up, when using bilingual subtitles and captions, a longer reading time for L2 unknown TWs may in general facilitate learners' knowledge of word forms. Moreover, a longer processing time for L2 forms could also facilitate better knowledge of word meanings, as reflected in the meaning recall and recognition scores in the bilingual subtitles and captions group, respectively. However, second-pass reading time for L2 unknown words was not useful for predicting any vocabulary learning gains.

Besides, reading L1 translations could not predict learners' vocabulary gains. These findings indicate the complexity of the relationship between learners' online processing and their learning gains. As argued by Montero Perez et al. (2015), eye-tracking data cannot paint a full picture of learners' engagement with unknown words, and it is not clear whether the reading time reflects a learning process, learning difficulty, or just superficial viewing behaviour. This further attests to the complexity of the relationship between processing times and outcome measures and points to the need to combine eye-movement data with other types of data, such as stimulated recall, to delve further into the different subprocesses involved (Godfroid & Schmidtke, 2013; Pellicer-Sánchez, 2020). The next chapter presents the qualitative part of the present research to further explore participants' engagement with unknown words during subtitled viewing.

## **Chapter 5. Qual Analyses and Results**

Analyses reported in Chapter 4 explored one aspect of engagement with unknown vocabulary, i.e., learners' amount of attention to the unknown words, as reflected in their eye movements. The analyses presented in this chapter addressed the following questions:

RQ5: How do learners engage with unknown TWs, as measured by their level of awareness and use of vocabulary processing strategies, during bilingual subtitled viewing, compared to captions and L1 subtitles as reported in stimulated recall interviews?

RQ6: Do participants' awareness and processing strategies of unknown TWs at the group level corroborate their vocabulary learning gains and their attention allocated to those words?

This chapter first reports analysis procedures of the qualitative data, followed by a summary of results and an interim discussion. It first explains steps taken to sample the stimulated recall data for coding, the coding procedures at two different levels (i.e., awareness and vocabulary processing strategies), and inter-coder reliability test results. Then, findings of the stimulated recall are presented. This chapter ends with a section discussing the results for each level of analysis, and the triangulation of qualitative and quantitative findings.

### **5.1. Stimulated Recall Analysis**

#### ***5.1.1. Data Sampling***

Several steps were taken to analyse stimulated recall data following the analysis guidance provided by Gass and Mackey (2017). First, all the qualitative data were fully transcribed using a Chinese transcription software Iflytek (<https://www.iflyrec.com/>). Transcriptions were then manually corrected by the researcher based on audio recordings. Then, the data were sampled for analysis. In the present study, 15 participants from each group (except for the no subtitles group) were randomly chosen for the stimulated recall analysis, resulting in an overall sample size of 45 participants. This number of samplings is in line with previous stimulated recall research. In a content analysis review of 88 journal articles using stimulated recall in applied linguistics research that had been published between 2012 and 2018, Hugo Santiago and Trevor (2020) found that the number of participants in stimulated recall studies were relatively small, with often no more than ten people, and only in a few studies it ranged from 30 to 77.

After choosing sampled participants, the next step was to choose the TWs for coding. Multiple exposures of a word are considered an influential factor to manipulate learners' engagement (Schmitt, 2008). Since most TWs in the present study appeared only once in the video, it was decided to focus on learners' initial engagement with each TW. Thus, five TWs out of the total 24 which appeared more than once in the viewing material were discarded, resulting in 19 TWs included in the coding and qualitative analysis. Following the aforementioned steps, 855 cases ( $45 \text{ participants} \times 19 \text{ TWs}$ ) were then prepared for coding. As suggested by Gass and Mackey (2017), the stimulated recall data can be separated into segments based on turn boundaries or idea units. For the data in the present research, turn boundaries were mostly clearly marked by each TW, as announced by the researcher during the stimulated recall aiming to elicit participants' recall. As shown in Example 1, when the researcher (denoted by *R*)

pronounced the TW *barneys*, participant A25's response can be clearly separated from the previous recall of the TW *confiscated*. In very few cases, confirmations were made during the stimulated recall where the turn boundary was not clear enough for the researcher to judge which word participants were commenting on.

### Example 1.

---

R: 这个 *confiscated*。 [This one, *confiscated*.]

A25: 嗯这个我是不认识的。嗯对，但我那时候在想是什么意思。 [Well, I didn't know this one. Yeah, but I was wondering what it means.]

R: 嗯，那你还记得你当时就是怎么想的吗？ [Okay, so do you remember what you were thinking at that time?]

A25: 嗯，就猜了一下它大概是什么意思，然后看了一下图片有没有帮助。 [Well, just guessed what it means, and then checked the picture to see if it helps.]

R: 嗯嗯，这个 *barneys* 当时有没有注意到这个词？ [Hmm, okay. This one, *barneys*, did you notice this word at that time?]

A25: 有听到，但是不知道什么意思，对，然后看一下画面也没懂。 [I listened to it, but I didn't know what it means. Yes, I looked at the picture but still didn't understand.]

---

### 5.1.2. Data Coding

Since the aim of the stimulated recall analysis was to explore learners' initial engagement with unknown words during subtitled viewing, those words that were familiar or partially familiar to each participant before the experiment were discarded for analysis. After coding the 855 cases (45 participants  $\times$  19 TWs) using NVivo 12

software, two steps were taken to ensure participants' unfamiliarity with the TWs prior to the experiment: 1) TWs that were scored correctly (as 1) on the form recognition pretest were discarded for each participant, resulting in 640 remaining cases (captions group = 215 cases; L1 subtitles group = 214 cases; and bilingual subtitles group = 211 cases); 2) as shown in Table 36, each TW was coded as "unknown", "partially known", or "known", based on participants' responses after the researcher's pronunciation of each TW or to the first stimulated recall question "Did you notice this word at that time?". Participants were told in the oral instructions that they should inform the researcher if a TW was already familiar to them before their viewing when it was first mentioned in the stimulated recall. Participants' familiarity with a TW was judged by their self-reports, without the need to demonstrate their knowledge by providing a translation. Among the remaining 640 cases, those that were reported as "partially known words" (4.53%) or "known words" (3.75%) in the stimulated recall were deleted for analysis, resulting in a final 587 cases in total (captions group = 204 cases; L1 subtitles group = 192 cases; bilingual subtitles group = 191 cases). This ensured that the analysis of learners' initial engagement with TWs was indeed about their engagement with items that were unknown to them, as reflected in both pretest scores and their own reports in the stimulated recall.

Table 36. Coding Categories and Examples of Participants' Familiarity with the Target Words in the Stimulated Recall

|               | Definitions                            | Examples                                    |
|---------------|----------------------------------------|---------------------------------------------|
| Unknown words | Reported their unfamiliarity of the TW | "I noticed this word, it was a novel word." |
| Partially     | Mentioned that they had                | "I felt this word... because this word      |

|             | Definitions                                | Examples                                                                     |
|-------------|--------------------------------------------|------------------------------------------------------------------------------|
| known words | seen/heard the TW before                   | looked familiar, but I didn't know its meaning, it seemed to be left aside." |
| Known words | Reported their prior familiarity of the TW | "This one I knew its meaning."                                               |

After excluding familiar TWs, each participant's engagement with each unknown TW was then coded. Svalberg's (2009) definition of engagement was adopted, as discussed in section 2.5.5. Learners' cognitive engagement in the present study was operationalised at three levels: *attention*, *awareness* and *vocabulary processing strategies*. Apart from the attention level, which has been reported in Chapter 4, the latter two levels are illustrated in Figure 15. Since no previous research has provided a framework to categorise learners' engagement with vocabulary in the viewing context, and almost all the vocabulary learning strategy taxonomies were developed for traditional classroom-based intentional learning (Segler, Pain, & Sorace, 2002), the present coding followed an inductive approach by adopting a qualitative content analysis method (Dörnyei, 2007; Selvi, 2020). This approach was particularly suitable for the current study for its emphasis on allowing categories to be derived inductively from the data (Bryman, 2012; Dörnyei, 2007). Therefore, themes and categories were generated in a data-driven approach. The coding procedures for these two levels are explained in turn.

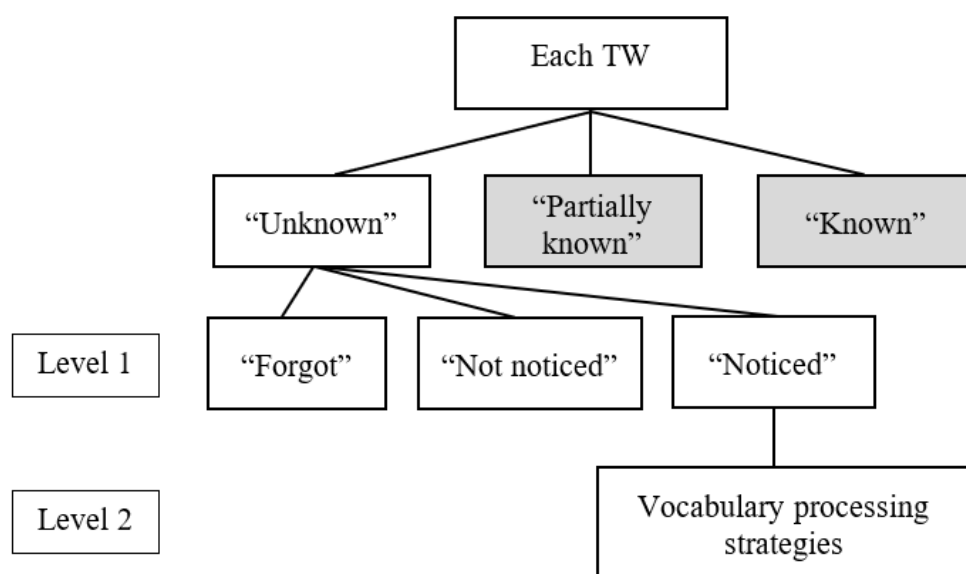


Figure 15. Illustration of Stimulated Recall Coding Procedures at Two Levels for Each Target Word

As shown in Figure 15, for Level 1 (awareness level), to explore participants' reported awareness of each TW, each unknown TW was coded as "forgot", "not noticed", or "noticed" based on participants' answers to the first stimulated recall question: "Did you notice this word at that time?". In the present study, *noticing* and *awareness* are used interchangeably as explained in section 2.5.5. I use the definition of *awareness* at the noticing level as proposed by Schmidt (1990), to focus on the "subjectively experienced" (p. 132) attribute, and thus noticing referred to learners' private experience. In the stimulated recall, the word *notice* was used as an everyday language indicating learners' self-reported awareness of either the written or auditory form of TWs. To be specific about the different categories at Level 1, "forgot" indicated the situations where participants mentioned that they forgot/did not remember/were uncertain about whether they had noticed or not noticed the TW while viewing. "Not noticed" indicated the cases where participants explicitly mentioned that they did not notice the TW during viewing, or they did not explicitly mention noticing by

commenting on something else relating to the content/images. “Noticed” was tagged to the cases where participants reported that they had noticed the English TW, either visually, aurally, or both, depending on the different subtitling conditions. The difference regarding the noticing modality was not distinguished in coding due to the difficulty for participants to specify when both forms were available. It should be noted that, however, for the L1 subtitles group, participants could only notice TWs via listening due to the lack of written L2 form in the subtitling area. In addition, it should be acknowledged that only participants’ self-reported noticing of L2 unknown TWs, rather than corresponding L1 translations, during viewing was considered as “noticed” in coding. Detailed definitions and examples are presented later in section 5.2.1.

As shown in Figure 15, the coding of Level 2 (vocabulary processing strategies), only applied to TWs that were coded as “noticed” at Level 1. To categorise the strategies used for all noticed unknown TWs, participants’ stimulated recall data were coded inductively following the steps suggested by Bryman (2012) and Selvi (2020): 1) concepts were generated by coding data at the level of open coding; 2) categories were generated through a constant comparison of concepts, micro-categories were grouped into more general categories; 3) saturated categories were listed; 4) categories were applied back to the stimulated recall data pertaining each TW. To name the categories emerging in the stimulated recall, previous works exploring L2 learners’ vocabulary learning strategies (e.g., Barcroft, 2009; Gu & Johnson, 1996; Lawson & Hogben, 1996; Nation, 2001; Schmitt & McCarthy, 1997) and vocabulary processing strategies in reading (e.g., Fraser, 1999; Knight, 1994; Rott, 2000; Rott, 2005) were used as references. For each unknown TW, all reported processing strategies attached to this word were coded, despite the fact that in some cases they were incorrect or inaccurately

applied. The number of codes for each strategy was calculated for each subtitling condition to generate group comparisons.

In summary, general levels and main codes generated from the stimulated recall data are presented in Table 37. Detailed coding scheme and examples of each category and sub-category are presented later in section 5.2.2.

Table 37. Two Levels and Main Codes Generated from Stimulated Recall Interviews for Participants' Engagement

| Two levels of participants' engagement   | Main codes                                                                                                                                  |
|------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Level 1 Awareness                        | Forgot<br>Not noticed<br>Noticed                                                                                                            |
| Level 2 Vocabulary processing strategies | Word feature analysis<br>Using context<br>Using L1 translations<br>Guessing without reported strategies<br>No reported strategies<br>Others |

To address RQ6, participants' awareness of unknown TWs and their vocabulary processing strategies in different subtitling conditions were compared and discussed with quantitative findings (including participants' vocabulary gains and their online processing of the unknown TWs). This allowed me to have a more thorough picture of participants' engagement with unknown words during subtitled viewing and examine whether the main findings corroborate with each other and to inspect potential disagreements.

### 5.1.3. *Inter-Coder Agreement in Coding*

After coding the stimulated recall data, in order to investigate group differences regarding learners' engagement at two levels, content analysis was used to compute a frequency count by adding up the number of TW falling into a particular category. Percentages for each category and sub-category were calculated for each subtitling group at the two engagement levels. To examine the reliability of the coding, 20% of the data, i.e., data from nine randomly selected participants (three from each group with on-screen text), was coded by the same second coder who scored meaning recall tests and also had experience of coding stimulated recall data. Inter-coder reliability was calculated using Cohen's kappa. The coding showed an agreement for awareness level of  $\kappa = .96$  (95% CI, .93 to 1),  $p < .001$  and for vocabulary processing strategy level of  $\kappa = .89$  (95% CI, .82 to .95),  $p < .001$ , indicating almost perfect agreement between the two coders. Any disagreements between the coders were resolved after discussion.

## 5.2. **Qual Results**

As discussed in previous sections, the aim of the stimulated recall was to delve further into learners' engagement with unknown words during subtitled viewing and to better understand the relationship between learners' engagement and their learning gains. RQ5 aimed to answer how learners noticed unknown TWs and what types of strategies they used to process unknown TWs during viewing, and how these may differ between the bilingual subtitles group and monolingual groups (i.e., captions and L1 subtitles). This section presents learners' engagement with unknown TWs at two levels: *awareness* and *vocabulary processing strategies*. However, during data coding, an

additional level (i.e., *intentionality of learning*) emerged with a few reported cases.

Therefore, the results of the analysis are presented for the three levels separately, with a general summary of each level in each group followed by comparisons among the three subtitling conditions.

### 5.2.1. Level 1: Awareness

After discarding partially known and known words, participants' reported awareness of the unknown TWs was explored. As explained in section 5.1.2, the stimulated recall data were grouped into three categories for unknown TWs (i.e., forgot, not noticed, noticed). During coding, the "not noticed" category was further divided into two subcategories. These categories are shown in Table 38.

Table 38. Results for Coding Categories of Participants' Reported Awareness (Level 1) of Unknown Target Words in Stimulated Recall Interviews

| Categories  | Subcategories                          | Definitions                                                                                                   | Examples                                                  |
|-------------|----------------------------------------|---------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|
| Forgot      |                                        | Forgot whether they had noticed the TW or not during viewing                                                  | "I don't remember."                                       |
| Not noticed | Not noticed (explicitly mentioned)     | Explicitly mentioned their lack of noticing of the TW during viewing                                          | "I didn't notice this word at that time."                 |
|             | Not noticed (not explicitly mentioned) | Did not comment on their noticing of the TW during viewing by commenting on video content, images, audio etc. | "I was working hard to understand the speakers' posture." |

| Categories | Subcategories | Definitions                                                                                | Examples                            |
|------------|---------------|--------------------------------------------------------------------------------------------|-------------------------------------|
| Noticed    |               | Noticed the TW, reported that they had seen/heard it or guessed its meaning during viewing | “I noticed this word at that time.” |

Table 39 presents the summary of participants’ awareness of unknown TWs in three subtitling conditions. In general, participants failed to recall 5.79% of the cases (i.e., forgot) during the stimulated recall due to memory decay. More than half of the unknown TWs were reported as “not noticed” (53.49%), compared to 40.72% of the “noticed” cases in all groups. Therefore, in general, participants noticed around 40% of unknown TWs that appeared once during subtitled viewing. However, this number varied across different subtitling conditions. As shown in Table 39, the captions group recorded the highest number of “noticed” cases (50.98%), followed by the bilingual subtitles group (41.88%), and the L1 subtitles group (28.65%). Moreover, among “not noticed” cases, most of them were explicitly indicated by participants, and most of “not noticed” cases were recorded in the L1 subtitles group (66.15%), followed by the bilingual subtitles group (54.45%) and the captions group (40.69%). It can be observed that only the captions group reported more noticed cases than not noticed ones.

Table 39. Results for the Frequency and Percentage of Participants’ Reported Awareness (Level 1) of Unknown Target Words in Stimulated Recall Interviews by Group

| Categories | Subcategories | Captions (%) | L1 (%)    | Bilingual (%) | Total frequency (%) |
|------------|---------------|--------------|-----------|---------------|---------------------|
| Forgot     |               | 17 (8.33)    | 10 (5.21) | 7 (3.66)      | 34 (5.79)           |

| Categories  | Subcategories                          | Captions (%) | L1 (%)      | Bilingual (%) | Total frequency (%) |
|-------------|----------------------------------------|--------------|-------------|---------------|---------------------|
| Not noticed |                                        | 83 (40.69)   | 127 (66.15) | 104 (54.45)   | 314 (53.49)         |
|             | Not noticed (explicitly mentioned)     | 76 (37.25)   | 95 (49.48)  | 102 (53.40)   | 273 (46.51)         |
|             | Not noticed (not explicitly mentioned) | 7 (3.43)     | 32 (16.67)  | 2 (1.05)      | 41 (7.00)           |
| Noticed     |                                        | 104 (50.98)  | 55 (28.65)  | 80 (41.88)    | 239 (40.72)         |
| Total (%)   |                                        | 204 (100)    | 192 (100)   | 191 (100)     | 587 (100)           |

### ***5.2.2. Level 2: Vocabulary Processing Strategies***

Analysis at Level 2 focused on processing strategies that participants used to construct their knowledge of the unknown TWs that they noticed during subtitled viewing. Only unknown TWs that were reported as noticed at Level 1 were further analysed at this level, resulting in a total of 239 cases. As Table 40 shows, there were six broad categories including 19 subcategories reported by sampled participants.

Table 40. Results for the Coding Categories of Participants' Vocabulary Processing Strategies (Level 2) of Unknown Target Words in Stimulated Recall Interviews

| Strategy categories       | Subcategories                     | Further subcategories | Definitions                                                                             | Example quote                                                                                                                                                                                  |
|---------------------------|-----------------------------------|-----------------------|-----------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Word features analysis | 1.1. Analysing part of speech     |                       | Analysed the part of speech or reported awareness of the part of speech of the TW       | "Hmm I guessed it [ <i>endearing</i> ] was a positive adjective..."                                                                                                                            |
|                           | 1.2. Analysing word-structure     |                       | Analysed the affixes, suffixes, and/or roots of the TW                                  | "‘ <i>Confiscated</i> ’... Because looking at the word-structures, ‘con’ is very common, ‘fiscated’, I knew how to pronounce this type of word by a glance, but just didn’t know its meaning." |
|                           | 1.3. Spelling association         |                       | Associated the TW with a known word sharing similar spelling or pronunciation of the TW | "... ‘ <i>buffering</i> ’, I immediately thought about ‘suffering’..."                                                                                                                         |
|                           | 1.4. Analysing word pronunciation |                       | Used the pronunciation of the word to understand its meaning                            | "I thought it [ <i>nuzzle</i> ] meant a nap, because the word ‘ <i>nuzzle</i> ’ sounded very intimate."                                                                                        |
|                           | 1.5 Word usage                    |                       | Reported their awareness of the usage of the TW                                         | "‘ <i>buffering of stress</i> ’. I heard this phrase. I had an impression of it at that time. I didn’t know it could be combined with ‘stress’, so I thought                                   |

| Strategy categories | Subcategories                    | Further subcategories | Definitions                                                                                         | Example quote                                                                                                                                                                                                        |
|---------------------|----------------------------------|-----------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                     |                                  |                       |                                                                                                     | about it.”                                                                                                                                                                                                           |
| 2. Using context    | 2.1. Using auditory cues         |                       | Used the audio/sound of the video to support their understanding of the TW                          | “Oh, I have a deep impression about this one, because the cat was purring at that time, then I knew this [ <i>purring</i> ] must mean purring [in Chinese].”                                                         |
|                     | 2.2. Using images                |                       | Used the images of the video to support their understanding of the TW                               | “At that time, I guessed it [ <i>barneys</i> ] should be... because the image showed the house, so I thought it should be the name of the residence or that kind of building.”                                       |
|                     | 2.3. Using global understanding  |                       | Used the video plot or background knowledge to support their understanding of the TW                | “I noticed this word [ <i>bizarre</i> ] at that time. The context and plots before and after all indicated the situation was weird.”                                                                                 |
|                     | 2.4. Using local contextual cues |                       | Used the word(s) that appeared just before or after the TW to support their understanding of the TW | “Oh, this word I guessed its [ <i>sedated</i> ] meaning when I watched. Because the following is ‘they are real’, so the ‘not’ before the word should mean they are not. It has the meaning of virtual, or digital.” |
| 3. Using L1         | 3.1. Using L1                    | 3.1.1. L2 triggered   | Noticed the L2 TW first, then                                                                       | “Yes, this word [ <i>sedated</i> ] I didn’t know, and so I                                                                                                                                                           |

| Strategy categories | Subcategories             | Further subcategories                 | Definitions                                                                                                          | Example quote                                                                                                                                        |
|---------------------|---------------------------|---------------------------------------|----------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| translations        | matched                   | reference to L1                       | checked its L1 translation                                                                                           | took a look at the subtitles.”                                                                                                                       |
|                     |                           | 3.1.2. L1 triggered reference to L2   | Noticed the L1 translation of the TW first, then checked how to express the meaning in L2                            | “At that time, I was curious about how to express confiscation [in Chinese], so I took a look at the English word [ <i>confiscated</i> ].”           |
|                     |                           | 3.1.3. Using L1 no sequence mentioned | Noticed both the L1 translation and L2 TW but not mentioned the particular sequence of noticing                      | “I heard ‘ <i>hump</i> ’, and then I knew it means hump [in Chinese].”                                                                               |
|                     | 3.2. Using L1 not matched | 3.2.1. L1 triggered other L2          | Noticed the L1 translation of the TW first, then thought about other L2 words sharing similar meaning but not the TW | “...[For TW <i>midwife</i> ] because I seemed to want to hear the pronunciation of ‘nurse’, but I didn’t hear ‘nurse’, as far as I remember, yes...” |
|                     |                           | 3.2.2. Mismatched L1 and L2           | Failed to match the L1 translation with the L2 TW                                                                    | “...Oh I saw the Chinese subtitles. She said how to buffer stress [in Chinese], but I didn’t know which word was ‘ <i>buffer</i> ’.”                 |
|                     |                           | 3.2.3. Lack of time to check L1       | Noticed the L2 TW but did not have time to check its L1 translation                                                  | “...That is [ <i>foraging</i> ], because the subtitles jumped fast, sometimes I may not have time to read the Chinese meaning.”                      |

| Strategy categories                     | Subcategories                       | Further subcategories | Definitions                                                                        | Example quote                                                                                                                                                        |
|-----------------------------------------|-------------------------------------|-----------------------|------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4. Guessing without reported strategies | 4.1. Meaning fully guessed          |                       | Reported a guessed meaning directly with no specific strategy mentioned            | "I heard it [ <i>surrogate</i> ] at that time, and then I knew it was, it was something like a surrogate mother [in Chinese]."                                       |
|                                         | 4.2. Meaning partially guessed      |                       | Reported a partially guessed meaning with no specific strategy mentioned           | "Hmm... something good, I didn't know the exact meaning, but I only knew it [ <i>endearing</i> ] means something nice."                                              |
|                                         | 4.3. Meaning unsuccessfully guessed |                       | Reported attempts to guess the meaning of the TW but not reported outcomes         | "I don't know what it [ <i>sedated</i> ] means, but at that time I didn't know the meaning either, but I have attempted to guess I think."                           |
| 5. Other strategies                     | 5.1. Pretest impact                 |                       | Mentioned that they remembered had seen the TW in the pretest during their viewing | "Because I have seen this word [ <i>midwife</i> ] last time when completing the test, and then this time when I saw this word, I paid attention to it particularly." |
|                                         | 5.2. Dictionary use                 |                       | Attempted to remember the TW and refer to dictionary afterwards                    | "I wanted to know its [ <i>barneys</i> ] meaning, and I even wanted to memorise it and to check the dictionary after."                                               |
|                                         | 5.3. Visualizing                    |                       | Created mental images for the TW                                                   | "...[For TW <i>hump</i> ] I saw it, it didn't show the way it was humped [in Chinese], but I could image what it would look like."                                   |

| Strategy categories       | Subcategories                    | Further subcategories | Definitions                                                                                        | Example quote                                                                                     |
|---------------------------|----------------------------------|-----------------------|----------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| 6. No reported strategies | 6.1. No reported meaning guessed |                       | Noticed the unknown word but did not report attempts to guess or make form-meaning link for the TW | “I didn’t know. I just felt I didn’t know the meaning (of the TW <i>foraging</i> ) at that time.” |
|                           | 6.2. Forgot thoughts             |                       | Reported that they forgot what they were thinking when they noticed the TW                         | “I saw this word [ <i>twirls</i> ]. But I don’t remember, yes.”                                   |

As can be seen in Table 40, six broad categories included: 1. Word feature analysis; 2. Using context; 3. Using L1 translations; 4. Guessing without reported strategies; 5. Other strategies; and 6. No reported strategies. It should be noted that in some cases, although participants firstly reported their noticing of an unknown TW, they did not report their guess/thoughts about the TW. This was named as “6. No reported strategies”. This category should be distinguished from: a) the “Not noticed” category at Level 1 (i.e., awareness level), where participants did not report their awareness of a TW; and b) the “4. Guessing without reported strategies” category at Level 2 (i.e., vocabulary processing strategies level), where participants provided a guessed meaning, or demonstrated their attempts to guess the meaning while viewing.

Among the 239 cases where participants noticed the unknown TWs, there were 4 cases (1.49%) where participants forgot what they were thinking while viewing after confirming their noticing of the unknown TWs (labelled as “6.2. Forgot thoughts” in Table 40). In terms of strategy combination, in most cases, participants used only one strategy for each unknown TW (89.12%). However, in 26 cases (10.88%), participants combined two (9.21%) or three (1.67%) strategies to engage with one TW. Overall, there were 269 instances of strategy use for the 239 cases of noticed unknown TWs. The bilingual subtitles group witnessed the highest number of combined strategies (5.02%), followed by the captions group (4.60%), and the L1 subtitles group (1.26%). Table 41 presents the number of instances that each of the strategies were combined with others in each group. In general, strategies that were most frequently combined with others were: 2.2. Using images (25.00%), 3.1. Using L1 matched (23.21%), and 2.3. Using global understanding (12.50%). For the captions group, more combination cases occurred when using context (i.e., subcategories 2.2, 2.3, and 2.4), while for the

bilingual and L1 subtitles groups, “3.1. Using L1 matched” and “2.2. Using images” were combined most frequently with other strategies.

Table 41. Results for the Frequency and Percentage of Participants’ Combination of Different Vocabulary Processing Strategies by Group

| Subcategories                        | Captions<br>(%) | L1 (%)    | Bilingual<br>(%) | Total (%)  |
|--------------------------------------|-----------------|-----------|------------------|------------|
| 1.1. Analysing part of speech        | 2 (8.00)        | 0 (0)     | 1 (4.00)         | 3 (5.36)   |
| 1.2. Analysing word-structure        | 2 (8.00)        | 0 (0)     | 0 (0)            | 2 (3.57)   |
| 1.3. Spelling association            | 2 (8.00)        | 0 (0)     | 4 (16.00)        | 6 (10.71)  |
| 1.4. Analysing word<br>pronunciation | 1 (4.00)        | 0 (0)     | 1 (4.00)         | 2 (3.57)   |
| 1.5. Word usage                      | 0 (0)           | 0 (0)     | 1 (4.00)         | 1 (1.79)   |
| 2.2. Using images                    | 7 (28.00)       | 2 (33.33) | 5 (20.00)        | 14 (25.00) |
| 2.3. Using global understanding      | 5 (20.00)       | 1 (16.67) | 1 (4.00)         | 7 (12.50)  |
| 2.4. Using local contextual cues     | 6 (24.00)       | 0 (0)     | 0 (0)            | 6 (10.71)  |
| 3.1. Using L1 matched                | 0 (0)           | 2 (33.33) | 11 (44.00)       | 13 (23.21) |
| 5.3. Visualizing                     | 0 (0)           | 1 (16.67) | 1 (4.00)         | 2 (3.57)   |
| Total (%)                            | 25 (100)        | 6 (100)   | 25 (100)         | 56 (100)   |

Table 42 summaries the frequency of which each strategy was used in the three subtitling groups. In general, from the six broad categories, categories 3. Using L1 translations (31.60%), 2. Using context (27.51%), and 6. No reported strategies (22.30%) were the most frequent categories that emerged across groups. To be specific, the top five most frequently used sub-strategies across groups were: 3.1. Using L1 matched (29.74%), 6.1. No reported meaning guessed (20.82%), 2.2. Using images (10.04%), 2.4. Using local contextual cues (8.55%), and 2.3. Using global understanding (5.95%), as shown in Table 42.

Table 42. Results for the Frequency and Percentage of Participants' Vocabulary Processing Strategies (Level 2) of Unknown Target Words in Stimulated Recall Interviews by Group

| Strategy categories                     | Subcategories                     | Captions (%) | L1 (%)     | Bilingual (%) | Total (%)  |
|-----------------------------------------|-----------------------------------|--------------|------------|---------------|------------|
| 1. Word feature analysis                |                                   | 13 (11.02)   | 2 (3.45)   | 7 (7.53)      | 22 (8.18)  |
|                                         | 1.1. Analysing part of speech     | 3 (2.54)     | 1 (1.72)   | 1 (1.08)      | 5 (1.86)   |
|                                         | 1.2. Analysing word-structure     | 5 (4.24)     | 0 (0)      | 0 (0)         | 5 (1.86)   |
|                                         | 1.3. Spelling association         | 3 (2.54)     | 0 (0)      | 4 (4.30)      | 7 (2.6)    |
|                                         | 1.4. Analysing word pronunciation | 2 (1.69)     | 1 (1.72)   | 1 (1.08)      | 4 (1.49)   |
|                                         | 1.5. Word usage                   | 0 (0)        | 0 (0)      | 1 (1.08)      | 1 (0.37)   |
| 2. Using context                        |                                   | 51 (43.22)   | 10 (17.24) | 13 (13.98)    | 74 (27.51) |
|                                         | 2.1. Using auditory cues          | 2 (1.69)     | 3 (5.17)   | 3 (3.23)      | 8 (2.97)   |
|                                         | 2.2. Using images                 | 13 (11.02)   | 6 (10.34)  | 8 (8.60)      | 27 (10.04) |
|                                         | 2.3. Using global understanding   | 13 (11.02)   | 1 (1.72)   | 2 (2.15)      | 16 (5.95)  |
|                                         | 2.4. Using local contextual cues  | 23 (19.49)   | 0 (0)      | 0 (0)         | 23 (8.55)  |
| 3. Using L1 translations                |                                   | 0 (0)        | 27 (46.55) | 58 (62.37)    | 85 (31.60) |
|                                         | 3.1. Using L1 matched             | 0 (0)        | 25 (43.10) | 55 (59.14)    | 80 (29.74) |
|                                         | 3.2. Using L1 not matched         | 0 (0)        | 2 (3.45)   | 3 (3.23)      | 5 (1.86)   |
| 4. Guessing without reported strategies |                                   | 22 (18.64)   | 0 (0)      | 0 (0)         | 22 (8.18)  |
|                                         | 4.1. Meaning fully guessed        | 5 (4.24)     | 0 (0)      | 0 (0)         | 5 (1.86)   |

| Strategy categories       | Subcategories                       | Captions (%) | L1 (%)     | Bilingual (%) | Total (%)  |
|---------------------------|-------------------------------------|--------------|------------|---------------|------------|
| 5. Other strategies       | 4.2. Meaning partially guessed      | 9 (7.63)     | 0 (0)      | 0 (0)         | 9 (3.35)   |
|                           | 4.3. Meaning unsuccessfully guessed | 8 (6.78)     | 0 (0)      | 0 (0)         | 8 (2.97)   |
|                           |                                     | 1 (0.85)     | 3 (5.17)   | 2 (2.15)      | 6 (2.23)   |
|                           | 5.1. Pretest impact                 | 0 (0)        | 2 (3.45)   | 1 (1.08)      | 3 (1.12)   |
|                           | 5.2. Want to use a dictionary       | 1 (0.85)     | 0 (0)      | 0 (0)         | 1 (0.37)   |
|                           | 5.3. Visualizing                    | 0 (0)        | 1 (1.72)   | 1 (1.08)      | 2 (0.74)   |
| 6. No reported strategies |                                     | 31 (26.27)   | 16 (27.59) | 13 (13.98)    | 60 (22.30) |
|                           | 6.1. No reported meaning guessed    | 31 (26.27)   | 15 (25.86) | 10 (10.75)    | 56 (20.82) |
|                           | 6.2. Forgot thoughts                | 0 (0)        | 1 (1.72)   | 3 (3.23)      | 4 (1.49)   |
| Total (%)                 |                                     | 118 (100)    | 58 (100)   | 93 (100)      | 269 (100)  |

When looking at the strategies used by each group, it should be noted that some strategies only applied to certain subtitling conditions. As shown in Table 42, the “3. Using L1 translations strategy” was not applicable to the captions group due to the lack of L1 translations in the subtitling area. Moreover, the “4. Guessing without reported strategies” did not apply to the L1 and bilingual subtitles groups because the presence of L1 translations made meaning guess less likely to occur during viewing.

To more easily compare strategy use across different subtitling groups, Table 43 summarises the top three most frequently used strategies in each subtitling condition. It is worth noting that, the “6. No reported strategies” was the second most frequently reported category in all three groups. This is possible due to the incidental nature of the viewing activity. The strategies used in the captions group were distinct from the other two groups. Using context (43.22%), especially using local contextual cues (19.49%) and global understanding (11.02%), was more frequently reported in the captions group than in the other two groups. In addition, the captions group reported the “4. Guessing without reported strategies” (18.64%) more frequently than the other two groups. Table 43 shows more similarities in strategy use between the bilingual and L1 subtitles groups as they shared the same three most frequent strategies. However, it should be noted that the bilingual subtitles group reported more cases of using L1 translations (62.37%) than the L1 subtitles group (46.55%). Moreover, there were fewer cases of “6. No reported strategies” in the bilingual subtitles group (13.98%) than in the L1 subtitles group (27.59%).

Table 43. Top Three Most Frequently Used Vocabulary Processing Strategy Categories and Subcategories in Three Subtitling Groups

| Group    | Rank            | Broad category (%)       | Subcategory (%)          |
|----------|-----------------|--------------------------|--------------------------|
| Captions | 1 <sup>st</sup> | 2. Using context (43.22) | 6.1. No reported meaning |

| Group     | Rank            | Broad category (%)                              | Subcategory (%)                                                      |
|-----------|-----------------|-------------------------------------------------|----------------------------------------------------------------------|
|           |                 |                                                 | guessed (26.27)                                                      |
|           | 2 <sup>nd</sup> | 6. No reported strategies (26.27)               | 2.4. Using local contextual cues (19.49)                             |
|           | 3 <sup>rd</sup> | 4. Guessing without reported strategies (18.64) | 2.2. Using images (11.02)<br>2.3. Using global understanding (11.02) |
| L1        | 1 <sup>st</sup> | 3. Using L1 translations (46.55)                | 3.1. Using L1 matched (43.10)                                        |
|           | 2 <sup>nd</sup> | 6. No reported strategies (27.59)               | 6.1. No reported meaning guessed (25.86)                             |
|           | 3 <sup>rd</sup> | 2. Using context (17.24)                        | 2.2. Using images (10.34)                                            |
| Bilingual | 1 <sup>st</sup> | 3. Using L1 translations (62.37)                | 3.1. Using L1 matched (59.14)                                        |
|           | 2 <sup>nd</sup> | 6. No reported strategies (13.98)               | 6.1. No reported meaning guessed (10.75)                             |
|           | 3 <sup>rd</sup> | 2. Using context (13.79)                        | 2.2. Using images (8.60)                                             |

The most frequently used strategy for the bilingual and L1 subtitles groups was “3. Using L1 translations”. This strategy is then analysed in detail to further explore potential group differences. As shown in Table 44, this category was divided into two subcategories and each of these two subcategories was further divided into three subcategories. Table 44 shows that, on most occasions, participants in the bilingual (59.14%) and L1 subtitles (43.10%) groups could successfully match L1 translations to unknown TWs during viewing, with more successful cases reported in the bilingual subtitles group. When looking at the further subcategories, “3.1.3. Using L1 no sequence mentioned” was the most frequently reported strategy, followed by “3.1.1. L2 triggered reference to L1” in both groups. However, for the bilingual subtitles group, the third most frequently used strategy was “3.1.2. L1 triggered reference to L2”

(8.62%), whereas this strategy was not mentioned by the L1 subtitles group. On the contrary, for the L1 subtitles group, “3.2.1. L1 triggered other L2” (7.41%) was recorded as the third most frequently used, which was not reported by the bilingual subtitles group.

Table 44. Results for the Frequency and Percentage of Subcategories and Further Subcategories Within the “3. Using L1 translations” Category in the L1 and Bilingual Subtitles Groups

| Subcategories             | Further subcategories                 | L1 (%)     | Bilingual (%) |
|---------------------------|---------------------------------------|------------|---------------|
| 3.1. Using L1 matched     |                                       | 25 (43.10) | 55 (59.14)    |
|                           | 3.1.1. L2 triggered reference to L1   | 7 (25.93)  | 8 (13.79)     |
|                           | 3.1.2. L1 triggered reference to L2   | 0 (0)      | 5 (8.62)      |
|                           | 3.1.3. Using L1 no sequence mentioned | 18 (66.67) | 42 (72.41)    |
| 3.2. Using L1 not matched |                                       | 2 (3.45)   | 3 (3.23)      |
|                           | 3.2.1. L1 triggered other L2          | 2 (7.41)   | 0 (0)         |
|                           | 3.2.2. Mismatched L1 and L2           | 0 (0)      | 1 (1.72)      |
|                           | 3.2.3. Lack of time to check L1       | 0 (0)      | 2 (3.45)      |
| Total                     |                                       | 27 (100)   | 58 (100)      |

### 5.2.3. A Third Level of Engagement that Emerged: Intentionality of Learning

Although participants’ intentionality of learning unknown TWs was not explicitly asked in the stimulated recall, a few participants mentioned their intentionality to learn unknown TWs. According to Svalberg’s (2009) definition of engagement, which is adopted in the present study, intentionality can be regarded as the *affective* level of engagement and is also considered an important contributor to learning (as discussed in section 2.5.1). Therefore, although this level was not systematically reported by the

participants, I decided to include a third level of coding to record the cases where participants' vocabulary learning intentionality was reported. However, it should be noted that no firm claims can be made concerning this level because it was not part of the stimulated recall protocol and participants were not explicitly asked to report about their intentionality. It could then be the case that some participants might have intentionally engaged with the TWs, but they did not mention it in the stimulated recall.

As shown in Table 45, among the 239 cases in which participants reported their awareness of unknown TWs at Level 1, participants mentioned their intentionality (or lack of intentionality) to learn unknown TWs while viewing in 25 cases (10.46%), with 11 times not intended to learn (4.60%) and 14 times intended to learn (5.86%). Interestingly, most of reported cases emerged in the bilingual subtitles group. However, it is important to note that no firm conclusions about which conditions lead to higher intentionality of learning could be made due to the limited number of reported cases.

Table 45. Results for the Frequency and Percentage of Participants' Reported Intentionality of Learning Unknown Target Words (Level 3) in Stimulated Recall Interviews by Group

| Categories            | Captions (%) | L1 (%)     | Bilingual (%) | Total (%)   |
|-----------------------|--------------|------------|---------------|-------------|
| Not intended to learn | 2 (1.92)     | 0 (0)      | 9 (11.25)     | 11 (4.60)   |
| Intended to learn     | 1 (0.96)     | 3 (5.45)   | 10 (12.50)    | 14 (5.86)   |
| Not mentioned         | 101 (97.12)  | 52 (94.55) | 61 (76.25)    | 214 (89.54) |
| Total                 | 104 (100)    | 55 (100)   | 80 (100)      | 239 (100)   |

### **5.3. Qual Interim Discussion**

This section first discusses the findings in relation to RQ5, which investigates participants' reported engagement with unknown TWs during subtitled viewing and comparisons between bilingual subtitles and other monolingual subtitle conditions. The findings are discussed individually at three levels (i.e., awareness, vocabulary processing strategies, and intentionality). Then, the qualitative findings are triangulated with the quantitative results (i.e., vocabulary tests and eye-tracking findings) to answer RQ6.

#### **5.3.1. *Level 1: Awareness***

Stimulated recall data first showed that in the majority of cases (94.21%), participants were able to report unknown TWs as noticed or not noticed, which suggests that stimulated recall is an appropriate method for examining learners' awareness of novel vocabulary.

In all three subtitling conditions, participants were more likely not to notice unknown TWs than to notice them. The overall noticed rate (i.e., reporting having noticed TWs) was 40.72%, which was still higher than the rate reported in the L2 reading study by Godfroid et al. (2013), with a recorded 31.1% of noticing cases. This suggests a possible advantage for subtitled viewing over reading-only conditions for increasing L2 learners' awareness of unknown words, at least in incidental learning settings. This finding echoes the claim that the use of multimedia can facilitate L2 acquisition by presenting information in different input modes, which increases the likelihood of noticing linguistic features (Plass & Jones, 2005).

When comparing the awareness rate between groups, most noticing cases were recorded in the captions group (50.98%), followed by the bilingual subtitles group (41.88%), whereas the L1 subtitles group only noticed 28.65% of unknown TWs. This result provides empirical evidence to support the Modality Principle in Multimedia Learning (Low & Sweller, 2005). According to this principle, learners' working memory capacity is believed to be effectively expanded by presenting the same information in a mixed visual and auditory mode, especially when information in a single mode is not understandable by itself. Therefore, presenting L2 unknown words in both written and auditory forms, which was the case in the captions and bilingual subtitles groups, might increase learners' noticing of unknown words and facilitate learning. The lower awareness rate in the bilingual subtitles group compared to the captions group also seems to suggest that the presence of L1 translations might potentially distract learners' attention away from L2 unknown TWs. The lowest awareness rate reported in the L1 subtitles group, supporting earlier claims that the use of L1 subtitles may prevent learners concentrating on spoken L2 and so bypass the L2 spoken form of words (e.g., Lambert & Holobow, 1984; Peters, 2019). Therefore, although the presence of an L1 line in bilingual subtitles seemed to lead to less noticing of L2 words than in the captions group, it still led to more noticing than in the L1 subtitles group.

### ***5.3.2. Level 2: Vocabulary Processing Strategies***

#### **5.3.2.1. Vocabulary Processing Strategies Across Groups**

Concerning participants' vocabulary processing strategies, the results show that the participants used various types of strategies to engage with unknown TWs that appeared

only once in the video. The findings support the reading study by Rott (2000), where learners were found to engage and infer unknown word meanings even on their first encounter during reading. Similar to previous reading studies (e.g., Fraser, 1999; Hu & Nassaji, 2012; Rott, 2000), participants in the present research also used strategies independently (89.12%) as well as in combination (10.88%) to construct their knowledge of unknown words. Especially, the bilingual subtitles group had the highest number of strategy combination cases, with using L1 translations being the most frequently combined strategy. The provision of more input when using bilingual subtitles seemed to encourage the participants to apply and combine more strategies to engage with a novel word.

As can be observed in Table 46, the types of strategies generated in the present study were similar to previous research that explored L2 learners' vocabulary processing/learning strategies during viewing and reading. Comparing to the only viewing study conducted by Syodorenko (2010), almost all the strategies reported there were seen in the present findings with one exception, i.e., "recognizing words that are similar to L1" (p. 61). This can be attributed to the lack of cognates between participants' L1 (i.e., Chinese) and L2 (i.e., English) in the present research. There are some strategies that emerged in the present study that were not recorded by Syodorenko (2010). This may be due to the different research designs and research methods applied. While Syodorenko (2010) used a questionnaire to explore L2 learners' general vocabulary learning strategies during captioned viewing, the present research used stimulated recall targeting each TW during subtitled viewing. Therefore, strategies concerning the use of L1 translations were not applicable in her study. Moreover, the use of a questionnaire only allowed participants to report the general strategies they could remember, whereas stimulated recall could capture the specific strategies used for

each unknown word. Additionally, the present findings also overlap with previous reading studies in terms of the strategies used. The major difference between reading studies and the present findings is participants' lack of using multimedia input to engage with unknown words in reading studies, for example, using auditory cues and images.

Table 46. Comparison of the Vocabulary Processing Strategies Generated by the Present Study and Previous Research

| Strategy category in the present study | Strategy subcategory in the present study | Vocabulary processing/learning strategy in previous reading and viewing studies                                                                                                                                      |
|----------------------------------------|-------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Word feature analysis               | 1.1. Analysing part of speech             | Morphological analysis; Mistaken ID (Huckin & Bloch, 1993);                                                                                                                                                          |
|                                        | 1.2. Analysing word-structure             | Analyzing; Analogy; Grammatical knowledge; Morphological knowledge                                                                                                                                                   |
|                                        | 1.3. Spelling association                 | (Nassaji, 2003); Word feature analysis (Lawson & Hogben, 1996);                                                                                                                                                      |
|                                        | 1.4. Analysing word pronunciation         | L1/L2 word identification (Fraser, 1999);                                                                                                                                                                            |
|                                        | 1.5. Word usage                           | Breaks TW into its two components; Tried different word categories (Rott, 2000); Form-focused (Hu & Nassaji, 2012); Reading captions; Using the roots of known words; Paying attention to grammar (Sydorenko, 2010). |
| 2. Using context                       | 2.1. Using auditory cues                  | \                                                                                                                                                                                                                    |
|                                        | 2.2. Using images                         | Matching visual images with words (Sydorenko, 2010).                                                                                                                                                                 |
|                                        | 2.3. Using global understanding           | Used context (Huckin & Bloch, 1993); Verifying; World knowledge; Discourse knowledge (Nassaji, 2003);                                                                                                                |
|                                        | 2.4. Using local contextual cues          | Parallelism; Sentence-bound cues;                                                                                                                                                                                    |

| Strategy category in the present study  | Strategy subcategory in the present study                                                           | Vocabulary processing/learning strategy in previous reading and viewing studies                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                         |                                                                                                     | Complex elaboration (Lawson & Hogben, 1996);<br>Sense creation (Fraser, 1999);<br>Inferences using immediate context; Use of grammatical knowledge; Use of background knowledge; Elaborating on the context; Lexically correct inferences; Conceptual inferences (Rott, 2000);<br>Searching for meaning in the context of the TW; Accessing background knowledge to make meaning of the TW (Rott, 2005);<br>Meaning-focused (Hu & Nassaji, 2012);<br>Reading captions; Paying attention to verbal context; Paying attention to grammar (Sydorenko, 2010). |
| 3. Using L1 translations                | 3.1. Using L1 matched<br>3.2. Using L1 not matched                                                  | Consult (a dictionary) (Fraser, 1999);<br>Referring to the glosses; Using existing knowledge sources to retrieve a synonym of the L1 gloss (Rott, 2005);<br>L1 knowledge (Nassaji, 2003)                                                                                                                                                                                                                                                                                                                                                                  |
| 4. Guessing without reported strategies | 4.1. Meaning fully guessed<br>4.2. Meaning partially guessed<br>4.3. Meaning unsuccessfully guessed | Simple elaboration (Lawson & Hogben, 1996);<br>Circumlocution of the meaning of the TW (Rott, 2000).                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 5. Other strategies                     | 5.1. Pretest impact<br>5.2. Want to use a dictionary<br>5.3. Visualizing                            | \                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

| Strategy category in the present study | Strategy subcategory in the present study | Vocabulary processing/learning strategy in previous reading and viewing studies                                                                                                                                  |
|----------------------------------------|-------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6. No reported strategies              | 6.1. No reported meaning guessed          | Potholes (Huckin & Bloch, 1993);<br>Ignore (Fraser, 1999);<br>Demonstrates awareness of TWs but does not infer meaning (Rott, 2000);<br>Monitor (Rott, 2005);<br>Monitoring (Hu & Nassaji, 2012; Nassaji, 2003). |
|                                        | 6.2. Forgot thoughts                      | \                                                                                                                                                                                                                |

In the present research, the most frequently reported strategy categories were using L1 translations (only in the L1 and bilingual subtitles groups), noticed words but did not report strategy to guess/link (i.e., “6. no reported strategy”), and using various types of contexts (e.g., images and contextual cues). In terms of using L1 translations, although no previous studies have explored learners’ vocabulary processing strategies during L1 subtitled viewing, research has shown that, when L1 is available during reading (e.g., using a dictionary or L1 glosses), learners frequently make use of L1 when encountering unknown words as a tool to aid their vocabulary learning and understanding (e.g., Fraser, 1999; Ouyang et al., 2020; Rott, 2005). Moreover, L1 translation is also believed to be the most effective method to support understanding and help to build up an initial form-meaning link (Nation, 2003; Schmitt, 2008), which could explain learners’ preference for using their L1 to check unknown words during viewing in those conditions where L1 lines were available.

The second most frequent category across groups was noticing unknown words but not reporting any strategy to guess/link. In this category, participants reported their noticing of a TW as an unknown word but did not report their guess or thoughts

concerning the meaning or form of the word. This could indicate that either participants did not use any strategy during viewing, or they did not remember. Similar categories have emerged in reading research but been named differently, such as “ignore” by (Fraser, 1999), which was distinguished from “no attention” (i.e., not noticed) and “infer (i.e., inferred word meaning)” (p. 229). Rott (2000, p. 261) labelled it as “demonstrates awareness of TWs but does not infer meaning”. A similar situation was later named as “monitor”, under “meta-cognitive processing” strategy, which was defined as “readers *monitored* their comprehension of the TW by mentioning that they were unsure of its meaning saying ‘I am not sure’” (Rott, 2005, p. 103). This strategy can be considered an exemplar of shallow processing according to the Depth of Processing Hypothesis ( Craik & Lockhart, 1972; Leow, 2015), as TWs were only registered at the noticing level, but no further engagement or cognitive effort was involved. Interestingly, compared to other strategies, this strategy has not been frequently mentioned in reading studies where vocabulary learning was the aim of reading (e.g., Fraser, 1999; Rott, 2005). Its frequent occurrence in the present research also attests to the nature of incidental learning, in which the main aim was comprehension rather than language learning (Hulstijn, 2001, 2003), and learners paid more attention to content rather than learning vocabulary.

Another frequently mentioned category was using context, which supports previous reading research findings (e.g., Fraser, 1999; Lawson & Hogben, 1996; Rott, 2000, 2005). However, it should be noted that *context* in this study includes both written/auditory contextual cues and images, which is different from reading research where *context* only refers to the written verbal context. When images were available as part of language input, they became a strong competitor for verbal context in supporting learners’ understanding of unknown words. Learners frequently used imagery to support

their understanding of unknown words during viewing. This finding echoes Syodorenko (2010), who found that using images was the most frequently used strategy to engage with unknown words during captioned viewing. It also provides evidence to support previous claims suggesting that on-screen imagery can benefit L2 learners' incidental vocabulary learning (e.g., Peters, 2019; Rodgers, 2018). The frequent use of images could also imply that the presentation of on-screen text, even with bilingual subtitles, did not prohibit viewers' processing of images, which supports previous viewing research (e.g., Lazareva & Loerts, 2017; Lwo & Lin, 2012).

#### **5.3.2.2. Vocabulary Processing Strategies: Between-Group Comparisons**

This section discusses the comparisons between the bilingual subtitles and two monolingual subtitles groups separately. When comparing the most frequently used strategies in the bilingual subtitles and captions group, four main differences were revealed. First, as expected, the use of L1 translations was the most frequently used strategy in the bilingual subtitles group when encountering an unknown word, whereas it was not recorded in the captions group, as they were not available in that group. In addition, participants in the captions group were not always able to articulate the strategies they used to guess the meanings of words (i.e., 4. Guessing without reported strategies), and this strategy did not emerge in the bilingual subtitles group. Third, although both groups frequently used the context (including auditory cues, imagery, and verbal context) to engage with unknown TWs, this strategy was reported much more frequently in the captions group (43.22% vs. 13.79% in the bilingual subtitles group). Especially, the bilingual subtitles group was less likely to infer word meanings using global understanding (2.15%) or local contextual cues (0%). In the reading context, Fraser (1999) found that L2 learners used inferring more than consulting a dictionary to

understand the meanings of unknown words. However, it is important to note that checking the L1 translation in a dictionary while reading involves extra actions, a higher level of intentionality, and interruption of the reading process, which is different from using L1 lines which were already available during viewing.

The last difference lay in the lower reported rate of not using strategies to engage with noticed unknown words in the bilingual subtitles group (10.75%) compared to the captions group (26.27%). This suggests that the presence of both L1 and L2 (i.e., using bilingual subtitles) led to more strategy use. However, when using captions, participants had a relatively higher chance of noticing unknown words but not engaging with word meanings by ignoring or giving up guessing. As shown in Example 2, when using captions, participant A16 noticed the TW *traumatised*, but thought the word looked complicated and thus did not attempt to guess. This happened frequently, especially during viewing where the presentation time of unknown words was limited.

### **Example 2.**

---

R: 这个“*traumatised*”。当时有没有注意到这个词？[This one “*traumatised*”.

Did you notice this word at that time?]

A16: 当时注意到这个词了，觉得好复杂，看不懂。[I noticed this word at that time. I thought it was very complicated and I couldn’t understand it.]

---

The aforementioned group differences between bilingual subtitles and captions indicated that with the presence of L1 translations in the bilingual subtitles group, participants were more likely to apply strategies to engage with unknown words, and they would reasonably turn to the use of L1 translations as a shortcut to facilitate their understanding of unknown words. However, they were less likely to infer unknown TWs’ meanings using global understanding or local contextual cues. When L1

translations were not available (i.e., using captions), participants were less likely to apply strategies to engage with unknown words, and if they did, they seemed to rely more on the verbal context or on other unmentioned sources.

Comparisons were then made between bilingual and L1 subtitles. There were great similarities between these two conditions in terms of the types of vocabulary processing strategies used. The three most frequent categories were the same for both groups (i.e., 3. Using L1 translations, 6. No reported strategies, and 2. Using context). However, participants in the two subtitling groups differed in their frequency of using these strategies. First, although using L1 translations was the most frequent strategy and, on most occasions, participants could successfully match L1 translations to L1 unknown TWs in both groups, this strategy was much more frequently reported in the bilingual subtitles group (62.37% vs. 46.55% in L1 subtitles). This suggests that the presence of both L1 and L2 lines during viewing encouraged participants' use of the translations to understand unknown words. This finding also echoes the comment from a participant in Li's (2016) study when using bilingual subtitles: "I can compare with the two lines of languages. I use L1 subtitles for getting the meaning and L2 subtitles for getting the words' spelling" (p. 195). Second, in the L1 subtitles group, there were twice as many cases as in the bilingual subtitles group where participants noticed an unknown word but did not report any strategies to engage with it. This indicates that when the written forms of L2 TWs were available, learners were more likely to engage with the meanings of unknown words and attempt to match L1 translations to L2 unknown words, rather than simply registering them without engaging with their meanings.

When looking in depth into participants' use of L1 translations, further subcategories revealed another difference between these two groups. When using bilingual subtitles, L1 translations could also serve as a trigger for learners' referring to

L2 unknown words, whereas this was not reported when using L1 subtitles. On the contrary, some participants using L1 subtitles reported that L1 translations triggered their memory of other L2 words. Due to L2 written forms not being available, learners were likely to mistakenly match a familiar L2 word to a presented L1 translation, resulting in paying less attention to the L2 auditory form or bypassing the auditory L2 input (Peters, 2019). However, when using bilingual subtitles, learners could access the written forms of both L1 and L2 at the same time, which increased their chances of analysing word features and linking auditory forms with their written forms.

These group differences between the bilingual and L1 subtitles groups indicate that while participants in both conditions relied on L1 translations of TWs, bilingual subtitles led to an even greater reliance on L1 translations and also enabled learners to have more engagement with L2 word forms, which was more likely to facilitate the initial establishment of form-meaning connections for unknown words.

### ***5.3.3. A Third Level of Engagement that Emerged: Intentionality of Learning***

Although stimulated recall was not designed to probe learners' intentionality to learn each TW, there were some cases (10.46%) where participants explicitly commented on their intentionality to learn TWs while viewing. It is important to acknowledge this aspect of learners' engagement with TWs since it is closely related to the *affective* aspect of engagement. As discussed in section 2.5.1, according to Svalberg's (2009) concept of engagement, an engaged individual has a positive, purposeful, willing, and autonomous disposition towards the object. Also, intentional learning could lead to greater and faster vocabulary gains with better learning retention compared to learning without intention (Schmitt, 2008). Therefore, learners'

intentionality to learn unknown words during viewing can also affect their learning gains.

The present findings provide empirical evidence to support the claim that intentional and deliberate learning can occur in incidental learning conditions (Hulstijn, 2001, 2003). However, it should be noted that the number of cases showing participants' intentionality of learning across conditions was very small (5.86%). This suggests that, while there may be some intentionality in incidental learning conditions, participants still focused mostly on comprehension, as intended by the design of the learning condition.

When looking at each subtitling group, participants in the bilingual subtitles group commented more on their intentionality to learn unknown vocabulary than the other two monolingual groups. This suggests that bilingual subtitles might be regarded as an English learning tool which is more likely to raise learners' vocabulary learning intention. This also supports Li's (2016) finding that bilingual subtitles have been ranked as the most useful subtitling type to facilitate L2 learners' vocabulary learning. However, it is important to note that this conclusion was based on a small number of cases and further research should be conducted to test this claim.

In summary, the stimulated recall findings suggest that regarding learners' awareness of unknown TWs (Level 1), participants in general noticed 40.72% of unknown TWs during subtitled viewing in the incidental learning context. The bilingual subtitles group noticed fewer unknown words than the captions group but more than the L1 subtitles group. Concerning the use of vocabulary learning strategies (Level 2), learners applied six general types of strategies (with 23 specific strategies) to engage with unknown words during subtitled viewing, with using L1 translations and using context being the most frequently reported strategies. Participants also frequently

reported that they noticed unknown TWs but did mention any strategies to engage with them. Various strategies were used in different subtitling conditions. The bilingual subtitles group shared more similarities with the L1 subtitles group than with the captions group. Comparing to monolingual subtitles, learners using bilingual subtitles more frequently used L1 translations to engage with unknown words, and more successful meaning matching cases were reported. The results provided empirical evidence showing the potential of bilingual subtitles to combine the advantages of monolingual subtitles. L2 written forms increased learners' noticing of unknown words compared to using L1 subtitles, and L1 translations helped learners to establish accurate initial form-meaning connections. However, the bilingual subtitles group was less likely to use the verbal context to engage with unknown words compared to the captions group. Concerning the level of analysis that emerged (i.e., Level 3: Intentionality of learning), more cases of intentional learning were reported in the bilingual subtitles condition, suggesting that bilingual subtitles may have the potential to increase learners' vocabulary learning intention during viewing.

#### ***5.3.4. Triangulation of Qualitative and Quantitative Findings***

RQ6 aimed to combine and compare stimulated recall findings with previous quantitative results to paint a more thorough picture of how learners' report engagement with unknown words relates to their incidental vocabulary learning gains and how it relates to their eye movements during subtitled viewing. Triangulation of participants' stimulated recall findings is first done with their vocabulary tests results, followed by triangulation with participants' eye movements results.

#### **5.3.4.1. Stimulated Recall Findings vs. Vocabulary Test Results**

The stimulated recall findings corroborated the vocabulary test results reported in section 4.2.1. The vocabulary test results indicated that the captions group significantly outperformed the other groups on form recognition. Similarly, the stimulated recall findings showed that participants in the captions group noticed more unknown TWs than the bilingual and L1 subtitles groups. This suggested that the higher level of awareness reported in the stimulated recall data was indeed reflected in greater form recognition gains.

In terms of learning word meanings, the bilingual subtitles group significantly outperformed the captions group on meaning recognition. This advantage of bilingual subtitles in facilitating meaning knowledge can be explained by the vocabulary processing strategy used. The bilingual subtitles group was more likely to apply strategies to engage with unknown TWs than the captions group, whereas the captions group had a relatively higher chance of noticing unknown words but not using strategies to engage with word meanings. Moreover, the bilingual subtitles group used L1 translations largely to construct their knowledge of unknown words while viewing. This may explain the general better performance of bilingual subtitles than captions in learning word meanings. This suggests that the frequent use of L1 translations did indeed facilitate learners' recognition of word meanings. With the presence of both L1 and L2, bilingual subtitles could facilitate the initial establishment of a form-meaning link, which was reflected in recognition tests. This finding also supports the claim that using L1 translations is an effective and beneficial method to establish an initial form-meaning link (Schmitt, 2008), which also echoes previous studies showing the benefits of using L1 glosses for incidental learning gains in both reading (e.g., Ouyang

et al., 2020; Teng, 2019) and viewing (e.g., Montero Perez et al., 2018). The advantages of L1 for facilitating vocabulary learning have also been noted by Nation (2003):

This [study always found L1 translation is the most effective] is probably because L1 translations are usually clear, short and familiar, qualities which are very important in effective definitions (McKeown 1993). When the use of an L1 translation is combined with the use of word cards for the initial learning of vocabulary, then learners have a very effective strategy for speeding up vocabulary growth. (p. 4)

However, it should be noted that the advantage of bilingual subtitles was not significant over captions in meaning recall. As explained in section 4.3.1, the form-meaning links established by referring to L1 translations in bilingual subtitles might have been too subtle to be observed in a recall test. This potential reason seems to be confirmed by the stimulated recall findings. Despite the fact that the captions group tended to apply fewer strategies to engage with unknown words, in cases where they did, they seemed to make more use of contextual cues, which might have led to deeper processing of word meanings. According to the Depth of Processing Theory (Leow, 2015), information can be better retained in memory with a larger amount of cognitive effort and higher level of analysis involved. However, according to Leow (2015), the use of L1 translation in vocabulary learning can be considered a lower depth of processing, leading to simple and shallow engagement with unknown words, compared to elaborations and deep cognitive processes in guessing word meanings. This echoes Hu and Nassaji's (2012) findings showing that when the meaning of a novel word is easily understood, L2 learners are likely to pay less attention to the word form, leading to weak form-meaning connections which may negatively affect word retention. In the present study, the different depth of processing was revealed by comparing the engagement of two participants with the same TW *surrogate* when using captions and bilingual subtitles as examples:

### Example 3. [Captions group]

---

A13: 就是图片中这个鸭子是跟着猫的。后面是“mum”，但是有个修饰词，所以这个修饰词就是。因为不是它们真的母亲，所以然后视频一直在讲它小的时候。就是这个鸭子小的时候，它被猫抚养，所以就应该是，应该是这个意思，就是被抚养的，就给了生命的妈妈这种，抚养它们的妈妈。 [In the picture, the duck is following the cat. There is a “mum” followed, but there is a modifier, so this modifier is [the word]. Because it’s [the cat] not their [the ducks’] real mother, and the video kept talking about when it [the duck] was young. When the duck was young, it was raised by the cat, so it should be, it should mean this, it should mean being raised, the type of mother who gave them life, the mother who raised them.]

---

### Example 4. [Bilingual subtitles group]

---

C4: 啊这个看了，这个就知道，就知道了它是“*surrogate*”是代理。 [Ah, I looked at this, when I looked at this I knew it’s “*surrogate*”, it means surrogate [in Chinese].]

---

As can be observed in Example 3, participant A13 in the captions group used images, local contextual cues, and a global understanding to infer the meaning of the TW *surrogate* during viewing, whereas in Example 4, participant C4 in the bilingual subtitles group used L1 translations to easily determine the meaning of the unknown word. The instant understanding of word meaning may involve a relatively low level of cognitive effort. However, when L1 translations were not available, learners attempted to generate word meanings which required a higher level of cognitive effort. Once a successful guess is made, it seems to be more likely to lead to better and longer vocabulary retention than simply checking L1 translations. However, it should be noted that, in general, the captions group was less likely to apply strategies to engage with unknown words, which limited their learning gains. In sum, the use of L1 translations

can be efficient to build up initial form-meaning connections which can be manifested in a recognition test but may not be strong enough to be captured in a recall test.

When comparing vocabulary test results between the bilingual and L1 subtitles groups, no significant group differences were revealed in terms of form and meaning recognition. This is supported by the similar types of processing strategies applied in these two subtitling groups. However, the bilingual subtitles group demonstrated a significant advantage over the L1 subtitles group on meaning recall and showed a superiority approaching significance in meaning recognition. This could be explained by the higher number of cases of noticing recorded in the bilingual subtitles group. Since the meaning recall test required learners to first recognise the form of TWs before providing translations, the provision of L2 written forms while viewing could facilitate learners' mapping of L1 translations to L2 word forms. Moreover, although using L1 translations was the most frequently used strategy in both groups, there were more successful matching cases recorded in the bilingual subtitles group. In addition, more strategy combination cases were reported in the bilingual subtitles group than in the L1 subtitles group, which may imply a deeper engagement being involved in the bilingual subtitles group compared to the L1 subtitles group. These findings echo previous research showing that a combination of different processing strategies leads to higher meaning recall gains than using strategies individually (e.g., Fraser, 1999; Hu & Nassaji, 2012; Lawson & Hogben, 1996).

#### **5.3.4.2. Stimulated Recall Findings vs. Eye-movement Findings**

As reported in section 4.3.3.3, eye-movement data revealed that the bilingual subtitles group spent significantly less time on L2 unknown TWs than the captions group, and significantly more time was spent on L1 translations of unknown TWs than

the L1 subtitles group. Stimulated recall data also showed that the bilingual subtitles group noticed fewer unknown TWs than the captions group but more than the L1 subtitles group. Thus, the greater attention paid to TWs seems to have been reflected in more noticing. It is important to note that the eye-tracking data only revealed the attention learners paid to the written form of TWs or L1 translations, which did not take into account participants' noticing of TWs in auditory form. Stimulated recall data compensated for this limitation of eye-tracking in viewing research by examining learners' noticing of TWs in both written and aural forms. This correspondence between learners' eye-movement findings and their reported awareness at group level echoes Godfroid and Schmidtke (2013), who found a significant, positive relationship between learners' attention (as measured by eye movements) and awareness (as measured by verbal report) during L2 reading. In their study, pseudowords that were reported as noticed during reading were associated with longer processing times than unnoticed ones.

Comparing the bilingual and L1 subtitles groups, eye-movement data revealed that the bilingual group spent significantly more time on L1 translations of unknown TWs than the L1 group. Stimulated recall reported that more unknown TWs were noticed in the bilingual subtitles group than in the L1 subtitles group. Taken together, these findings seem to suggest that the presentation of both written L1 and L2 led to more noticing of unknown TWs than presenting written L1s alone, and increased noticing was also reflected in the increased amount of attention learners paid to L1 translations of unknown words. The longer processing times for L1 translations in bilingual subtitles could signal learners' attempts to check the meanings of unknown words after noticing them. This may also be supported by the vocabulary processing strategies learners reported in these two groups. The bilingual group reported more cases of successfully

matching L1 translations to unknown TWs than the L1 group, while the L1 subtitles group reported more cases of not using strategies to engage with words or more cases of wrongly matching other L2 words to L1 translations of TWs than the bilingual subtitles group.

The stimulated recall findings also offer an explanation for the discrepancy between learners' processing time for L2 unknown TWs and their vocabulary gains. As reported in section 4.3.4, the longer second-pass reading time for unknown TWs in the present study did not lead to higher learning gains, which agrees with the findings of Montero Perez et al. (2015). It has been conjectured that a longer second-pass reading time can reflect either readers' "increased intention to commit the word to memory" (Montero Perez et al., 2015, p. 324) or processing problems caused by readers' incomplete lexical integration processes (Montero Perez et al., 2015; Rayner, 1998). However, these different cognitive processes cannot be distinguished by merely referring to eye-movement data (Montero Perez et al., 2015; Pellicer-Sánchez, 2020a). The present stimulated recall findings provide empirical evidence to support Montero Perez et al.'s (2015) speculation by revealing a great variety of strategies that learners used to engage with unknown words during viewing. Moreover, in this incidental learning setting, there was a considerable percentage of cases (20.82%) where participants noticed unknown words during viewing but did not use any strategies to engage with them. In addition, the stimulated recall results showed that not all the strategies used could lead to successful inferencing or matching between L1 translations and L2 forms, as also pointed out by other researchers (e.g., Huckin & Bloch, 1993; Lawson & Hogben, 1996; Nassaji, 2003). A similar issue was also reported by Syodorenko (2010), where participants reported a high frequency of using a guessing strategy during captioned viewing but demonstrated uncertainty about the outcome of

their guess. Therefore, a longer processing time could have encompassed different processing strategies, and some of these strategies might not necessarily lead to successful learning.

As presented in Example 5, participants A14, A16, and A27 in the captions group all spent a relatively long second-pass reading time on the TW *confiscated*, at 1,208ms, 911ms, and 763ms, respectively. However, different processing strategies were applied by different participants. A14 did not notice the TW, A16 noticed the TW but did not use any strategies to engage with it, whereas A27 noticed the TW and also partially successfully inferred its meaning during viewing. These different processing strategies were reflected in similar second-pass reading times, which might have led to different learning gains and caused discrepancies between learners' attention and learning outcomes.

**Example 5. [Captions group]**

---

A14: 我没有注意它，可能我看它，但是我没注意它。当时好像跳过去了，我的注意力。[I didn't pay attention to it, maybe I saw it, but I didn't pay attention to it. At that time it seemed to be skipped, my attention.]

A16: 对，看到这个词不熟。不熟，不知道什么意思。[Yes, I saw this word but it wasn't familiar. I'm not familiar with it, I didn't know what it means.]

A27: 这个我不认得。当时我就在想我在猜这个词是不是关押。[I don't know this word. At the time, I was thinking I was guessing if this word was imprisoned.]

---

Another discrepancy between eye-movement data and vocabulary learning gains lies in the lack of a significant relationship between the processing time of L1 translations in bilingual subtitles and vocabulary learning gains as reported in section

4.3.4. The stimulated recall data showed that the longer processing times for L1 translations in bilingual subtitles did not necessarily indicate that participants were actively making form-meaning links. Instead, different types of engagement with unknown words were involved. The longer processing times for L1 translations could indicate that learners were simply taking in information without establishing form-meaning links. As shown in Example 6, participant C27 spent a total of 1,267ms reading time on the L1 translation of the TW *foal* but 0ms on the L2 form, and also reported no awareness of the TW. As revealed in other cases, the longer processing times for L1 could also indicate learners' simple matching of L1 translations and L2 words without deep engagement (as shown in Example 7). Moreover, not all successful matches necessarily lead to vocabulary learning and retention, especially when most of the unknown TWs only appeared once in the viewing material.

**Example 6.** [Bilingual subtitles group]

---

R: 好，这个 “*foal*” 。 [OK, this one “*foal*”.]

C27: 嗯，没，好像没怎么注意这个词。 [Hmm, no, I did not seem to pay much attention to this word.]

---

**Example 7.** [Bilingual subtitles group]

---

R: 这个， “*traumatised*” 。 [This one, “*traumatised*”.]

C17: 我当时想的是这个词很长，然后但是我看了一下中文，我觉得，嗯就了解了，就没有去思考英文，对。 [What I thought at the time was that the word was very long, but then I looked at Chinese, I felt that, um, I understood it, so I didn't think about the English, right.]

R: 这个， “*barneys*” 。 [This one, “*barneys*”.]

---

---

C26: 嗯对我注意了。我就也是之前不认识，然后看了一下对应的（中文翻译）。

[Yes, I noticed it. I didn't know this one before, then I checked its corresponding (Chinese translation).]

---

Therefore, the lack of relationship between learners' attention and their vocabulary learning gains could be attributed to the mixed cognitive processes underlying participants' eye movements as revealed in the stimulated recall findings. Moreover, these different processing strategies did not necessarily account for successful inferencing or learning gains that could be reflected in higher test scores.

In summary, the stimulated recall findings presented in this chapter also support the quantitative findings of the vocabulary tests and eye-movement data. The stimulated recall findings provided explanations of the vocabulary test results. The higher awareness rate reported in the captions group explained their better performance on form recognition. The bilingual subtitles group was found to use L1 translations more frequently and successfully than other groups, which could explain their higher meaning gains. However, participants using bilingual subtitles were found to be less likely to use verbal contextual cues to infer word meanings compared to the captions group, which was reflected in the lack of superiority of bilingual subtitles over captions in meaning recall. Stimulated recall also compensated for the limitations of eye-tracking by revealing participants' noticing of unknown TWs during viewing. The bilingual subtitles group was found to notice more unknown TWs than L1 subtitles, but fewer than the captions group. The stimulated recall data also showed that participants engaged in a number of different strategies, which confirms previous claims that different cognitive processes are involved in eye-movement data and in the attention paid to lexical items. The inconclusive relationship between attention and learning gains can thus be explained by the fact that not all strategies necessarily lead to successful

guesses and learning gains.

## **Chapter 6. General Discussion and Conclusion**

This research was motivated by the widespread use of bilingual subtitles among Chinese learners of English. The main purpose of this study was to test the relative effectiveness of bilingual subtitles on L2 learners' incidental vocabulary learning, in comparison to captions, L1, and no subtitles. The relative effects of bilingual subtitles on comprehension were also examined. Most previous studies used offline tests to examine learners' vocabulary learning outcomes, but few of them explored learners' cognitive engagement with unknown words during subtitled viewing. As such, a secondary purpose of this study was to investigate L2 learners' processing of different sources of input in bilingual subtitles, in comparison to other subtitling conditions, especially their engagement with unknown TWs during subtitled viewing with the help of eye-tracking and stimulated recall.

This chapter starts by summarising the main findings of the study. The results and discussion concerning each RQ have been presented in Chapters 4 and 5, so the general discussion is organised based on specific topics rather than each RQ. The chapter then presents a discussion of the theoretical, methodological and pedagogical implications of the findings. It finishes with a discussion of the limitations of the present study and suggestions for future directions.

### **6.1. Summary of the Main Findings**

**1. The effects of bilingual subtitles on L2 learners' incidental vocabulary learning (RQ1) and comprehension (RQ2), compared to captions, L1 subtitles, and no subtitles.**

One of the aims of the present study was to compare the effects of bilingual subtitles on incidental vocabulary learning and comprehension. A pretest-posttest experimental design was employed to examine learners' learning gains after viewing with different subtitling types. Overall, the results demonstrate the benefits of using bilingual subtitles to facilitate word meaning knowledge and comprehension. Bilingual subtitles seemed to be less effective than captions in form recognition, but more beneficial for learning word meanings compared to other subtitling types. In addition, their benefits for vocabulary learning did not overshadow their superiority in comprehension, which led to significantly better comprehension than captions and no subtitles. Learners with a larger vocabulary size also tended to achieve higher vocabulary and comprehension scores. Moreover, word length and frequency of occurrence predicted meaning recall and meaning recognition gains, respectively.

## **2. L2 learners' attention distribution towards different areas during viewing using bilingual subtitles, compared to captions, L1 subtitles, and no subtitles (RQ3).**

The eye-tracking data demonstrated that both the captions and bilingual subtitles groups tended to spend more time processing the subtitling area than the image area, while the L1 subtitles and no subtitles groups spent more time on images. In terms of processing the subtitling area, presenting more input did not necessarily lead to longer processing times. Learners spent similar amounts of time on the subtitling area when using bilingual subtitles (60% of total reading time) and captions (64%), which were both significantly longer than when using L1 (44%) and no subtitles (4%). By looking in detail at learners' use of L1 and L2 lines in bilingual subtitles, the participants were found to use both lines, but the time spent on L1 lines (42%) was double than that spent on L2 lines (20%). Moreover, the time spent on L1 lines was even longer than that in

the L1 subtitles group (34%), whereas the time spent on L2 lines was shorter than that in the captions group (56%). These findings suggest that when using bilingual subtitles, learners rely more on L1 lines but still selectively process L2 lines. In addition, learners' vocabulary size had a negative effect on their processing of captions but no effects on reading the bilingual and L1 subtitles.

### **3. L2 learners' engagement, including attention (RQ3), awareness (RQ5), and vocabulary processing strategies (RQ5), with unknown TWs during viewing using bilingual subtitles, compared to captions and L1 subtitles.**

In order to investigate L2 learners' engagement with unknown words during viewing, learners' attention to each unknown TW was measured through various eye-movement measures while viewing. Their awareness of and processing strategies for each unknown TW were also examined using stimulated recall after viewing. Eye-movement data showed that when they encountered unknown TWs, learners using bilingual subtitles spent significantly longer on the L1 translations of unknown words than on L2 word forms. Moreover, words with higher frequency of occurrence were more likely to be attended to. Their processing time for L2 unknown words was shorter than that in the captions group, but more time was spent on L1 translations of unknown words than those who used L1 subtitles. This suggests that the use of bilingual subtitles led to increased attention to L1 translations, compared to only L1 subtitles being presented. The presentation time and word length of unknown TWs showed positive effects on learners' processing time.

Similar patterns of group differences were observed when looking at the stimulated recall data. In terms of learners' awareness of unknown TWs, learners in the bilingual subtitles group noticed fewer unknown TWs than those using captions. However, the

use of bilingual subtitles led to more noticing of unknown TWs than L1 subtitles. The presence of L1 input seemed to have distracted the learners from L2 unknown words, resulting in less noticing of unknown words when using bilingual and L1 subtitles compared to using captions.

Learners' processing strategies used to engage with each unknown TW were also explored. Learners were found to use a variety of processing strategies during subtitled viewing. The most frequent common strategies used across groups were using L1 translations (for the bilingual and L1 subtitles groups), using context, and not reporting any strategy to engage with noticed words. Results also flagged up group differences. When using captions, inferring word meanings based on context (especially verbal context) was the most frequently used strategy. For the bilingual and L1 subtitles groups, using L1 translations to check the meanings of unknown words was the most frequent strategy. Bilingual subtitles were more likely to trigger learners to apply strategies, and more form-meaning matching cases were reported in the bilingual subtitles group than in the monolingual subtitles groups.

#### **4. Potential relationship between L2 learners' engagement with unknown TWs and their incidental vocabulary gains (RQ4 & RQ6).**

To further explore the relationship between learners' engagement with unknown words and their learning gains, vocabulary test results, eye-movement data, and stimulated recall findings were triangulated and discussed. In general, learners' attention to L2 unknown words (but not L1 translations) could to some extent predict their vocabulary learning gains, despite the fact that this relationship was not consistent in all eye-tracking measures. This inconclusive relationship might be attributed to learners' different underlying cognitive processes during viewing. This was confirmed by the

stimulated recall data, revealing a variety of vocabulary processing strategies reported by the participants, and not all vocabulary processing strategies led to successful guessing. The stimulated recall data also supported the groups noticing more unknown words achieving higher form recognition scores. The better performance of the bilingual subtitles group on meaning-related tests was related to their higher noticing of unknown words than the L1 subtitles group. Besides, their more frequent use of L1 translations to refer to word meanings than the captions group also helped to establish initial form-meaning connections.

## **6.2. Implications of the Study**

### ***6.2.1. Theoretical Implications***

The results of the present thesis have a number of theoretical implications. First, the findings highlight the importance of attention and further confirm its complex relationship with vocabulary learning gains. Although the present study is situated in the viewing context, which is different from most of the previous eye-tracking studies focusing on reading (e.g., Godfroid et al., 2018; Godfroid & Schmidtke, 2013; Mohamed, 2018; Pellicer-Sánchez, 2016; Williams & Morris, 2004), the results also support previous findings showing that attention (operationalised as eye movements) is a vital predictor of vocabulary gains. By using three vocabulary tests and various early and late eye-tracking measures, this study has shown that attention has a positive correlation with learning word forms, echoing previous reading and viewing studies (e.g., Godfroid & Schmidtke, 2013; Mohamed, 2018; Montero Perez et al., 2015). However, more attention being paid to a word does not always guarantee successful learning, since the time to reread L2 words (i.e., second-pass reading time) did not

result in higher form or meaning scores. Moreover, the predictive role of attention in learning word meanings was inconclusive. Stimulated recall data revealed that learning word meanings might also relate to learners' underlying cognitive processes with words rather than just the amount of attention. In other words, it was what learners did and thought about words, rather than how long they looked at them, that might determine their learning of word meanings, at least in the incidental learning context.

Second, the present research has addressed an important but under-researched concept in vocabulary learning, i.e., engagement. The present thesis adapted Svalberg's (2009) definition of *engagement with language* to vocabulary learning, defining engagement as a construct consisting of three main aspects: cognitive, affective, and social aspects. With a particular focus on the cognitive aspect, the present study has operationalised cognitive engagement as attention, awareness, and vocabulary processing strategies. This study contributes to our understanding of engagement in incidental vocabulary research and provides a framework for future studies. The findings of the present research echo Godfroid and Schmidtke (2013), demonstrating a close relationship between learners' attention and awareness. Furthermore, this research has filled a research gap by probing learners' vocabulary processing strategies during subtitled viewing. The findings revealed various processing strategies that learners used to construct their knowledge of unknown words during viewing with different subtitling types, which enriches our understanding of learners' engagement with novel words during viewing and how they might relate to their learning gains. The list of strategies that emerged from the present research can also inform future studies and be applied in future investigations of vocabulary learning from viewing. To the best of my knowledge, this is the first study to reveal how learners engage with each unknown word in the subtitled viewing context by tapping into learners' attention, awareness, and word

processing strategies together. The findings also highlight the importance of investigating different aspects of learners' engagement rather than merely focusing on attention and awareness to fully understand learners' engagement.

Third, this study contributes to our understanding of the operationalisation of incidental learning. Incidental learning has been criticised as being problematic due to its exclusion of learners' subjective experience during activity (Gass, 1999; Hulstijn, 2001). A learning condition can be incidental but the type of learning that learners engage with may be intentional. The results of this study provide evidence to support that the type of learning that accrued from this incidental learning condition was largely incidental, as revealed in stimulated recall. Although there were a few reported cases where participants intended to learn words, for more than half of unknown words, participants did not notice them. In addition, in 22% of noticed cases, participants did not use any processing strategies to engage with words. Consequently, the operationalisation of incidental learning in the present study, by not informing learners' of upcoming vocabulary tests, did indeed result in learning processes that were largely incidental.

Fourth, by investigating the use of bilingual subtitles, this study has also touched on the controversy of using L1 in vocabulary learning. The present findings show that L1 translations were helpful for the initial establishment of form-meaning connections, as reflected in the high meaning recognition scores, but this connection tended to be weak, as shown in the moderate meaning recall gains, at least in incidental learning from viewing settings. These findings resonate with previous arguments about the use of L1 equivalents. While some researchers have criticised the use of L1 equivalents for encouraging learners' laziness rather than deeply engaging with L2 words, others have suggested the benefit of using L1 lexical transfer to establish a form-meaning link

(Jiang, 2002; Schmitt, 2010). According to the Depth of Processing Hypothesis (Craik & Lockhart, 1972; Leow, 2015), the form-meaning link established by using L1 equivalents may not be strong or resistant enough, resulting in shallower processing of form-meaning connections. However, L1 equivalents can help in establishing an initial correct form-meaning link and reducing the danger of mistaken guessing when using the context (e.g. Prince, 1996; Zou, 2016). This is helpful since “the first step in the vocabulary acquisition process is establishing an initial form-meaning link” (Schmitt, 2008, p. 335), and the use of L1 equivalents can provide tremendous help in setting up this first step. In sum, the use of L1 in L2 learning should never be taken as a clear-cut issue. The use of L1 translations can be facilitative of vocabulary learning, but this benefit can also be constrained without further engagement with words.

Lastly, the findings of the present study have important implications for theories of multimodal and multimedia learning. The results support Dual Coding Theory (Paivio, 1986, 2007) and the Cognitive Theory of Multimedia Learning (Mayer, 2005a), which suggest the benefits of using multimedia to trigger both verbal and imagery systems, facilitating information processing and further enhancing learning outcomes. However, the results also challenge the redundancy principle (Chandler & Sweller, 1991; Sweller, 2005b). According to this principle, the presentation of captions can be regarded as redundant by repeating aural information, which may result in cognitive overload and learning difficulties (Chandler & Sweller, 1991; Mayer et al., 2014; Sweller, 2005b). Moreover, this redundancy may be further maximised when using bilingual subtitles since they also include written L1 input. Nevertheless, care should be taken when interpreting multimedia learning principles for L2 learning contexts since they were originally devised for L1 learning contexts (Montero Perez, 2020b). Numerous studies have shown that in the SLA context, the combination of audio and corresponding

written texts can ease L2 learners' cognitive load by providing supportive written input, which could raise learners' awareness of input, help them decode and segment speech and diminish extraneous cognitive load (Frumuselu, 2018; Montero Perez, 2020b). The present findings support this claim by showing the advantages of using captions and bilingual subtitles for vocabulary learning and comprehension. Moreover, the use of bilingual subtitles did not prevent learners processing images, as revealed by eye-tracking data. Instead of evenly distributing attention to all presented information, participants were found to pay most attention to L1 lines but selectively use L2 lines. Therefore, care should be taken when applying multimedia learning principles in L2 learning contexts. Whether or not input is redundant should always be judged by considering different learning contexts as well as learners' subjective experience and needs.

### **6.2.2. *Methodological Implications***

The present study also has some important implications for research methodology. First, the biggest novelty of this research lies in the systematic analysis of performance measures, eye movements, and stimulated recall data. Pre- and post-vocabulary tests can examine learning outcomes, but they cannot capture learners' attention and learning processes. With the help of eye-movement data, learners' real-time attentional processing can be accurately revealed (Conklin et al., 2018). However, eye-movement data are still limited in disclosing learners' cognitive processes, as they cannot reveal the different subprocesses reflected in eye movements (Montero Perez et al., 2015; Pellicer-Sánchez, 2020a). The added stimulated recall data addressed the limitations of quantitative data by examining whether items had been registered at the level of awareness and what exactly learners did when engaging with items. They also provide

potential explanations for the quantitative findings. This research also addresses the lack of qualitative and mixed methods research on vocabulary, as lately noted by Webb (2020a). By triangulating different data sources to examine the same phenomenon, the inherent weaknesses of each method can be reduced, and complementary strengths can be added, which can maximise both the internal and external validity of research (Dörnyei, 2007; Johnson & Christensen, 2012). The triangulation of three data sources in the present study provides us with different perspectives to examine the same phenomenon, allowing us to tap into not only learning outcomes and the amount of attention that learners paid to unknown words, but also learners' underlying cognitive processes. Triangulation also paints a fuller picture of learners' learning processes and a better understanding of the relationship between engagement and learning outcomes.

In addition, the findings underscore the importance of using different vocabulary tests to examine learners' vocabulary learning gains. There is a consensus that vocabulary acquisition is an incremental process. A word cannot be simply labelled as learned or not learned (Melka, 1997; Nation, 2013; Schmitt, 2000; Webb & Nation, 2017). Therefore, using various tests to measure the knowledge of one word can reveal different kinds of knowledge gained as well as the strength of knowledge gained (Nation & Webb, 2011), which can provide a more accurate evaluation and minimise the risk of underestimating learning gains (Nation & Webb, 2011). By focusing on both form and meaning aspects, and examining knowledge of word meaning both receptively and productively, the effects of different subtitling types can be better captured and distinguished.

The research findings also highlight the benefits of using a variety of eye-tracking measures when investigating the relationship between attention and vocabulary learning. Early measures capture initial processing, which are believed to reflect automatic and

non-strategic reading procedures (Conklin et al., 2018; Godfroid, 2020a), while late measures reveal more controlled cognitive behaviour which can signal interruption or reanalysis following an initial processing difficulty (Godfroid, 2020a). In the present research, although the different measures strongly correlated with each other, different results were revealed in terms of the predictive role of different eye-tracking measures in vocabulary gains. Participants' early processing of words was found to be closely related to the learning of word forms, in line with previous studies (e.g., Mohamed, 2018; Montero Perez et al., 2015). The total reading time for L2 word forms could also to some extent predict the gains but was not consistent for different subtitling groups. These mixed findings could be related to the ineffectiveness of second-pass reading time for predicting learning gains. As a "pure late-processing measure" (Godfroid, 2020a, p. 224) included in total reading time, second-pass reading time signalled the complex subprocesses underlying eye movements, which might not lead directly to learning. The combination of both early and late measures can paint a more complete picture of learners' viewing behaviour, which can provide converging evidence for research (Conklin et al., 2018; Godfroid, 2020a).

An additional methodological implication concerns advanced statistical analysis. Most previous studies used parametric statistics (e.g., t-tests or ANOVA) to analyse quantitative data by using averaged values, i.e., one averaged value for each participant, without taking into account item-level differences. Thus, their findings are less robust for generalisation to different types of target items. Mixed-effects models have the advantage of accommodating nested data with various fixed (i.e., variables that are studied or controlled for) and random variables (i.e., variables resulting from random sampling and affecting outcomes) in a single analysis (Baayen et al., 2008; Cummings, 2012; Godfroid, 2020). Mixed-effects models are also powerful in handling multiple

continuous and categorical independent variables, by taking different variables at the participant level (e.g., learner' pretest scores, vocabulary size) and item level (e.g., frequency of occurrence, word length, part of speech) into account, the results can be better generalised beyond the participants and vocabulary items included in a given study (Baayen et al., 2008; Linck & Cunnings, 2015). Therefore, mixed-effects models should be used when dealing with data collected from randomly selected participants and items in order to produce more robust and generalizable results. The analytical procedures followed in the present study could inform future studies with randomly selected participants and randomly selected target language items.

### ***6.2.3. Pedagogical Implications***

The first pedagogical implication of this research is the advocacy of using audio-visual materials to facilitate vocabulary learning. This study has added empirical evidence, along with numerous previous studies, showing the benefits of watching audio-visual materials for L2 vocabulary acquisition (e.g., Feng & Webb, 2020; Montero Perez, 2020a; Peters & Webb, 2018; Puimège & Peters, 2020; Rodgers & Webb, 2019). As an entertainment activity packed with rich authentic language input, the potential of viewing should be emphasized (Webb & Rodgers, 2009). Especially, since frequency of occurrence has been found to increase the likelihood of meaning recognition in this study, students should be encouraged to watch some related TV programmes and documentaries which contain more repeated occurrences of the same vocabulary to optimise learning from viewing (Rodgers, 2018; Rodgers & Webb, 2011). Teachers should choose authentic L2 audio-visual materials that are suitable for learners' L2 proficiency and vocabulary size and encourage students to watch outside the classroom to increase their exposure to L2 input. For learners with higher

proficiency and larger vocabulary size, they should be self-motivated to watch audio-visual materials as entertainment since they could benefit more from viewing incidentally. Moreover, the present study adds to the growing body of evidence showing that the use of different on-screen text is indeed useful for learning vocabulary (e.g., Koolstra & Beentjes, 1999; Montero Perez et al., 2014; Montero Perez et al., 2018; Peters, 2019; Syodorenko, 2010; Teng, 2018). Especially, the presence of written L2 forms increases the chance of noticing and engaging with unknown words. Thus, the presentation of written L2 should be advocated during viewing to raise learners' awareness of words and facilitate their form knowledge.

Another important implication concerns the potential of using bilingual subtitles. The present research has provided empirical evidence to support the use of bilingual subtitles and facilitate L2 learners' vocabulary learning without a trade-off to hamper learners' processing of images or video comprehension. Bilingual subtitles, presenting both L1 and L2 written forms along with visual images and auditory support, allow viewers to choose different input resources to meet their own needs. The presence of L1 translations did not prevent learners from processing L2 input, and more initial form-recognition connections could be established. The potential of bilingual subtitles in vocabulary learning has only been recognised in recent years (e.g., García, 2017; Hao et al., 2021; Lazareva & Loerts, 2017; Li, 2016; Lwo & Lin, 2012; Y. Wang, 2019). The increasing popularity of online streaming platforms in Western countries (for example, Netflix, Disney+, HBO, Amazon Prime Video, etc.) and in China (for example, Youku, iQiyi, LeTV, Tencent Video, etc.), together with video sharing websites such as YouTube and Bilibili has been witnessed in the past two decades. These platforms often enable viewers to modify the input by adding on-screen text. Therefore, bilingual subtitles can also be considered as a new subtitling type to serve as

an option to meet viewers' different preferences. For some platforms on which bilingual subtitles are available, language learners, at least those with intermediate to high proficiency levels, should be encouraged to replace L1 subtitles with bilingual subtitles to further facilitate their meaning knowledge or use captions to facilitate their form knowledge.

Apart from encouraging the use of bilingual subtitles out-of-class, the rich input offered by bilingual subtitles can also be exploited as a language teaching and learning tool. The disadvantage of bilingual subtitles over captions is that they led to less learning of word forms, due to the increased attention paid to L1 translations. This disadvantage could be overcome by giving instructions to learners and trying to direct their attention not only to L1 translations but also to L2 forms. Therefore, the potential of bilingual subtitles may be ameliorated through the integration of activities in the L2 classroom by adopting form-focused techniques, such as textual enhancement and test announcements. Although practitioners are cautious about employing L1 in the L2 classrooms, the benefit of using L1 as an effective way to quickly make correct form-meaning links has been repeatedly suggested (e.g., Laufer & Shmueli, 1997; Nation & Webb, 2011; Prince, 1996; Ramachandran & Rahim, 2004). The initial definitional encounter of an unknown word is also a good foundation for future learning of that word in context, which can be regarded as the first step in vocabulary learning (Nagy & Herman, 1987). Since vocabulary learning is a complex process where no single method can contribute to all learning gains, the benefits of bilingual subtitles can be maximised by combining with other learning activities, and learners should also be encouraged to use various learning strategies to engage with unknown words during viewing to reinforce their memory of them. The strategies and activities introduced in class could also raise learners' awareness of the language learning function of bilingual

subtitles, which can in turn benefit their incidental learning. It has been emphasized that incidental and intentional learning should complement each other to optimise vocabulary learning (Laufer, 2003; Webb, 2020a).

### **6.3. Limitations and Future Directions**

This study inevitably has a number of limitations that need to be acknowledged. They can be grouped into four broad aspects: 1) overall research design; 2) selection of material and TWs; 3) data analysis; and 4) generalisation of findings. In terms of the overall research design, a pretest-posttest design was employed to examine participants' vocabulary learning gains, whereas no delayed posttests were conducted. Thus, no conclusions can be made concerning the effects of different subtitles on word retention. Future research should also employ delayed posttests, administered several days or weeks after an immediate posttest (Nation & Webb, 2011), to see whether the advantages of bilingual subtitles for learning word meanings can be sustained. Moreover, the longitudinal effects of bilingual subtitles should also be attested in future research.

In additions, stimulated recall showed that, although very few, there were some cases in which learners reported having remembered items from the pretest. Therefore, a control group that only completes tests should be included in future research to control for potential test effects. However, it is important to note that any test effects present in the current study would equally apply to all conditions and, therefore, the results of the between-group comparisons reported in this research would still hold. Another way to reduce test effects and ensure participants' unfamiliarity with TWs is to replace real words with pseudowords, which has also been done in previous empirical studies (e.g.,

Boers, Warren, Grimshaw, & Siyanova-Chanturia, 2017; Godfroid et al., 2013; Mohamed, 2018; Montero Perez, 2020a).

An additional limitation lies in the reliability and reactivity issue of the stimulated recall method. Reliability concerns how reliable learners' self-reports are to reflect their actual cognitive processes. It is still debatable whether verbal reports can accurately reflect mental events (Ericsson & Simon, 1993; Gass & Mackey, 2017). Reactivity includes how learners' self-reports can be altered by other interventions such as a posttest that draws their attention to a particular linguistic structure (Gass & Mackey, 2017). In order to address these two issues, it is suggested to conduct simulated recall interviews as close to actual events as possible, and these should be conducted before posttests (Gass & Mackey, 2017). However, stimulated recall in the present research was conducted after posttests, since asking participants' thoughts about each TW would disclose the purpose of the study and increase participants' exposure to TWs, leading to problematic posttests scores. Although participants were asked to recall the thoughts that they had during viewing, it is still arguable that their memory recall might have been influenced by having completed vocabulary posttests. As suggested by Godfroid and Schmidtke (2013), future studies can consider using half stimulated recall and half posttests to tackle this issue.

Another limitation lies in the lack of investigation of participants' processes of auditory input. The eye-tracking method can only detect participants' visual attention. In stimulated recall, no distinctions were made between the noticing of target items aurally, visually, or both. It is therefore difficult to hypothesise how awareness, activated through different channels, can contribute to participants' learning gains. This impossibility of distinguishing between the processing of written and auditory forms of TWs is indeed a limitation of most studies on audio-visual input. The methods used in

the present study, and those used in previous studies, did not allow for an examination of the different affordances of written and auditory input in this context. Future research can apply more tightly controlled experimental stimuli and more precise measures to further tap into potential differences.

In terms of the selection of material and TWs, this study has only focused on one relative short documentary clip in one context. Replication studies using audio-visual materials with different genres or length are therefore warranted. For the selection of TWs, to ensure ecological validity, the present study used real words that appeared in the video as TWs. This meant that the number of occurrences of each word could not be modified. Previous eye-tracking research has shown that learners' attention to unknown words in reading decreased dramatically after a few encounters (e.g., Mohamed, 2018; Pellicer-Sánchez, 2016). To have a fuller understanding of the incremental learning process in a multimedia learning context, it would also be interesting for future eye-tracking research to explore how many encounters are sufficient for incidental vocabulary learning to take place, and how word repetition might affect learners' engagement with words during subtitled viewing.

Furthermore, the present research has only focused on the learning of previously unmet words, without taking into account the strengthening and enriching of words that were partially known by learners before viewing. Some participants in stimulated recall mentioned their use of bilingual subtitles to confirm their partial knowledge of a familiar word or to learn another meaning of a polysemous word. Thus, it is worth not only looking at the learning of new words but also looking into the development of partially known words (Nation & Webb, 2011; Waring, 2003). Thus, future research can also take into account learners' partial knowledge of TWs when selecting them and designing vocabulary tests to better capture the effects of viewing. In addition, apart

from only examining the learning of single words, the effects of bilingual subtitles on learning multiword items, informal and colloquial language can also be explored in future research (e.g., Frumuselu, 2018; Frumuselu et al., 2015; Pavia et al., 2019; Pellicer-Sánchez, 2017; Teng, 2019).

In terms of data analysis, while this study contributes greatly to our understanding of L2 learners' processing and learning of vocabulary in subtitled viewing by triangulating different types of data and using advanced statistical analysis, this analysis is not without its limitations. First, there was a lack of consideration of imagery support for different TWs. Previous studies have shown that words with more imagery support in video could be better learned (e.g., Peters, 2019; Rodgers, 2018). Future research should also take this variable into account when investigating the effects of viewing.

The second limitation concerns the triangulation of data. As pointed out by Creswell (2003), one difficulty in triangulating concurrent mixed methods research data lies in the comparisons made between results of two analyses using data in different forms, i.e., quantitative and qualitative. To the best of my knowledge, the present research is the first to use offline, online, and qualitative data to investigate L2 learners' incidental vocabulary learning through viewing. Mixed-effects models were used to examine the relationship between vocabulary test scores and eye-movement data, but the triangulation of qualitative data (i.e., stimulated recall) and quantitative data (i.e., vocabulary test scores and eye movements) was based on the general patterns of each group, rather than running statistical analyses focusing on each TW. This triangulation is sufficient and informative to answer the present research questions. However, future studies exploring the effects of awareness and different processing strategies on vocabulary learning could also transform qualitative data into quantitative data and run inferential statistics analysis to further explore these potential relationships.

In terms of the generalisation of findings, care should be taken when interpreting the findings and pedagogical implications of this research for the following reasons. First, the present research only focused on high-intermediate to advanced level Chinese learners of English who were familiar with bilingual subtitles and had experience of using bilingual subtitles. Consequently, the findings may not be generalisable to a different population. Viewers who have some experience of subtitles seem to use them in an effortless way (d'Ydewalle & Gielen, 1992), but this may not be the case for learners who lack experience of using subtitles. Additionally, learners with different proficiency levels and L1s may also use subtitles differently (e.g., Vanderplank, 1988; Winke et al., 2010). Thus, to attest the effects of bilingual subtitles on a larger population, more research is needed to evaluate their effectiveness with lower level L2 learners as well as learners with different L1 backgrounds. Lastly, the present research has revealed that participants spent more time processing L1 lines than L2 lines in bilingual subtitles. Processing patterns could be a consequence of the order in which L1 and L2 lines were presented. Although presenting L1 lines in the first line is the most common type of bilingual subtitles applied in China, it would be interesting in future research to conduct a study reversing the order of L1 and L2 lines to examine whether the same processing patterns would emerge.

Another limitation that deserves attention concerns the lack of control for individual differences among the participants. Working memory, in particular, contributes greatly to learners' cognitive processes and abilities (Linck, Osthus, Koeth, & Bunting, 2014). It is believed to distinguish learners' different abilities to store and manipulate information in complex cognitive activities (Baddeley, 2003), which may be closely relevant to learners' attention allocation during subtitled viewing (Gass et al., 2019). Working memory, especially phonological short-term memory, has also been

found to correlate positively with L2 vocabulary learning (Peters, 2020). Apart from working memory, although the present study did not plan to examine learners' intentionality to learn each TW, this was mentioned by a small number of participants in stimulated recall, which may also have affected their learning gains. Therefore, future incidental learning research can take into account participants' individual differences by including measures of their working memory, learning intention, and motivation, among others.

#### **6.4. Concluding Remarks**

We are living in a world surrounded by multimedia information and characterised by easy access to multimedia materials. We are also provided with different options and have the opportunity to choose how we want to engage with those. With L2 videos, viewers can freely choose which types of on-screen text they want to use to support their viewing experience. Therefore, it is the researcher's role to evaluate the pros and cons of each option, before making suggestions to guide the use and application of novel products and materials. My first goal for the present study was to investigate to what extent bilingual subtitles, as a popular subtitling type among Chinese learners of English, could improve L2 learners' incidental vocabulary learning. To do so, I compared bilingual subtitles with other common subtitling types. The findings demonstrate that bilingual subtitles seem to be superior to other subtitling types for the acquisition of word meaning, whereas they are less effective as captions for form recognition. Bilingual subtitles were also found to be as useful as L1 subtitles for comprehension.

A secondary goal of this study was to probe how bilingual subtitles were processed to further explore how learners' engagement with unknown words during subtitled

viewing related to their learning gains. With the help of eye-tracking and stimulated recall data, it has been shown that when using bilingual subtitles, learners spent similar amounts of time in the subtitling area as using captions. Moreover, learners relied more on L1 lines compared to L2 lines, and they also paid more attention to L1 translations of unknown vocabulary than to L2 word forms. Learners were also found to use a variety of processing strategies to engage with unknown words during subtitled viewing, with referring to L1 being the most frequently used strategy when using bilingual subtitles.

Then, data were triangulated to obtain a better understanding of how learners' engagement with words related to their learning gains. In general, more attention being paid to L2 unknown TWs (but not their L1 translations) tended to lead to higher vocabulary learning gains. The group that reported a higher awareness rate for unknown TWs also performed better on a form recognition test. When using bilingual subtitles, the simultaneous presentation of target L2 words and their translations increased learners' noticing of unknown words more than L1 subtitles and facilitated the establishment of initial form-meaning links, which supported learners in acquiring the meaning of novel vocabulary, at least as observed at the meaning recognition level. It has also been pointed out that more attention being paid to an unknown word (or its L1 translation) did not always lead to successful learning, potentially due to the different vocabulary processing strategies that learners apply.

This study has conducted quite a comprehensive view of the use of bilingual subtitles in L2 learners' incidental vocabulary learning. It has revealed the aspects of vocabulary knowledge that seem to benefit most from the use of bilingual subtitles, and it has signalled their potential pitfalls. As demonstrated previously, this study is not without its limitations, and more studies are warranted to thoroughly understand the benefits and accurate application of bilingual subtitles. Nevertheless, the results of the

present investigation contribute to our understanding of the benefits and potential of using multimedia to facilitate language learning.

## REFERENCES

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *psychology of learning and motivation* (Vol. 2, pp. 89-195). London, UK: Academic Press.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12-28. <http://doi.org/10.21500/20112084.807>
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon.
- Baddeley, A. D. (1999). *Essentials of human memory*. Hove, UK: Psychology Press.
- Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189-208. [https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)
- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford, UK: Oxford University Press.
- Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *The Canadian Modern Language Review*, 56(1), 31-48. <https://doi.org/10.3138/cmlr.56.1.31>
- Barcroft, J. (2009). Strategies and performance in intentional L2 vocabulary learning. *Language Awareness*, 18(1), 74-89. <https://doi.org/10.1080/09658410802557535>
- Bartolomé, A. I. H., & Cabrera, G. M. (2005). New trends in audiovisual translation: The latest challenging modes. *Miscelánea: a journal of English and American Studies*(31), 89-104. Retrieved from <https://www.miscelaneajournal.net/index.php/misc/article/view/438>
- Barton, K. (2020). MuMIn: Multi-model inference. R package (Version 1.43.17). Retrieved from <https://CRAN.R-project.org/package=MuMIn>

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.  
<https://doi.org/10.18637/jss.v067.i01>
- Bianchi, F., & Ciabattini, T. (2008). Captions and subtitles in EFL learning: An investigative study in a comprehensive computer environment. In A. Baldry, M. Pavesi, & C. T. Torsello (Eds.), *From didactas to ecolingua* (pp. 69–80). Trieste, Italy: Edizioni Università di Trieste.
- Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS One*, 11(6), e0158409.  
<https://doi.org/10.1371/journal.pone.0158409>
- Bisson, M. J., van Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 35(02), 399-418. <https://doi.org/10.1017/s0142716412000434>
- Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 143-157). London, UK: Routledge.
- Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning*, 30(7), 709-725.  
<http://doi.org/10.1080/09588221.2017.1356335>
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191. <http://doi.org/10.1017/S0267190500003536>
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Berg, A. M. C. W., Nielsen, A., Skaug, H. J., . . . Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378-400. <https://doi.org/10.3929/ethz-b-000240890>
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a foreign language*, 20(2), 136-163. <http://doi.org/10.125/66816>
- Bruton, A., Lopez, M. G., & Mesa, R. E. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly*, 45(4), 759-768.  
<http://doi.org/10.5054/tq.2011.268061>

- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford, UK: Oxford University Press.
- Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.
- Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 660-675. <http://doi.org/10.1037/0096-1523.34.3.660>
- Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics*, 14(2), 25-44.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293-332.  
[http://doi.org/10.1207/s1532690xci0804\\_2](http://doi.org/10.1207/s1532690xci0804_2)
- Chang, A. C. S. (2019). Effects of narrow reading and listening on L2 vocabulary learning: Multiple dimensions. *Studies in Second Language Acquisition*, 41(4), 769-794. <http://doi.org/10.1017/S0272263119000032>
- Chen, Y. (2021). Comparing incidental vocabulary learning from reading-only and reading-while-listening. *System*, 97, 102442.  
<https://doi.org/10.1016/j.system.2020.102442>
- Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement*. New York: Springer.
- Chung, J. M. (1999). The effects of using video texts supported with advance organizers and captions on Chinese college students' listening comprehension: An empirical study. *Foreign Language Annals*, 32(3), 295-308.  
<http://doi.org/10.1111/j.1944-9720.1999.tb01342.x>
- CNNIC. (2021). The 47th China Statistical Report on Internet Development. Retrieved from  
<http://www.cnnic.cn/hlwfzyj/hlwzxbg/hlwtjbg/202102/P020210203334633480104.pdf>
- Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge, UK: Cambridge University Press.
- Cohen, A. D. (1990). *Language learning: Insights for learners, teachers, and researchers*. New York: Newbury House.

- Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, 32(3), 453-467.  
<http://doi.org/10.1177/0267658316637401>
- Conklin, K., Pellicer-Sánchez, A., & Carroll, G. (2018). *Eye-tracking: A guide for applied linguistics research* Cambridge, UK: Cambridge University Press.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching*, 5(1), 161-170.  
<http://doi.org/10.1515/iral.1967.5.1-4.161>
- Corel. (2018). Corel VideoStudio Pro 2018 [Computer software]. Retrieved from <https://www.videostudiopro.com/en/pages/videostudio-2018/>
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684.  
[https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X).
- Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268-294. <http://doi.org/10.1037/0096-3445.104.3.268>
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). London, UK: SAGE Publications.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). London, UK: SAGE Publications.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369-382.  
<http://doi.org/10.1177/0267658312443651>
- D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, 23(1), 109-122.  
<http://doi.org/10.1080/10862969109547729>
- d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist*, 12(3), 196-205.  
<http://doi.org/10.1027/1016-9040.12.3.196>
- d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 415-427). New York: Springer.

- d'Ydewalle, G., & Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition? In P. Winterhoff-Spurk & T. H. A. van der Voort (Eds.), *New horizons in media psychology: Research cooperation and projects in Europe* (pp. 145-155). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.
- d'Ydewalle, G., Praet, C., Verfaillie, K., & Rensbergen, J. V. (1991). Watching subtitled television: Automatic reading behavior. *Communication research*, 18(5), 650-666. <https://doi.org/10.1177%2F009365091018005005>
- d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28(3), 227-244. <https://doi.org/10.1023/A:1023202130625>
- Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, 42(4), 497-527. <http://doi.org/10.1111/j.1467-1770.1992.tb01042.x>
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta*, 49(1), 67-77. <http://doi.org/10.7202/009021ar>
- de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, 68(4), 906-941. <http://doi.org/10.1111/lang.12296>
- Desmet, P., & Cornillie, F. (2012). Multimedia CALL. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2368-2371). Boston: Springer.
- Dizon, G., & Thanyawatpokin, B. (2021). Language learning with Netflix: Exploring the effects of dual subtitles on vocabulary learning and listening comprehension. *Computer Assisted Language Learning Electronic Journal*, 22(3), 52-65. Retrieved from <http://callej.org/journal/22-3/Dizon-Thanyawatpokin2021.pdf>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford, UK: Oxford University Press.
- Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10), 28-28. <http://doi.org/10.1167/10.10.28>
- Duquette, L., & Painchaud, G. (1996). A comparison of vocabulary acquisition in audio and video contexts. *Canadian modern language review*, 53(1), 143-172. <https://doi.org/10.3138/cmlr.53.1.143>

- Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, 88, 1-13.  
<https://doi.org/10.1016/j.system.2019.102166>
- Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 40(2), 341-366.  
<http://doi.org/10.1017/S0272263117000109>
- Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24(2), 174-187. <https://doi.org/10.2307/747863>
- Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In S. Fotos (Ed.), *Learning a second language through interaction* (pp. 35-61). Amsterdam, Netherlands: John Benjamins.
- Ericsson, K. A. (2002). Towards a procedure for eliciting verbal expression of non-verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology*, 16(8), 981-987. <https://doi.org/10.1002/acp.925>
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, UK: Mass.
- Fazilatfar, A. M., Ghorbani, S., & Samavarchi, L. (2011). The effect of standard and reversed subtitling versus no subtitling mode on L2 vocabulary learning. *Journal of Teaching Language Skills*, 30(1), 43-64.  
<https://dx.doi.org/10.22099/jtls.2012.393>
- Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 42(3), 499-523. <http://doi.org/10.1017/S0272263119000494>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, UK: SAGE Publications.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). California: SAGE Publications.
- Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21(2), 225-241.  
<http://doi.org/10.1017/S0272263199002041>

- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109. <https://doi.org/10.3102/00346543074001059>
- Frumuselu, A. D. (2018). The implications of Cognitive Load Theory and exposure to subtitles in English Foreign Language (EFL). *Translation and Translanguaging in Multilingual Contexts*, 4(1), 55-76. <http://doi.org/10.1075/ttmc.00004.fru>
- Frumuselu, A. D., De Maeyer, S., Donche, V., & Colon Plana, M. D. M. G. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, 32, 107-117. <http://doi.org/10.1016/j.linged.2015.10.001>
- Gampe, A., Liebal, K., & Tomasello, M. (2012). Eighteen-month-olds learn novel words through overhearing. *First Language*, 32(3), 385-397. <http://doi.org/10.1177/0142723711433584>
- García, B. (2017). Bilingual subtitles for second-language acquisition and application to engineering education as learning pills. *Computer Applications in Engineering Education*, 25(3), 468-479. <https://doi.org/10.1002/cae.21814>
- Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24(3), 239-258. <http://doi.org/10.1111/j.1944-9720.1991.tb00469.x>
- Gass, S. M. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21(2), 319-333. <https://doi.org/10.1017/S02722263199002090>
- Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research* (2nd ed.). New York: Routledge.
- Gass, S. M., Svetics, I., & Lemelin, S. (2003). Differential effects of attention. *Language Learning*, 53(3), 497-546. <https://doi.org/10.1111/1467-9922.00233>
- Gass, S. M., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology*, 23(2), 84-104. <http://hdl.handle.net/10125/44684>
- Gesa Vidal, F. (2019). *L1/L2 subtitled TV series and EFL learning: A study on vocabulary acquisition and content comprehension at different proficiency levels*. (Unpublished doctoral dissertation). University of Barcelona, Barcelona, Spain.

- Gilabert, R., Vasylets, L., Moskvina, N., Barón, J., Levkina, M., Suárez, M. d. M., & Feijóo, S. (2018). *The impact of genre on L2 vocabulary learning through captioned video*. Paper presented at the EuroSLA 28, University of Münster.
- Godfroid, A. (2020a). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. New York: Routledge.
- Godfroid, A. (2020b). Sensitive measures of vocabulary knowledge and processing. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 433-453). London, UK: Routledge.
- Godfroid, A., Ahn, J., Choi, I. N. A., Ballard, L., Cui, Y., Johnston, S., . . . Yoon, H.-J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism*, 21(3), 563-584.  
<http://doi.org/10.1017/S1366728917000219>
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35(03), 483-517.  
<http://doi.org/10.1017/s0272263113000119>
- Godfroid, A., & Hui, B. (2020). Five common pitfalls in eye-tracking research. *Second Language Research*, 36(3), 277-305. <http://doi.org/10.1177/0267658320921218>
- Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183-205). Honolulu: University of Hawaii, National Foreign Language Resource Center.
- Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (Vol. 48, pp. 325-348). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Gottlieb, H. (2004). Subtitles and international anglicization. *Nordic Journal of English Studies*, 3(1), 219-230. Retrieved from  
<https://ojs.ub.gu.se/index.php/njes/article/download/244/241>
- Goulden, R., Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341-363. <https://doi.org/10.1093/applin/11.4.341>

- Grace, C. A. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *The Modern Language Journal*, 82(4), 533-544.  
<http://doi.org/10.2307/330223>
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274. <http://doi.org/10.2307/1163620>
- Gu, Y. (2020). Strategies for learning vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 271-287). London, UK: Routledge.
- Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46(4), 643-679.  
<https://doi.org/10.1111/j.1467-1770.1996.tb01355.x>
- Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological methods*, 3(3), 339-353.  
<https://doi.org/10.1037/1082-989X.3.3.339>
- Hao, T., Sheng, H., Ardasheva, Y., & Wang, Z. (2021). Effects of dual subtitles on Chinese students' English listening comprehension and vocabulary learning. *The Asia-Pacific Education Researcher*.  
<https://doi.org/10.1007/s40299-021-00601-w>
- Hashemi, M. R., & Babaii, E. (2013). Mixed methods research: Toward new research designs in applied linguistics. *The Modern Language Journal*, 97(4), 828-852.  
<https://doi.org/10.1111/j.1540-4781.2013.12049.x>
- Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2021). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, 0(0), 1-30.  
<https://doi.org/10.1177/13621688211001289>
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a foreign language*, 11, 207-223.
- Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *The Canadian Modern Language Review*, 56(2), 308-328. <http://doi.org/10.3138/cmlr.56.2.308>

- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 346-363.  
<https://doi.org/10.1002/bimj.200810425>
- Hu, H.-c. M., & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies, and their relationship with the retention of word meanings inferred from context. *Canadian modern language review*, 68(1), 54-77.  
<http://doi.org/10.3138/cmlr.68.1.054>
- Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153-178). Norwood, New Jersey: Ablex Publishing Corporation.
- Hugo Santiago, S., & Trevor, G. (2020). Stimulated recall. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics*. Abingdon, UK: Routledge.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London, UK: Macmillan Academic and Professional Ltd.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge, UK: Cambridge University Press.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*. Oxford, UK: Blackwell.
- In'Nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26(2), 219-244.  
<http://doi.org/10.1177/0265532208101006>
- Jay, M. (2019). generalhoslem: Goodness of fit tests for logistic regression models. R package version 1.3.4. Retrieved from  
<https://CRAN.R-project.org/package=generalhoslem>

- Jelani, N. A. M., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video: Does test modality matter? *ITL - International Journal of Applied Linguistics*, 169(1), 169-190. <https://doi.org/10.1075/itl.00011.jel>
- Jiang, N. (2002). Form–meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 24(4), 617-637. <https://doi.org/10.1017/S0272263102004047>
- Johnson, B., & Christensen, L. B. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). London, UK: SAGE Publications.
- Joyce, P. (2018). L2 vocabulary learning and testing: The use of L1 translation versus L2 definition. *The Language Learning Journal*, 46(3), 217-227. <http://doi.org/10.1080/09571736.2015.1028088>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354. <http://doi.org/10.1037/0033-295X.87.4.329>
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1-2), 262-284. <http://doi.org/10.1080/09541440340000213>
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, 78(3), 285-299. <http://doi.org/10.2307/330108>
- Koolstra, C. M., & Beentjes, J. W. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development*, 47(1), 51-60.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Essex, UK: Longman Group Limited.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440-464.

- Kruger, J. L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, 7(5). <http://doi.org/10.16910/jemr.7.5.4>
- Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, 35(1), 65-85.  
<http://doi.org/10.1080/17439880903561876>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1-26.  
<https://doi.org/10.18637/jss.v082.i13>
- Lambert, W. E., & Holobow, N. (1984). Combinations of printed script and spoken dialogue that show promise for students of a foreign language. *Canadian Journal of Behavioural Science*, 16(1), 1-11.  
<https://psycnet.apa.org/doi/10.1037/h0080775>
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian modern language review*, 59(4), 567-587. <http://doi.org/10.3138/cmlr.59.4.567>
- Laufer, B., & Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *The Modern Language Journal*, 101(4), 729-741. <http://doi.org/10.1111/modl.12431>
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202-226.  
<http://doi.org/10.1191/0265532204lt277oa>
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.  
<http://doi.org/10.1111/j.0023-8333.2004.00260.x>
- Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26. <http://doi.org/10.1093/applin/22.1.1>
- Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC journal*, 28(1), 89-108.  
<http://doi.org/10.1177/003368829702800106>

- Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning*, 46(1), 101-135.  
<https://doi.org/10.1111/j.1467-1770.1996.tb00642.x>
- Lazareva, E., & Loerts, H. (2017). Double subtitles as an effective tool for vocabulary learning. *Koninskie Studia Jezykowe-Konin Language Studies*, 5, 61-92.
- Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study *The Modern Language Journal*, 102(3), 557-577. <https://doi.org/10.1111/modl.12503>
- Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, 42(3), 625-651.  
<http://doi.org/10.1017/S0272263120000108>
- Lenhart, J., Lenhard, W., Vaahtoranta, E., & Suggate, S. (2018). Incidental vocabulary acquisition from listening to stories: A comparison between read-aloud and free storytelling approaches. *Educational Psychology*, 38(5), 596-616.  
<https://doi.org/10.1080/01443410.2017.1363377>
- Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, 47(3), 467-505. <http://doi.org/10.1111/0023-8333.00017>
- Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach* (1st ed.). New York: Routledge.
- Leow, R. P., Grey, S., Marijuan, S., & Moorman, C. (2014). Concurrent data elicitation procedures, processes, and the early stages of L2 learning: A critical overview. *Second Language Research*, 30(2), 111-127.  
<http://doi.org/10.1177/0267658313511979>
- Li, M. (2012). *L1, L2 and L1+ L2 subtitles in videos in L2 classroom settings: An investigation of teachers' and students' attitudes towards the use of L1, L2 and L1+ L2 subtitles in the L2 classroom*. Saarbrücken, Germany: LAP LAMBERT Academic Publishing.
- Li, M. (2016). *Investigation into the differential effects of subtitles (first language, second language, and bilingual) on second language vocabulary acquisition*. (Unpublished doctoral dissertation). The University of Edinburgh, Edinburgh.
- Liao, S., Kruger, J.-L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of*

- Specialised Translation*, 33. Retrieved from [https://www.jostrans.org/issue33/art\\_liao.php](https://www.jostrans.org/issue33/art_liao.php)
- Linck, J., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65(S1), 185-207. <https://doi.org/10.1111/lang.12117>
- Linck, J., Osthus, P., Koeth, J., & Bunting, M. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861-883. <http://doi.org/10.3758/s13423-013-0565-2>
- Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147-158). Cambridge, UK: Cambridge University Press.
- Lüdecke, D. (2020). sjPlot: Data visualization for statistics in social science. R package (Version 2.8.4). Retrieved from <https://CRAN.R-project.org/package=sjPlot>
- Lunin, M., & Minaeva, L. (2015). Translated subtitles language learning method: A new practical approach to teaching English. *Procedia - Social and Behavioral Sciences*, 199, 268-275. <https://doi.org/10.1016/j.sbspro.2015.07.516>
- Lwo, L., & Lin, M. C.-T. (2012). The effects of captions in teenagers' multimedia L2 learning. *ReCALL*, 24(02), 188-208. <http://doi.org/10.1017/s0958344012000067>
- MacFadden, K., Barrett, K., & Horst, M. (2009). What's in a television word list? A corpus-informed investigation. *Concordia Working Papers in Applied Linguistics*, 2, 78-98.
- Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 40(03), 651-675. <http://doi.org/10.1017/s0272263117000341>
- Markham, P. L., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems*, 31(3), 331-341. <https://doi.org/10.2190%2FBHUH-420B-FE23-ALA0>
- Markham, P. L., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, 34(5), 439-445. <http://doi.org/10.1111/j.1944-9720.2001.tb02083.x>

- Mayer, R. E. (2005a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge, UK: Cambridge University Press.
- Mayer, R. E. (2005b). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-16). Cambridge, UK: Cambridge University Press.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, 28(5), 653-660.  
<https://doi.org/10.1002/acp.3050>
- Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In M. McCarthy & N. Schmitt (Eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge, UK: Cambridge University Press.
- Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge, UK: Cambridge University Press.
- Mohamed, A. A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, 40(2), 269. <http://doi.org/10.1017/S0272263117000092>
- Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: An eye-tracking study. *The Language Learning Journal*, 47(4), 460-478.  
<http://doi.org/10.1080/09571736.2019.1638623>
- Montero Perez, M. (2020a). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 42(4), 749-773.  
<http://doi.org/10.1017/S0272263119000706>
- Montero Perez, M. (2020b). Multimodal input in SLA research. *Studies in Second Language Acquisition*, 42(3), 653-663.  
<http://doi.org/10.1017/S0272263120000145>
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, 18(1), 118-141.  
<http://dx.doi.org/10125/44357>

- Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, 99(2), 308-328. <https://doi.org/10.1111/modl.12215>
- Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(1-2), 1-26. <http://doi.org/10.1080/09588221.2017.1375960>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720-739. <http://doi.org/10.1016/j.system.2013.07.013>
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge, UK: Cambridge University Press.
- Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, 67, 77-86. <https://doi.org/10.1016/j.system.2017.04.015>.
- Naghizadeh, M., & Darabi, T. (2015). The impact of bimodal, Persian and no-subtitle movies on Iranian EFL learners' L2 vocabulary learning. *Journal of Applied Linguistics and Language Research*, 2(2), 66-79. Retrieved from [http://www.jallr.com/index.php/JALLR/article/view/33/pdf\\_31](http://www.jallr.com/index.php/JALLR/article/view/33/pdf_31)
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270. <http://doi.org/10.3102/00028312024002237>
- Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The Nature of vocabulary acquisition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253. <https://doi.org/10.2307/747758>
- Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL*, 20(1), 3-20. <http://doi.org/10.1017/S0958344008000219>

- Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4), 645-670. <https://doi.org/10.2307/3588216>
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*: Boston, MA: Heinle & Heinle Publishers.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, I. S. P. (2003). The role of first language in foreign language learning. *The Asian EFL journal*, 5(2), 1-8. Retrieved from <https://www.asian-efl-journal.com/main-editions-new/the-role-of-the-first-language-in-foreign-language-learning/index.htm>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian modern language review*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2012). Vocabulary Size Test information and specifications. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Nation, I. S. P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 15-29). London, UK: Routledge.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13. Retrieved from [https://jalt-publications.org/tlt/issues/2007-07\\_31.7](https://jalt-publications.org/tlt/issues/2007-07_31.7)
- Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts [Computer software] (Version 3). Retrieved from <https://www.lex tutor.ca/cgi-bin/range/texts/index.pl> .
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Navarro, D. J. (2015). Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.5). Retrieved from <http://ua.edu.au/ccs/teaching/lsr>

- Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27(1), 95-106. <https://doi.org/10.2307/747835>
- Newmark, P. (1981). *Approaches to translation* (1st ed.). Oxford, UK: Pergamon Press.
- Niegemann, H. M., & Heidig, S. (2012). Multimedia learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2372-2375). Boston, MA: Springer US.
- Ouyang, J., Huang, L., & Jiang, J. (2020). The effects of glossing on incidental vocabulary learning during second language reading: Based on an eye-tracking study. *Journal of Research in Reading*, 43(4), 496-515. <https://doi.org/10.1111/1467-9817.12326>
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.
- Paivio, A. (2007). Basic principles of dual coding theory. In A. Paivio (Ed.), *Mind and its evolution: A dual coding theoretical approach* (pp. 25-57). New York, NY: Psychology Press.
- Paivio, A. (2014). Bilingual dual coding theory and memory. In R. R. Heredia & J. Altarriba (Eds.), *Foundations of bilingual memory* (pp. 41-62). New York, NY: Springer.
- Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology*, 34(4), 388-399. <https://psycnet.apa.org/doi/10.1037/h0081101>
- Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (6th ed.). Berkshire, UK: Open University Press.
- Pattemore, A., & Muñoz, C. (2020). Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors. *System*, 93, 102303. <https://doi.org/10.1016/j.system.2020.102303>
- Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, 41(4), 745-768. <http://doi.org/10.1017/S0272263119000020>
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97-130. <http://doi.org/10.1017/S0272263115000224>

- Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381-402.  
<http://doi.org/10.1177/1362168815618428>
- Pellicer-Sánchez, A. (2020a). Expanding English vocabulary knowledge through reading: Insights from eye-tracking studies. *RELC journal*, 51(1), 134-146.  
<http://doi.org/10.1177/0033688220906904>
- Pellicer-Sánchez, A. (2020b). Learning single words vs. multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 158-173). London, UK: Routledge.
- Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, 71(1), 162-203.  
<https://doi.org/10.1111/lang.12430>
- Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a foreign language*, 22(1), 31-55.
- Pellicer-Sánchez, A., & Siyanova-Chanturia, A. (2018). Eye movements in vocabulary research: Current findings and future directions. *ITL - International Journal of Applied Linguistics*, 169(1), 5-29. <https://doi.org/10.1075/itl.00004.pel>
- Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M. P. H., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, 1-22. <http://doi.org/10.1017/S0272263120000091>
- Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, 13(3), 243-272.  
<http://doi.org/10.1080/15213269.2010.502873>
- Peters, E. (2007). Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning & Technology*, 11(2), 36-58.  
<http://dx.doi.org/10125/44103>
- Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, 18(1), 75-94. <http://doi.org/10.1177/1362168813505384>

- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, 169(1), 142-168. <https://doi.org/10.1075/itl.00010.pet>
- Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, 53(4), 1008-1032. <http://doi.org/10.1002/tesq.531>
- Peters, E. (2020). Factors affecting the learning of single-word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 125-142). London, UK: Routledge.
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134-148. <http://doi.org/10.1016/j.system.2016.10.002>
- Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, 59(1), 113-151. <https://doi.org/10.1111/j.1467-9922.2009.00502.x>
- Peters, E., & Muñoz, C. (2020). Introduction to the special issue: Language learning from multimodal input. *Studies in Second Language Acquisition*, 42(3), 489-497. <http://doi.org/10.1017/S0272263120000212>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 1-27. <http://doi.org/10.1017/s0272263117000407>
- Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50-72. <http://doi.org/10.1017/S0267190515000094>
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a foreign language*, 18(1), 1-28.
- Pitts, M., White, H., & Krashen, S. (1989). Acquiring the second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a foreign language*, 5(2), 271-275.
- Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467-488). Cambridge, UK: Cambridge University Press.

- Plonsky, L., & Oswald, F. L. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912.  
<https://doi.org/10.1111/lang.12079>
- PortableSoft. (2012). SrtEdit [Computer software] (Version 6.3). Retrieved from  
<http://www.portablesoft.org/srtedit-portable/>
- Price, K. (1983). Closed-captioned TV: An untapped resource. *MATSOL Newsletter*, 12(2), 1-8. Retrieved from  
<https://www.matsol.org/assets/documents/Currentsv12no2Fall1983.pdf>
- Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, 80(4), 478-493. <https://doi.org/10.1111/j.1540-4781.1996.tb05468.x>
- Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, 47(4), 424-438.  
<http://doi.org/10.1080/09571736.2019.1638630>
- Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 42(3), 525-549. <http://doi.org/10.1017/S027226311900055X>
- Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47(4), 479-496.  
<http://doi.org/10.1080/09571736.2019.1616806>
- Pujadas, G., & Muñoz, C. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, 42(3), 551-575. <http://doi.org/10.1017/S0272263120000042>
- Qian, D. D., & Lin, L. H. F. (2020). The relationship between vocabulary knowledge and language proficiency. In *The Routledge handbook of vocabulary studies* (pp. 66-80). London, UK: Routledge.
- R Development Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.

- Ramachandran, S. D., & Rahim, H. A. (2004). Meaning recall and retention: The impact of the translation method on elementary level learners' vocabulary learning. *RELC journal*, 35(2), 161-178. <http://doi.org/10.1177/003368820403500205>
- Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85(3), 618-660. <http://doi.org/10.1037/0033-2909.85.3.618>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506. <http://doi.org/10.1080/17470210902816461>
- Rayner, K., Li, X., Juhasz, B., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review*, 12(6), 1089-1093. <http://doi.org/10.3758/BF03206448>
- Rayner, K., Li, X., & Pollatsek, A. (2007). Extending the E-Z reader model of eye movement control to Chinese readers. *Cognitive Science*, 31(6), 1021-1033. <https://doi.org/10.1080/03640210701703824>
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Reichle, E. D. (2006). Computational models of eye-movement control during reading: Theories of the "eye-mind" link. *Cognitive Systems Research*, 7(1), 2-3. <http://doi.org/10.1016/j.cogsys.2005.07.001>
- Rice, M., Huston, A., Truglio, R., & Wright, J. (1990). Words from "Sesame Street": Learning vocabulary while viewing. *Developmental Psychology*, 26(3), 421-428. <https://psycnet.apa.org/doi/10.1037/0012-1649.26.3.421>
- Rice, M., & Woodsmall, L. (1988). Lessons from television: Children's word learning when viewing. *Child Development*, 59(2), 420-429. <http://doi.org/10.2307/1130321>
- Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Malden, MA: Blackwell Publishing Ltd.

- Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2011). Attention and awareness in second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247-267). Abingdon, UK: Routledge.
- Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.
- Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics*, 169(1), 191-211. <http://doi.org/10.1075/itl.00012.rod>
- Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45(4), 689-717. <http://doi.org/10.5054/tq.2011.268062>
- Rodgers, M. P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO Journal*, 34(1), 20-38. <http://doi.org/10.1558/cj.29522>
- Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through viewing television. *ITL - International Journal of Applied Linguistics*, 171(2), 191-220. <http://doi.org/10.1075/itl.18034.rod>
- Rose, H., McKinley, J., & Briggs Baffoe-Djan, J. (2020). *Data collection research methods in applied linguistics*. London, UK: Bloomsbury.
- Rott, S. (2000). Relationships between the process of reading, word inferencing, and incidental word acquisition, in assigning meaning to form. In J. E. Lee & A. Valdman (Eds.), *Form and meaning: Multiple perspectives* (pp. 255-282). Boston, MA: Heinle & Heinle Publishers.
- Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a foreign language*, 17(2), 95-124.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158. <http://doi.org/10.1093/applin/11.2.129>

- Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, 13, 206-226. <http://doi.org/10.1017/S0267190500002476>
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11-26.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (Vol. 9, pp. 1-63). Honolulu, Hawaii: University of Hawaii Press.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363. <http://doi.org/10.1177/1362168808089921>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London, UK: Palgrave Macmillan.
- Schmitt, N., & McCarthy, M. (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge, UK: Cambridge University Press.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88. <http://doi.org/10.1177/026553220101800103>
- Segler, T. M., Pain, H., & Sorace, A. (2002). Second language vocabulary acquisition and learning strategies in ICALL environments. *Computer Assisted Language Learning*, 15(4), 409-422. <http://doi.org/10.1076/call.15.4.409.8272>
- Selvi, A. F. (2020). Qualitative content analysis. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 440-452). Abingdon, UK: Routledge.
- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69(3), 493-525. <http://doi.org/10.1177/0013164408323233>

- SR Research. (2011). Experiment Builder [Apparatus and software] (Version 1.10.165). Mississauga, Ontario, Canada: SR Research Ltd. Retrieved from <https://www.sr-research.com/experiment-builder/>
- SR Research. (2016). Eyelink 1000 plus [Apparatus and software]. Mississauga, Ontario, Canada: SR Research Ltd. Retrieved from <https://www.sr-research.com/eyelink1000plus.html>
- SR Research. (2018). EyeLink Data Viewer [Apparatus and software] (Version 3.1.246). Mississauga, Ontario, Canada: SR Research Ltd. Retrieved from <https://www.sr-research.com/data-viewer/>
- Suggate, S. P., Lenhard, W., Neudecker, E., & Schneider, W. (2013). Incidental vocabulary acquisition from stories: Second and fourth graders learn more from listening than reading. *First Language*, 33(6), 551-571.  
<http://doi.org/10.1177/0142723713503144>
- Sun, F., & Feng, D. (1999). Eye movements in reading Chinese and English text. In J. Wang, H.-C. Chen, R. Radach, & A. Inhoff (Eds.), *Reading Chinese script: A cognitive analysis*. Mahwah, NJ: Psychology Press.
- Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, 28(3), 387-422.  
<http://doi.org/10.1017/S0272263106060177>
- Svalberg, A. M. L. (2009). Engagement with language: Interrogating a construct. *Language Awareness*, 18(3-4), 242-258.  
<http://doi.org/10.1080/09658410903197264>
- Svalberg, A. M. L. (2018). Researching language engagement; current trends and future directions. *Language Awareness*, 27(1-2), 21-39.  
<http://doi.org/10.1080/09658416.2017.1406490>
- Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261-285.  
<http://doi.org/10.3102/00346543069003261>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.  
[http://doi.org/10.1016/0364-0213\(88\)90023-7](http://doi.org/10.1016/0364-0213(88)90023-7)

- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295-312.  
[https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Sweller, J. (2005a). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). Cambridge, UK: Cambridge University Press.
- Sweller, J. (2005b). The redundancy principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159-168). Cambridge, UK: Cambridge University Press.
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50-73. <http://doi.org/10125/44214>
- Tanenhaus, M. K. (2007). Spoken language comprehension: insights from eye movements. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 209-329). Oxford, UK: Oxford University Press.
- Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. Traxler, M. Gernsbacher, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 863-900). Amsterdam, Netherlands: Elsevier Science & Technology.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. London, UK: SAGE Publications.
- Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, 12(3), 274-288.  
<http://doi.org/10.1080/17501229.2016.1203328>
- Teng, F. (2019). The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education*, 142, 1-14.  
<http://doi.org/10.1016/j.compedu.2019.103655>
- Teng, F. (2020a). Retention of new words learned incidentally from reading: Word exposure frequency, L1 marginal glosses, and their combination. *Language Teaching Research*, 24(6), 785-812. <http://doi.org/10.1177/1362168819829026>

- Teng, F. (2020b). Vocabulary learning through videos: Captions, advance-organizer strategy, and their combination. *Computer Assisted Language Learning*, 1-33. <http://doi.org/10.1080/09588221.2020.1720253>
- Thompson, I. (1995). Testing listening comprehension. *AATSEEL Newsletter*, 37(5), 24-31. Retrieved from <https://eric.ed.gov/?id=ED386916>
- Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16(2), 183-203. <http://doi.org/10.1017/S0272263100012870>
- Tremblay, A., & Ransijn, J. (Producer). (2020). LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models. *R package 3.0*. Retrieved from <https://cran.r-project.org/web/packages/LMERConvenienceFunctions/LMERConvenienceFunctions.pdf>
- Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Clarendon Press.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53(1), 1-25. <http://doi.org/10.1146/annurev.psych.53.100901.135114>
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559-599. <http://doi.org/10.1111/lang.12343>
- Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41(3), 609-624. <http://doi.org/10.1016/j.system.2013.07.012>
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT journal*, 42(4), 272-281. <https://doi.org/10.1093/elt/42.4.272>
- Vanderplank, R. (2010). "Deja Vu"? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, 43(1), 1-37. <http://doi.org/10.1017/S0261444809990267>
- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56-373. <http://doi.org/10.1093/applin/24.1.56>
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219-258. <http://doi.org/10.1111/j.1467-9922.2010.00593.x>

- Vu, D. V., & Peters, E. (2020). Learning vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC journal*, 1-16.  
<https://doi.org/10.1177%2F0033688220911485>
- Wang, D. (2012). Self-directed English language learning through watching English television drama in China. *Changing English*, 19(3), 339-348.  
<http://doi.org/10.1080/1358684x.2012.704584>
- Wang, D., & Zhang, X. (2017). Fansubbing in China: Technology-facilitated activism in translation. *Target*, 29(2), 301-318. <http://doi.org/10.1075/target.29.2.06wan>
- Wang, Y. (2019). Effects of L1/L2 captioned TV programs on students' vocabulary learning and comprehension. *CALICO Journal*, 36(3), 204-224.  
<http://dx.doi.org/10.1558/cj.36268>
- Waring, R. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a foreign language*, 15(2), 130-163.  
<http://doi.org/10.125/66776>
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33-52. <http://doi.org/10.1017/s0272263105050023>
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65. <http://doi.org/10.1093/applin/aml048>
- Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies*, 11(1), 117-135. <http://doi.org/10.6018/ijes/2011/1/137131>
- Webb, S. (2012). Depth of vocabulary knowledge. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley.
- Webb, S. (2020a). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225-239). London, UK: Routledge.
- Webb, S. (2020b). Introduction. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 1-12). London, UK: Routledge.
- Webb, S., & Chang, A. C.-S. (2012). Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review*, 68(3), 267-290. <http://doi.org/10.3138/cmlr.1204.1>

- Webb, S., & Chang, A. C. S. (2015a). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37(4), 651-675.  
<http://doi.org/10.1017/S0272263114000606>
- Webb, S., & Chang, A. C. S. (2015b). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, 19(6), 667-686.  
<http://doi.org/10.1177/1362168814559800>
- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford, UK: Oxford University Press.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91-120. <http://doi.org/10.1111/j.1467-9922.2012.00729.x>
- Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, 59(2), 335-366.  
<http://doi.org/10.1111/j.1467-9922.2009.00509.x>
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian modern language review*, 53(1), 13-40. <https://doi.org/10.3138/cmlr.53.1.13>
- Wilkins, D. A. (1972). *Linguistics in language teaching*. London, UK: Edward Arnold.
- Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, 16(1-2), 312-339.  
<http://doi.org/10.1080/09541440340000196>
- Winke, P., Gass, S. M., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65-86. <http://dx.doi.org/10125/44203>
- Winke, P., Gass, S. M., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1), 254-275. <http://doi.org/10.1111/j.1540-4781.2013.01432.x>
- Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, 35(2), 182-191.  
<https://doi.org/10.1016/j.system.2006.12.009>

- Yanagisawa, A., & Webb, S. (2020). Measuring depth of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 371-386). London, UK: Routledge.
- Yang, J. C., & Chang, P. (2013). Captions and reduced forms instruction: The impact on EFL students' listening comprehension. *ReCALL*, 26(01), 44-61.  
<http://doi.org/10.1017/s0958344013000219>
- Zhang, X. (2013). Fansubbing in China. *MultiLingual*, July/August 2013. Retrieved from <https://multilingual.com/articles/fansubbing-in-china/>
- Zhou, S. A., Hiver, P., & Al-Hoorie, A. H. (2020). Measuring L2 engagement: A review of issues and applications. In P. Hiver, A. H. Al-Hoorie, & S. Mercer (Eds.), *student engagement in the language classroom* (pp. 75-98). Bristol, UK: Multilingual Matters.
- Zou, D. (2016). Comparing dictionary-induced vocabulary learning and inferencing in the context of reading. *Lexikos*, 26, 372-390. <https://doi.org/10.5788/26-1-1345>

## APPENDICES

### List of Appendices

- Appendix S1. Initial Online Viewing Habits Questionnaire
- Appendix S2. Approved Ethics Form
- Appendix S3. Information Sheet
- Appendix S4. Consent Form
- Appendix S5. Meaning Recognition Test
- Appendix S6. Comprehension Test
- Appendix S7. 3K Vocabulary Levels Test
- Appendix S8. Background Questionnaire for Participants in the Main Study
- Appendix S9. Stimulated Recall Interview Oral Instructions
- Appendix S10. Meaning Recall Test Scoring Scheme
- Appendix S11. Mixed-Effects Models for Four Eye-Movement Measures for Level 1 Overall Subtitling area (Level 1)
- Appendix S12. Mixed-Effects Models for Four Eye-Movement Measures for Subtitling area Within the Bilingual Subtitles Group (Level 2)
- Appendix S13. Mixed-Effects Models for Eye-Movement Measures for Four Subtitling area Between the Bilingual Subtitles and the Captions Groups (Level 2)
- Appendix S14. Mixed-Effects Models for Four Eye-Movement Measures for Subtitling area Between the Bilingual Subtitles and the L1 Subtitles Groups (Level 2)
- Appendix S15. Mixed-Effects Models for Seven Eye-Movement Measures Within the Bilingual Subtitles Group (Level 3)
- Appendix S16. Mixed-Effects Models for Seven Eye-Movement Measures Between the Bilingual Subtitles and the Captions Groups (RQ3)
- Appendix S17. Mixed-Effects Models for Seven Eye-Movement Measures Between the Bilingual Subtitles and the L1 Subtitles Groups (RQ3)
- Appendix S18. Mixed-Effects Models with Significant Results for the Relationship Between Eye-Movement Measures on the L2 Unknown Target Words and Vocabulary Tests in the Bilingual Subtitles Group (RQ4)
- Appendix S19. Mixed-Effects Models with Significant Results for the Relationship Between Eye-Movement Measures on the L2 Unknown Target Words and Vocabulary Tests in the Captions Group (RQ4)
- Appendix S20. Mixed-Effects Models for the Relationship Between Eye-Movement Measures on the L1 Translations of Unknown Target Words and Vocabulary Tests (RQ4)

## Appendix S1. Initial Online Viewing Habits Questionnaire

### 视频观看习惯调查问卷

#### Online Viewing Habits Questionnaire

本问卷包含 10 道题，预计总用时 3 分钟。本问卷旨在了解中国的英语学习者使用字幕的基本情况。本问卷采取匿名形式，所收集到的数据将只被用于我的博士论文研究。您的个人资料以及问卷答案将会被严格保密。非常感谢您的慷慨帮助! :-)

关于此问卷如您有任何疑问，可通过以下方式联系我：

研究者：王安荻

邮箱：xxxxx@ucl.ac.uk

#### 注意 Notice:

该项目的数据将由伦敦大学学院（UCL）进行管理。UCL 数据保护办公室负责监督涉及个人数据处理的活动，您可通过 data-protection@ucl.ac.uk 与其联系。有关 UCL 如何使用参与者信息的更多信息，请访问：

[www.ucl.ac.uk/legal-services/privacy/participants-health-and-care-research-privacy-notice](http://www.ucl.ac.uk/legal-services/privacy/participants-health-and-care-research-privacy-notice)。有关数据自主权的详细联系方式和详细信息，请访问信息专员办公室 ICO 网站：<https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/>

我已阅读并理解上述内容，并明白我有权考虑以上信息，提出疑问并得到充分解释。I confirm that I have read and understood the above statement, and have had the opportunity to consider the information, ask questions, and have had these questions adequately answered. [单选题] \*

☐ 是 Yes

☐ 否 No

我的参与是自愿的，我有权无理由地随时退出该项研究。I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. [单选题] \*

☐ 是 Yes

☐ 否 No

我明白我可以拒绝回答部分或全部问题，并且可以随时退出。I know that I can refuse to answer any or all of the questions and that I can withdraw from the questionnaire at any point. [单选题] \*

☐ 是 Yes

☐ 否 No

请仔细阅读并勾选合适的选项 Please carefully read and tick as appropriate  
我的年龄大于 18 岁。I am aged above 18. [单选题] \*

☐ 是 Yes

☐ 否 No

我的母语是中文，我学过/正在学英语。My native language is Chinese, and I learn English as a foreign language. [单选题] \*

☐ 是 Yes

☐ 否 No

1. 您的性别是? What is your gender? [单选题] \*

☐ A. 男 Male

☐ B. 女 Female

☐ C. 不愿透露 Prefer not to say

2. 您的年龄是? What is your age? [单选题] \*

☐ A. 18 - 25

☐ B. 26 - 30

☐ C. 31 - 40

☐ D. 41 - 50

☐ E. 50 +

3. 您认为您的英语水平在哪个级别? What do you think your English level is? [单选题] \*

☐ A. 初学者 Beginner

☐ B. 低中等 Low intermediate

☐ C. 高中等 High intermediate

○D. 准高级 Low advanced

○E. 高级 High advanced

4. 我的英语成绩和考试年份是: My overall score in the language proficiency examinations, and the year of taking it is:[矩阵文本题] \*

注: 没有参与某项考试的请填 0

|                                |       |
|--------------------------------|-------|
| 雅思成绩 IELTS:                    | _____ |
| 雅思考试年份 Year:                   | _____ |
| 托福成绩 TOEFL:                    | _____ |
| 托福考试年份 Year:                   | _____ |
| 大学英语 6 级成绩 CET-6:              | _____ |
| 6 级考试年份 Year:                  | _____ |
| 其他, 请注明 Other, please specify: | _____ |

5. 您喜欢在空闲时间看英文视频吗(美/英剧,电影,纪录片等)? Do you like watching English videos (films, series, documentaries etc.) as an entertainment? [单选题] \*

○A. 是 Yes

○B. 否 No

6. 您喜欢在英语课上看英文视频吗(美/英剧,电影,纪录片等)? Do you like watching English videos (films, series, documentaries etc.) in the classroom? [单选题] \*

○A. 是 Yes

○B. 否 No

7. 在观看英文视频(美/英剧, 电影, 纪录片等)时 When watching English videos: [矩阵文本题] [输入 1 到 6 的数字] \*

注: 1 为“几乎不 hardly ever”; 6 为“非常频繁 very often”。

|                                                                                                                               |       |
|-------------------------------------------------------------------------------------------------------------------------------|-------|
| 您不使用任何字幕的频率为 How often do you watch them without subtitles?                                                                   | _____ |
| 您使用字幕(包括中、英和双语字幕)的频率为 How often do you watch them with any type of subtitles (including English, Chinese, or dual subtitles)? | _____ |
| 您使用双语字幕(中英文同时出现)的频率为 How often                                                                                                | _____ |

|                                                                 |       |
|-----------------------------------------------------------------|-------|
| do you watch them with dual subtitles?                          |       |
| 您使用中文字幕的频率为 How often do you watch them with Chinese subtitles? | _____ |
| 您使用英文字幕的频率为 How often do you watch them with English subtitles? | _____ |

8. 您在下述两种情况下看英文视频的频繁程度大致为 How often do you watch English videos (film, series, documentaries, etc.) in the following situations?[矩阵文本题] [输入 1 到 6 的数字] \*

注：1 为“几乎不 hardly ever”；6 为“非常频繁 very often”。

|                          |       |
|--------------------------|-------|
| 在英语课上 In the classroom   | _____ |
| 在空闲时间 In your spare time | _____ |

9. 您以前是否听说过“双语字幕”(中英文字幕同时出现)? Have you ever heard of “dual/bilingual subtitles” (English and Chinese presented at the same time) before? [单选题] \*

○A. 是 Yes

○B. 否 No

10. 您对下列字幕的喜爱程度为 How much do you like the following types of subtitles:[矩阵文本题] [输入 1 到 6 的数字] \*

注：1 为“一点也不喜欢 not at all”；6 为“非常喜欢 very much”。

|                                 |       |
|---------------------------------|-------|
| 中文字幕 Chinese subtitles          | _____ |
| 英文字幕 English subtitles          | _____ |
| 双语字幕（中英文同时出现）<br>Dual subtitles | _____ |
| 无字幕 No subtitles                | _____ |

11. 如您有任何关于视频观看习惯或字幕使用习惯的想法，请在下方告诉我 If you have any comments about your viewing habits and subtitle use, please feel free to write down here: [填空题]

## Appendix S2. Approved Ethics Form

### Doctoral Student Ethics Application Form

Anyone conducting research under the auspices of the Institute of Education (staff, students or visitors) where the research involves human participants or the use of data collected from human participants, is required to gain ethical approval before starting. This includes preliminary and pilot studies. Please answer all relevant questions in simple terms that can be understood by a lay person and note that your form may be returned if incomplete.

#### **Registering your study with the UCL Data Protection Officer as part of the UCL Research Ethics Review Process**

If you are proposing to collect personal data i.e. data from which a living individual can be identified **you must be registered with the UCL Data Protection Office before you submit your ethics application for review.** To do this, email the complete ethics form to [data-protection@ucl.ac.uk](mailto:data-protection@ucl.ac.uk). Once your registration number is received, add it to the form\* and submit it to your supervisor for approval.

If the Data Protection Office advises you to make changes to the way in which you propose to collect and store the data this should be reflected in your ethics application form.

| Section 1 Project details |                                                 |                                                                                                          |
|---------------------------|-------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| a.                        | Project title                                   | An investigation of the effect of dual subtitles on Chinese EFL learners' incidental vocabulary learning |
| b.                        | Student name and ID number (e.g. ABC12345678)   | Andi Wang<br>xxxxxxxx                                                                                    |
| c.                        | <b>*UCL Data Protection Registration Number</b> | Z6364106/2018/11/09                                                                                      |
| c.                        | Supervisor/Personal Tutor                       | Dr Ana Pellicer-Sánchez                                                                                  |

|                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                            |                                   |                              |                          |                                                               |                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|-----------------------------------|------------------------------|--------------------------|---------------------------------------------------------------|-------------------|
| d.                                                                                                                                                                                                                                                                                                                                                                                                                                       | Department                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                            | Culture, Communication, and Media |                              |                          |                                                               |                   |
| e.                                                                                                                                                                                                                                                                                                                                                                                                                                       | Course category (Tick one)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | PhD <input checked="" type="checkbox"/><br>DEdPsy <input type="checkbox"/> | EdD <input type="checkbox"/>      |                              |                          |                                                               |                   |
| f.                                                                                                                                                                                                                                                                                                                                                                                                                                       | If applicable, state who the funder is and if funding has been confirmed.                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                            |                                   |                              |                          |                                                               |                   |
| g.                                                                                                                                                                                                                                                                                                                                                                                                                                       | Intended research start date                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                            | October 1st, 2018                 |                              |                          |                                                               |                   |
| h.                                                                                                                                                                                                                                                                                                                                                                                                                                       | Intended research end date                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                            | January 31st, 2022                |                              |                          |                                                               |                   |
| i.                                                                                                                                                                                                                                                                                                                                                                                                                                       | Country fieldwork will be conducted in<br><br><i>If research to be conducted abroad please check <a href="http://www.fco.gov.uk">www.fco.gov.uk</a> and submit a completed travel risk assessment form (see guidelines). If the FCO advice is against travel this will be <b>required</b> before ethical approval can be granted:</i><br><br><a href="http://ioe-net.inst.ioe.ac.uk/about/profservices/international/Pages/default.aspx">http://ioe-net.inst.ioe.ac.uk/about/profservices/international/Pages/default.aspx</a> |                                                                            |                                   |                              |                          |                                                               |                   |
| j.                                                                                                                                                                                                                                                                                                                                                                                                                                       | Has this project been considered by another (external) Research Ethics Committee?<br><table border="1"> <tr> <td>Yes <input type="checkbox"/></td> <td>External Committee Name:</td> </tr> <tr> <td>No <input checked="" type="checkbox"/> <i>go to Section 2</i></td> <td>Date of Approval:</td> </tr> </table>                                                                                                                                                                                                               |                                                                            |                                   | Yes <input type="checkbox"/> | External Committee Name: | No <input checked="" type="checkbox"/> <i>go to Section 2</i> | Date of Approval: |
| Yes <input type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                             | External Committee Name:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                            |                                   |                              |                          |                                                               |                   |
| No <input checked="" type="checkbox"/> <i>go to Section 2</i>                                                                                                                                                                                                                                                                                                                                                                            | Date of Approval:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                            |                                   |                              |                          |                                                               |                   |
| <b>If yes:</b> <ul style="list-style-type: none"> <li>– Submit a copy of the approval letter with this application.</li> <li>– Proceed to Section 10 Attachments.</li> </ul>                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                            |                                   |                              |                          |                                                               |                   |
| <b>Note:</b> Ensure that you check the guidelines carefully as research with some participants will require ethical approval from a different ethics committee such as the <a href="#">National Research Ethics Service</a> (NRES) or <a href="#">Social Care Research Ethics Committee</a> (SCREC). In addition, if your research is based in another institution then you may be required to apply to their research ethics committee. |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                            |                                   |                              |                          |                                                               |                   |

## Section 2 Research methods summary (tick all that apply)

|                                                                                                                                               |                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> Interviews<br><input type="checkbox"/> Focus groups<br><input checked="" type="checkbox"/> Questionnaires | <input checked="" type="checkbox"/> Controlled trial/other intervention study<br><input type="checkbox"/> Use of personal records |
|-----------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> Action research<br><input type="checkbox"/> Observation<br><input checked="" type="checkbox"/> Literature review | <input type="checkbox"/> Systematic review <i>if only method used go to Section 5.</i><br><input type="checkbox"/> Secondary data analysis <i>if secondary analysis used go to Section 6.</i><br><input type="checkbox"/> Advisory/consultation/collaborative groups<br><input type="checkbox"/> Other, give details:<br><br>The controlled trial/intervention will involve recording participants' eye movements with a head-free eye-tracker. |
|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Please provide an overview of the project, focusing on your methodology. This should include some or all of the following: purpose of the research, aims, main research questions, research design, participants, sampling, data collection (including justifications for methods chosen and description of topics/questions to be asked), reporting and dissemination. Please focus on your methodology; the theory, policy, or literary background of your work can be provided in an attached document (i.e. a full research proposal or case for support document). *Minimum 150 words required.*

**Purpose of the research:**

Vocabulary is one of the key components in language learning. Previous studies have shown that watching subtitled foreign language videos could benefit foreign language learners' vocabulary development. The effectiveness of different subtitle conditions has been examined in previous research. Dual subtitles is a combination of native and foreign language subtitles appearing at the bottom of the screen, and it has been widely used in China. Current research findings about the benefits of this dual subtitle condition are rather mixed. Therefore, the purpose of this study is to investigate whether the use of dual subtitles while watching foreign language videos could benefit viewers' vocabulary learning as well as their comprehension, compared with no subtitles and other subtitling conditions. Factors affecting the effectiveness of dual subtitles will also be examined.

**Aims**

The aim of the proposed study is to gain a clear picture of the effectiveness of dual subtitles, when comparing with L1 (first language), L2 (second language), and no subtitles, on Chinese EFL (English as a Foreign Language) learners' incidental vocabulary learning. Besides, eye-tracking method will be applied to explore learners' processing of different sources of information when using dual subtitles. The relationships between the processing of subtitles and learners' vocabulary gains as well as comprehension scores will also be investigated, taking learners' vocabulary size and working memory capacity into account.

### **Main research questions**

1. Does the use of dual subtitles incidentally facilitate vocabulary acquisition of Chinese EFL learners compared with L1, L2 and no subtitles?
2. How does the on-line processing (eye-tracking data) in dual subtitles compare to other three types of subtitle conditions in relation to allocation of attention to the several sources of input?
3. Does the use of dual subtitles lead to higher comprehension scores among Chinese EFL learners compared with other subtitle groups?
4. What is the relationship between the on-line processing and the off-line measures of vocabulary learning?
5. Does learners' vocabulary size affect their on-line processing, vocabulary learning and comprehension?
6. Does learners' working memory capacity affect their on-line processing, vocabulary learning and comprehension?

### **Method**

#### **a. Online questionnaire (details please see *Attachment 3*)**

Due to the limited number of empirical studies investigating the use of dual subtitles, a short online questionnaire will be conducted to first gain a general understanding of the use of subtitles among Chinese EFL learners. The participants at this stage will be Chinese EFL learners whose native language is Mandarin, aged between 18 and 30. They will be recruited online and will only be asked to complete the online questionnaire, and will not be involved in the sessions identified below. The expected number of the participants at this stage is more than 200 people. Basic personal information (gender, age, nationality, native language, length of studying English, English proficiency level) and participants' habits of using different types of subtitles (for example the frequency of use, participants' attitudes towards dual subtitles etc.) will be collected anonymously. The questionnaire will be designed by the researcher, and an online questionnaire platform (Wenjuanxing Website: <https://www.wjx.cn/>) will be used to help with online data collection. This agency has been used by many Chinese universities and research institutions for research purposes (for example Peking University, Tsinghua University, Shanghai Jiao Tong University etc.), thus, it is believed to be reliable and confidential for research purposes. According to the official website (<https://www.wjx.cn/wjx/license.aspx>), this platform will not disclose any personal data for any purposes.

The recruitment of participants at this stage will take place online, the questionnaire link will be sent out via Wechat and Weibo (two Chinese social-media APPs which have been widely used in China, known as the Chinese version of Facebook and Twitter) in order to recruit participants. The questionnaire link is plan to be open between November 2018 and June 2019. The purpose of this questionnaire and the consent requirement will be embedded in the online questionnaire. Once the potential participants click on the survey link, the first thing they will see is a paragraph

explaining the purpose of the study, then they will be asked to read five statements about their participation and tick them if they agree to participate (for details, please see *Attachment 3*). Data will only be collected with the participants who meet the above-mentioned criteria (Chinese EFL learners aged between 18 and 30), they will not be able to proceed to the questions until they have ticked “Yes” with those statements to show their qualifications and willingness to participate.

b. Experiment and interview (sample interview questions please see *Attachment 5*)

***Participants and Recruitment:*** The participants will be around 80 Chinese postgraduate students in London. They will be adult speakers of English as a second language with native language as Mandarin Chinese. To recruit participants, an email will be distributed to MA students in the Department of Culture, Communication, and Media. The email will be sent by the programme administrators and will include a brief description of what participation in the study will involve. Potential participants will be asked to contact the researcher to get more information about the study and arrange a time for their participation. If this fails to recruit enough participants, recruitment information will be posted on social-media platforms, and flyers of the recruitment information will be posted on the permitted areas in universities in London for recruitment. Participants will also be encouraged to introduce their friends who meet the criteria (Chinese as first language, intermediate level of English proficiency) and have an interest to participate in this study. Interested potential participants will be given opportunities to ask questions about the project before they decide whether to take part or not. Information Sheet and Consent Form will be provided for interested participants prior to participating in the study.

A pilot study will be conducted with an extra eight participants several months before the main study with the purpose of verifying the study procedure and make any necessary amendments. The recruitment information will be posted on my Wechat to recruit 8 qualified participants for the pilot study. Apart from the recruiting method, these eight participants will go through the same procedure as will be explained below, and receive the same information. They will also receive the same incentives as the participants in the main study.

***Design:*** The experiment consists of two sessions. In the first session, participants will be asked to complete a set of tests, including: a pre vocabulary test (paper and pencil), a vocabulary size test (computer based), and a working memory test (computer based). In the second session, participants will be asked to watch an English video clip (around 30 minutes) with one of four subtitling conditions (L1, L2, dual, or no subtitles) while their eye-movements are recorded with an EyeLink 1000+ Eye-tracker individually. This is a head-free remote eye-tracker. Participants will be randomly assigned to one of these four viewing conditions (20 participants per group). After the eye-tracking experiment, participants will be asked to complete a post vocabulary test, a reading

comprehension test, and a language background questionnaire (see *Attachment 4*). Finally, a stimulated recall interview will be conducted. Participants will be shown their eye movement video recorded in the previous viewing session, and they will be interviewed about their thoughts and experience when viewing with subtitles (or not), how vocabulary knowledge affected their comprehension, and their attitudes towards the particular type of subtitle on vocabulary learning. The interviews will be audio recorded.

**Procedure:** This empirical study mainly consists of the following steps:

1. Piloting: Apart from the 80 participants, around 8 participants will be asked to pilot the following steps intended to modify the test procedure.
2. The first session: A convenient time will be scheduled for the first face-to-face session. Participants will first be provided with the Information Sheet (see *Attachment 1*) in hardcopy. If they agree to participate, they will be asked to sign the Consent Form (see *Attachment 2*). Then a quick paper and pencil pre-vocabulary test will be conducted, followed by computerised vocabulary size test and working memory test. The whole procedure will take around 40 minutes in the eye-tracking Lab (The Space, Level 4, IoE).
3. The second session (experiment and interview): Time of individual session will be scheduled at participants' convenience (also taking into consideration the lab availability) at/after the first session for the experiment and interview. On the scheduled day, data will be collected individually in the eye-tracking Lab (xxxxx, IoE) by the researcher. The equipment will then be set up and calibrated. Then participants will be asked to watch an English video clip on the computer screen while the eye-tracker records their eye movements. After the viewing activity, they will be asked to complete a post vocabulary test, a reading comprehension test, and a language background questionnaire. Then they will watch their own eye movement recording during the viewing activity, and be interviewed to recall their viewing experience in order to gather their opinions about the subtitles and unknown vocabulary in the video, and their attitudes towards the particular type of subtitle on vocabulary learning. The specific interview questions will be refined after the pilot test, and they will also be based on individual eye movement behaviour (see *Attachment 5* for sample preliminary questions). Participants will be aware that the interviews will be audio-recorded as indicated in the Information Sheet and Consent form. The viewing activity, the series of tests, and the interviews will be completed in the same session and will last around 1.5 hour. Participants could take a break after the viewing activity at any time if they need. Each participant will receive a £10 Amazon Voucher for their participation.

**Analysis:** The experiment and eye tracking data will be coded and analysed by the researcher using a variety of programmes (DataViewer, Excel, and SPSS). The interview data will be transcribed and analysed by the researcher.

### Reporting and dissemination

The results will be reported in my upgrading document and my PhD thesis, and the results may also be presented at professional conferences and reported in research publications. A summary on the overall results of the study will be emailed to the participants when the project is completed. The identity of participants will always remain anonymous.

### Section 3 Research Participants (tick all that apply)

- ☐ Early years/pre-school
- ☐ Ages 5-11
- ☐ Ages 12-16
- ☐ Young people aged 17-18

- ☒ Adults *please specify below*
- ☐ Unknown – specify below
- ☐ No participants

The participants will be Chinese adult speakers of English as a second language. At the questionnaire stage, the participants will be recruited online. At the experiment stage, participants will be recruited among postgraduate students at UCL and other universities in London.

**NB:** Ensure that you check the guidelines carefully as research with some participants will require ethical approval from a different ethics committee such as the [National Research Ethics Service](#) (NRES) or [Social Care Research Ethics Committee](#) (SCREC).

### Section 4 Security-sensitive material (only complete if applicable)

Security sensitive research includes: commissioned by the military; commissioned under an EU security call; involves the acquisition of security clearances; concerns terrorist or extreme groups.

|    |                                                                                                                       |                                   |                                        |
|----|-----------------------------------------------------------------------------------------------------------------------|-----------------------------------|----------------------------------------|
| a. | Will your project consider or encounter security-sensitive material?                                                  | Yes<br><input type="checkbox"/> * | No <input checked="" type="checkbox"/> |
| b. | Will you be visiting websites associated with extreme or terrorist organisations?                                     | Yes<br><input type="checkbox"/> * | No <input checked="" type="checkbox"/> |
| c. | Will you be storing or transmitting any materials that could be interpreted as promoting or endorsing terrorist acts? | Yes<br><input type="checkbox"/> * | No <input checked="" type="checkbox"/> |

\* Give further details in **Section 8 Ethical Issues**

### Section 5 Systematic reviews of research (only complete if applicable)

|    |                                                        |                                           |                                        |
|----|--------------------------------------------------------|-------------------------------------------|----------------------------------------|
| a. | Will you be collecting any new data from participants? | Yes <input checked="" type="checkbox"/> * | No <input type="checkbox"/>            |
| b. | Will you be analysing any secondary data?              | Yes <input type="checkbox"/> *            | No <input checked="" type="checkbox"/> |

\* Give further details in **Section 8 Ethical Issues**

*If your methods do not involve engagement with participants (e.g. systematic review, literature review) **and** if you have answered **No** to both questions, please go to **Section 8 Attachments**.*

### Section 6 Secondary data analysis (only complete if applicable)

|    |                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                        |                              |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|------------------------------|
| a. | Name of dataset/s                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                        |                              |
| b. | Owner of dataset/s                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                        |                              |
| c. | Are the data in the public domain?                                                                                                                                                                                                                                                                                                                                                               | Yes <input type="checkbox"/> No <input type="checkbox"/>                                                               |                              |
|    |                                                                                                                                                                                                                                                                                                                                                                                                  | <i>If no, do you have the owner's permission/license?</i><br>Yes <input type="checkbox"/> No* <input type="checkbox"/> |                              |
| d. | Are the data special category personal data (i.e. personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation)? | Yes* <input type="checkbox"/>                                                                                          | No <input type="checkbox"/>  |
| e. | Will you be conducting analysis within the remit it was originally collected for?                                                                                                                                                                                                                                                                                                                | Yes <input type="checkbox"/>                                                                                           | No* <input type="checkbox"/> |
| f. | <b>If no</b> , was consent gained from participants for subsequent/future analysis?                                                                                                                                                                                                                                                                                                              | Yes <input type="checkbox"/>                                                                                           | No* <input type="checkbox"/> |
| g. | <b>If no</b> , was data collected prior to ethics approval process?                                                                                                                                                                                                                                                                                                                              | Yes <input type="checkbox"/>                                                                                           | No* <input type="checkbox"/> |

\* Give further details in **Section 8 Ethical Issues**

*If secondary analysis is only method used **and** no answers with asterisks are ticked, go*

to *Section 9 Attachments*.

## Section 7 Data Storage and Security

*Please ensure that you include all hard and electronic data when completing this section.*

a. **Data subjects - Who will the data be collected from?**

All participants will be Chinese adult speakers of English as a second language. At the questionnaire stage, the participants will be recruited online, including undergraduate, postgraduate, and graduated students.

At the experiment stage, participants will only be recruited among postgraduate students at UCL and other universities in London.

b. **What data will be collected?** Please provide details of the type of personal data to be collected

At the online questionnaire stage, only the questionnaire data will be collected, which include their gender, age, self-rating of English level, habits of watching English videos, and their familiarities of dual subtitles, etc. (details see *Attachment 3*)

At the experiment stage, different types of data will be collected from the participants:

- Eye-movement data (collected in measures of fixations and saccades by the eye-tracker Eye-Link 1000+ during their viewing processes. It will be collected by the eye-tracker and automatically extracted by the eye-tracking software.)
- Performance data (scores on pre- and post- vocabulary test, reading comprehension test)
- Vocabulary size test (quick computerized vocabulary size test)
- Working memory test (quick computerized working memory test)
- Language background questionnaire (collecting background information about their gender, age, first language, second or any additional language, self-rating of linguistic skills, IELTS scores, habits of using subtitles, etc.) (details see *Attachment 4*)

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p><b>c Is the data anonymised?</b></p> <p>Yes <input checked="" type="checkbox"/> No* <input type="checkbox"/></p> <p>Do you plan to anonymise the data?</p> <p>Yes* <input checked="" type="checkbox"/> No <input type="checkbox"/></p> <p>Do you plan to use individual level data?</p> <p>Yes* <input type="checkbox"/> No <input checked="" type="checkbox"/></p> <p>Do you plan to pseudonymise the data?</p> <p>Yes* <input type="checkbox"/> No <input checked="" type="checkbox"/></p> <p><i>* Give further details in <b>Section 8 Ethical Issues</b></i></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|  | <p><b>i. Disclosure</b> – Who will the results of your project be disclosed to?</p> <p>My supervisors and I will have access to the data and personal information. The results of the project will be reported in my upgrading document and thesis. The results may also be presented at professional conferences and reported in journal articles.</p> <p><b>ii. Disclosure</b> – Will personal data be disclosed as part of your project?</p> <p>No personal data will be disclosed. The identity of participants will always remain anonymous.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|  | <p><b>Data storage</b> – Please provide details on how and where the data will be stored i.e. UCL network, encrypted USB stick**, encrypted laptop** etc.</p> <p>Electronic data will be stored in two main locations: the UCL computer where we have the eye-tracking software in the lab and the researchers' personal laptop. Data in the lab computer will only be accessed from the university premises. This computer is accessed through a specific username and password, and only the researchers using the lab have access to. The personal laptop will be encrypted, and specific documents will be saved with password protection. Encrypted USB stick will also be used to temporarily store data for conference report use, and the data will be deleted when they are no longer needed. All the paper-based data will be stored in a locked cabinet in my accommodation. The results of the background questionnaire, performance measures, computerized vocabulary size measure, and computerized working memory measure will be entered in an Excel file. This file will be stored as the electronic data as explained above.</p> |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| <p>** Advanced Encryption Standard 256 bit encryption which has been made a security standard within the NHS</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                            |
| <p>g. <b>Data Safe Haven (Identifiable Data Handling Solution)</b> – Will the personal identifiable data collected and processed as part of this research be stored in the UCL Data Safe Haven (mainly used by SLMS divisions, institutes and departments)?</p>                                                                                                                                                                                                                                                                                                                                | <p>Yes <input checked="" type="checkbox"/> No <input type="checkbox"/></p> |
| <p>h. How long will the data and records be kept for and in what format?</p> <p>The hardcopy data will be kept up to three years, and digital format data will be kept for up to ten years. Data will be destroyed after the specified time.</p> <p>Will personal data be processed or be sent outside the European Economic Area? (If yes, please confirm that there are adequate levels of protections in compliance with GDPR and state what these arrangements are)</p> <p>No.</p> <p>Will data be archived for use by other researchers? (If yes, please provide details.)</p> <p>No.</p> |                                                                            |
| <p>i. If personal data is used as part of your project, describe what measures you have in place to ensure that the data is only used for the research purpose e.g. pseudonymisation and short retention period of data'</p> <p>Participants will be assigned a participant number and this participant number will be given to all the tests and interview so that we can then relate the performance in the different measures, without enclosing their identities.</p>                                                                                                                      |                                                                            |
| <p>* Give further details in <b>Section 8 Ethical Issues</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                            |

## Section 8 Ethical issues

Please state clearly the ethical issues which may arise in the course of this research and how will they be addressed.

**All** issues that may apply should be addressed. Some examples are given below, further information can be found in the guidelines. *Minimum 150 words required.*

|                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>– Methods</li> <li>– Sampling</li> <li>– Recruitment</li> <li>– Gatekeepers</li> <li>– Informed consent</li> <li>– Potentially vulnerable participants</li> <li>– Safeguarding/child protection</li> <li>– Sensitive topics</li> </ul> | <ul style="list-style-type: none"> <li>– International research</li> <li>– Risks to participants and/or researchers</li> <li>– Confidentiality/Anonymity</li> <li>– Disclosures/limits to confidentiality</li> <li>– Data storage and security both during and after the research (including transfer, sharing, encryption, protection)</li> <li>– Reporting</li> <li>– Dissemination and use of findings</li> </ul> |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### **Informed consent**

As explained in previous sections, for the online questionnaire, the research purpose and consent information will be embedded in the online questionnaire. Data will only be collected after obtaining the consent from participants, participants will not be able to proceed to the questions unless they provide their consent by ticking the right boxes to show their willingness to participate. As for the experiment and interview part, printed Information Sheet and Consent Form will be provided on the first session. Before signing the consent form, participants will be given sufficient time to read the documents and they will have opportunities to ask questions about the research. They will be free to decide whether they would like to participate in the research or not. Data will only be collected after receiving the signed consent form.

### **Potentially vulnerable participants**

No vulnerable groups will be targeted. All research participants will be adults.

### **Sensitive topics**

No 'sensitive' data under the definition of the Data Protection Act 1998 will be collected as part of the elicitation instruments. The video clip used in this project will not include any sensitive topic.

### **Risks to participants and/or researchers**

First, since the main aim of this research focuses on incidental vocabulary learning, the “incidental” nature will not allow participants to be informed that they will be tested on vocabulary, which is believed may attract their extra attention to the vocabulary while viewing. Consequently, in the Information Sheet, the project title will be modified into: *An investigation of the use of video subtitles in Chinese EFL learners’ English learning*, in order not to disclose the real purpose of the research. In the experiment instructions, participants will not be informed the post vocabulary test before the viewing activity, but only be informed the existence of a reading comprehension test, language background questionnaire, and interview. This might “surprise” them when they are asked to complete the post vocabulary tests after their viewing. However, there will not

be any potential harm in the post vocabulary tests, besides, clear explanations of this research will be given after the whole experiment. In addition, all participants will be made aware of their rights to withdraw their participation at any time.

Second, the participants will be asked to watch a video clip on screen, and this might be tiring for some participants. Therefore, the length of the viewing activity will be kept to around 30 minutes, which is believed shorter than an episode of a general English series. Besides, the selection of video will accommodate potential participants' preferences and will be adapted to their English level. In addition, participants will be made aware of their freedom to take a break after the viewing activity.

Third, some of the participants might not be familiar with eye tracker, so they will be informed and reassured that viewing with their eye movements recorded will impose no more risk than using computers.

### **Confidentiality and anonymity**

All data will honour assurances of confidentiality and anonymity. All the data in this research will be anonymised. After the students have signed on the consent forms, they will be given a number for identification. It is important to be able to relate performance of a participant across the different tests. Assigning a participant number will allow the researcher to do this but none of the tests will be linked to a particular identity.

### **Data storage/security**

All research data will be stored in line with the UCL's Information Security Management Policy. As outline above, the data will be kept in security during and after the project. All data will be kept securely: digital data will be kept in a password protected systems and laptop (and USB if needed) to which only I have access. Paper-based data will be stored in a locked cabinet in my accommodation.

### **Incentives**

Participants who have completed the first and second session will receive a £10 Amazon Voucher as financial incentive for their participation in the two sessions. Participants will only receive the incentive at the end of the second session, which means if the participant decides to leave and withdraw from the experiment, they will not receive the incentive, this be made clear to the participants in the Information Sheet. Participants will also be informed that the incentive will be the same regardless of their performance in the different experimental tests.

### **Reporting**

The results will be reported in my upgrading document and PhD thesis, and the results may also be presented at professional conferences and in research publications. The identity of the participants will always remain anonymous.

### **Dissemination and use of findings**

Participants will receive a summary about the overall results of the research via email upon request when the research is completed. No individual level information will be provided. The findings of this research will further our understanding about the effectiveness of using subtitles when watching English videos. Chinese EFL learners and EFL classroom practitioners may also be potential consumers of the findings in order to improve vocabulary learning and teaching.

Please confirm that the processing of the data is not likely to cause substantial damage or distress to an individual Yes ☒

## Section 9 Attachments Please attach the following items to this form, or explain if not attached

|                                                         |                                                                                                                                                  |                                         |                             |
|---------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|-----------------------------|
| a.                                                      | Information sheets, consent forms and other materials to be used to inform potential participants about the research<br>(List attachments below) | Yes <input checked="" type="checkbox"/> | No <input type="checkbox"/> |
| Attachment 1: Information Sheet                         |                                                                                                                                                  |                                         |                             |
| Attachment 2: Consent Form                              |                                                                                                                                                  |                                         |                             |
| Attachment 3: Online questionnaire                      |                                                                                                                                                  |                                         |                             |
| Attachment 4: English language background questionnaire |                                                                                                                                                  |                                         |                             |
| Attachment 5: Preliminary interview questions           |                                                                                                                                                  |                                         |                             |
|                                                         | <b>If applicable/appropriate:</b>                                                                                                                |                                         |                             |
| b.                                                      | Approval letter from external Research Ethics Committee                                                                                          | Yes <input type="checkbox"/>            |                             |
| c.                                                      | The proposal ('case for support') for the project                                                                                                | Yes <input type="checkbox"/>            |                             |
| d.                                                      | Full risk assessment                                                                                                                             | Yes <input type="checkbox"/>            |                             |

## Section 10 Declaration

I confirm that to the best of my knowledge the information in this form is correct and that this is a full description of the ethical issues that may arise in the course of this project.

|                                                                                                                                                                                                          |                                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|
| es                                                                                                                                                                                                       | No                                  |
| I have discussed the ethical issues relating to my research with my supervisor.                                                                                                                          |                                     |
| <input checked="" type="checkbox"/>                                                                                                                                                                      | <input type="checkbox"/>            |
| I have attended the appropriate ethics training provided by my course.                                                                                                                                   |                                     |
| <input type="checkbox"/>                                                                                                                                                                                 | <input checked="" type="checkbox"/> |
| <p><b>I confirm that to the best of my knowledge:</b></p> <p>The above information is correct and that this is a full description of the ethics issues that may arise in the course of this project.</p> |                                     |
| Name                                                                                                                                                                                                     | Andi Wang                           |
| Date                                                                                                                                                                                                     | 21 <sup>st</sup> September, 2018    |

**Please submit your completed ethics forms to your supervisor for review.**

**Notes and references**

### Professional code of ethics

You should read and understand relevant ethics guidelines, for example:

[British Psychological Society](#) (2018) *Code of Ethics and Conduct*

or

[British Educational Research Association](#) (2018) *Ethical Guidelines*

or

[British Sociological Association](#) (2017) *Statement of Ethical Practice*

Please see the respective websites for these or later versions; direct links to the latest versions are available on the Institute of Education

<http://www.ucl.ac.uk/ioe/research/research-ethics>

### Disclosure and Barring Service checks

If you are planning to carry out research in regulated Education environments such as Schools, or if your research will bring you into contact with children and young people (under the age of 18), you will need to have a Disclosure and Barring Service (DBS) CHECK, before you start. The DBS was previously known as the Criminal Records Bureau (CRB) . If you do not already hold a current DBS check, and have not registered with the DBS update service, you will need to obtain one through at IOE.

Ensure that you apply for the DBS check in plenty of time as will take around 4 weeks, though can take longer depending on the circumstances.

### Further references

The [www.ethicsguidebook.ac.uk](http://www.ethicsguidebook.ac.uk) website is very useful for assisting you to think through the ethical issues arising from your project.

Robson, Colin (2011). *Real world research: a resource for social scientists and practitioner researchers* (3rd edition). Oxford: Blackwell.

This text has a helpful section on ethical considerations.

Alderson, P. and Morrow, V. (2011) *The Ethics of Research with Children and Young People: A Practical Handbook*. London: Sage.

This text has useful suggestions if you are conducting research with children and young people.

Wiles, R. (2013) *What are Qualitative Research Ethics?* Bloomsbury.

A useful and short text covering areas including informed consent, approaches to research ethics including examples of ethical dilemmas.

### Departmental use

If a project raises particularly challenging ethics issues, or a more detailed review would be appropriate, the supervisor **must** refer the application to the Research Development

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| <p>Administrator (via <a href="mailto:ioe.researchethics@ucl.ac.uk">ioe.researchethics@ucl.ac.uk</a> so that it can be submitted to the IOE Research Ethics Committee for consideration. A departmental research ethics coordinator or representative can advise you, either to support your review process, or help decide whether an application should be referred to the REC. If unsure please refer to the guidelines explaining when to refer the ethics application to the IOE Research Ethics Committee, posted on the committee's website.</p> |                                                                                                          |
| Student name                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Andi Wang                                                                                                |
| Student department                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Culture, Communication, and Media                                                                        |
| Course                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | PhD                                                                                                      |
| Project title                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | An investigation of the effect of dual subtitles on Chinese EFL learners' incidental vocabulary learning |
| <b>Reviewer 1</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                          |
| Supervisor/first reviewer name                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Ana Pellicer-Sanchez                                                                                     |
| Do you foresee any ethical difficulties with this research?                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | I believe all the main ethical issues have been considered in this form.                                 |
| Supervisor/first reviewer signature                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                          |
| Date                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 5/11/2018                                                                                                |
| <b>Reviewer 2</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                          |
| Second reviewer name                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Andrea Revesz                                                                                            |
| Do you foresee any ethical difficulties with this research?                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | I don't see any ethical issues arising other than the ones addressed in the application.                 |
| Supervisor/second reviewer signature                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                          |
| Date                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 5/11/2018                                                                                                |
| <b>Decision on behalf of reviews</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                          |
| Decision                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Approved <input checked="" type="checkbox"/>                                                             |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Approved subject to the following additional measures <input type="checkbox"/>                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Not approved for the reasons given below <input type="checkbox"/>                                        |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Referred to REC for review <input type="checkbox"/>                                                      |
| Points to be noted by other reviewers and in report to REC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                          |
| Comments from reviewers for the applicant                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                          |

*Once it is approved by both reviewers, students should submit their ethics application form to the Centre for Doctoral Education team: [IOE.CDE@ucl.ac.uk](mailto:IOE.CDE@ucl.ac.uk).*

## Appendix S3. Information Sheet

### **Participant Information Sheet for Chinese EFL Learners**

UCL Research Ethics Committee Approval ID Number: Z6364106/2018/11/09

**Title of Study:** An investigation of the use of video for Chinese EFL learners' English learning

**Department:** Communication, Culture and Media

**Name and contact details of the researcher:** Andi Wang (xxxxx@ucl.ac.uk)

**Supervisor:** Dr Ana Pellicer-Sánchez (xxxxx@ucl.ac.uk)

You are being invited to take part in a PhD research project. Before you decide it is important for you to understand the purpose of the research and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

The aim of this project is to examine the effects of different video subtitles on Chinese EFL (English as a Foreign Language) learners' English learning and reading comprehension. The cognitive processes and behaviours are also taken into investigation with the help of an eye-tracker. An additional aim is to assess whether the viewing processes are also influenced by viewers' working memory and their English proficiency level. The results of this study will inform research on the application of multimedia in second language learning, and provide suggestions to facilitate Chinese EFL learners' English language learning. The estimated end of the whole project will be in January 2022.

**Around 80 Chinese learners of English** will be invited to participate. All of them shall be **Chinese postgraduate students who are currently studying in London.**

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. You are free to withdraw from the study at any time without reason and without any impact on your studies or well-being. If you decide to withdraw, any data collected from you will be destroyed. However, you will not receive the incentive as will be mentioned below. If you have any queries about the study, please feel free to ask.

If you decide to participate, two meetings with the researcher are expected in order to complete the following sessions:

In the first session, you will be asked to finish **several vocabulary size tests.**

In the second session, you will be asked to **watch an English video** (around 23 minutes) on a computer screen while your eye movements are recorded with a head-free eye-tracker. After the viewing activity, you will be asked to complete **a reading comprehension test, an online questionnaire and other relevant tests. An interview** will also be conducted to recall and gather your opinions about your viewing experience. The interview will be audio recorded. Both sessions will be completed in the eye-tracking lab (xxxxx, IoE).

The whole procedure will last approximately 2 hours in total, with the first session 30 minutes and the second 1.5 hours. You could take a break after the viewing activity at any time if needed. You will receive a £10 Amazon Voucher for your participation in two sessions, and the voucher will be given at the end of the second session. You have the right to leave and withdraw from the experiment at any time, however, due to the shortage of funds, if this happens, you will not receive the voucher. The incentive will be the same regardless of your performance in the different experimental tests. Travel expenses will not be reimbursed.

The audio recordings of the interview will be used only for analysis and for illustration in conference presentations. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings. They will be transcribed and your name or any other personal details will never be recorded. You will not be able to be identified in any ensuing reports or publications. Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case, we would inform you of any decisions that might limit your confidentiality. At the end of the research, a summary of the general results of the study will be provided upon request.

Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study [at which point you may withdraw your data from the study].

If you have any further questions before you decide whether to take part, please reach me at [xxxxx@ucl.ac.uk](mailto:xxxxx@ucl.ac.uk). If you have any complaints about the project, please contact my Supervisor, Dr Ana Pellicer-Sánchez at [xxxxx@ucl.ac.uk](mailto:xxxxx@ucl.ac.uk). If your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – [ethics@ucl.ac.uk](mailto:ethics@ucl.ac.uk).

#### **Notice:**

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at [data-protection@ucl.ac.uk](mailto:data-protection@ucl.ac.uk). If you remain unsatisfied, you

may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at:  
<https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/>

Further information on how UCL uses participant information can be found here:  
[www.ucl.ac.uk/legal-services/privacy/participants-health-and-care-research-privacy-notice](http://www.ucl.ac.uk/legal-services/privacy/participants-health-and-care-research-privacy-notice).

**Thank you very much for taking the time to read this information sheet and for your consideration to participate!! :-)**

## Appendix S4. Consent Form

### **Evaluation of An Investigation of the Use of Video for Chinese EFL Learners' English Learning Consent for Interviews: Chinese EFL Learners**

(tick as appropriate)

I confirm that I have read and understood this information sheet, and have had the opportunity to consider the information, ask questions, and have had these questions adequately answered.

☐

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

☐

I know that I can refuse to answer any or all of the questions and that I can withdraw from the interview at any point.

☐

I agree for the interview to be recorded, and that recordings will be kept secure and destroyed at the end of the project. I know that all data will be kept under the terms of the General Data Protection Regulation (GDPR).

☐

I agree that small direct quotes may be used in reports (these will be anonymised).

☐

In understand that in exceptional circumstances anonymity and confidentiality would have to be broken, for example, if it was felt that practice was putting children at risk, or there were concerns regarding professional misconduct. In these circumstances advice would be sought from a senior manager from another local authority who will advise us as to the appropriate course of action and as to whether we need to inform the authority of what you have told us.

☐

Name:.....

Signature: ..... Date: .....

Name of researcher:.....

Signature: ..... Date: .....

## Appendix S5. Meaning Recognition Test

请从下列选项中选出每个词的正确中文翻译。如果是你没有见过的词, 请选 E。Please select the correct translation for each item. If you have not met the word before, please choose E.

**1. hump**

- A. 向后甩    B. 被注射镇静剂    C. 培养    D. 弓起    E. 我不知道

**2. bizarre**

- A. 豢养的    B. 古怪的    C. 挤满人的    D. 不可抗拒的    E. 我不知道

**3. rhino**

- A. 角马    B. 溃疡    C. 犀牛    D. 损伤    E. 我不知道

**4. ridiculous**

- A. 蓄意的    B. 相同的    C. 勇敢的    D. 荒唐的    E. 我不知道

**5. surrogate**

- A. 代理    B. 预防药    C. 枢轴    D. 缓解    E. 我不知道

**6. affection**

- A. 喜爱    B. 荣誉    C. 大街    D. 管道    E. 我不知道

**7. interfere**

- A. 协作    B. 干涉    C. 看待    D. 发现    E. 我不知道

**8. bonkers**

- A. 愚蠢的    B. 嗜睡的    C. 充满激情的    D. 迷恋的    E. 我不知道

**9. confident**

- A. 自信的    B. 精密的    C. 成熟的    D. 精确的    E. 我不知道

**10. endearing**

- A. 小巧的    B. 特有的    C. 迷恋的    D. 可爱的    E. 我不知道

**11. knowledge**

- A. 知识    B. 忠诚    C. 生物    D. 忠实    E. 我不知道

**12. fecund**

- A. 愚蠢的    B. 多产的    C. 被关押的    D. 整齐的    E. 我不知道

**13. appear**

A. 实现    B. 断定    C. 说明    D. 出现    E. 我不知道

**14. waddled**

A. 把持    B. 摇摇摆摆地走    C. 向后甩    D. 发出咕噜声    E. 我不知道

**15. traumatised**

A. 被关押的    B. 特有的    C. 受创伤的    D. 古怪的    E. 我不知道

**16. separate**

A. 想象    B. 理解    C. 看待    D. 分开    E. 我不知道

**17. enamoured**

A. 多产的    B. 挤满人的    C. 迷恋的    D. 嗜睡的    E. 我不知道

**18. buffering**

A. 缓解    B. 喧闹    C. 严重破坏    D. 偷猎    E. 我不知道

**19. sedated**

A. 被注射镇静剂    B. 犹豫不决    C. 撞    D. 快速旋转    E. 我不知道

**20. ulcers**

A. 肾上腺    B. 雄象    C. 溃疡    D. 小牛    E. 我不知道

**21. poaching**

A. 改变方向    B. 偷猎    C. 觅食    D. 舒适地躺下    E. 我不知道

**22. captive**

A. 可爱的    B. 豢养的    C. 应急的    D. 愚蠢的    E. 我不知道

**23. nuzzle**

A. 用鼻子蹭    B. 严重破坏    C. 喧闹    D. 代理    E. 我不知道

**24. marvellous**

A. 绝妙的    B. 荒唐的    C. 精密的    D. 未解决的    E. 我不知道

**25. boggling**

A. 培养    B. 把持    C. 犹豫不决    D. 低声说    E. 我不知道

**26. frisson**

A. 枢轴    B. 角马    C. 预防药    D. 兴奋感    E. 我不知道

**27. individual**

A. 个体    B. 知识    C. 忠实    D. 大街    E. 我不知道

**28. calf**

A. 角马    B. 烦扰    C. 小牛    D. 腺体    E. 我不知道

**29. cortisol**

A. 皮质醇    B. 犀牛    C. 幼鸟    D. 肾上腺    E. 我不知道

**30. same**

A. 怪异的    B. 严重的    C. 相同的    D. 蓄意的    E. 我不知道

**加油~还有一半就结束啦! 😊**

**31. gland**

A. 小鹿    B. 腺体    C. 雄象    D. 治疗师    E. 我不知道

**32. obvious**

A. 可共存的    B. 明显的    C. 自信的    D. 极好的    E. 我不知道

**33. confiscated**

A. 跌跌撞撞    B. 弓起    C. 抚摸    D. 没收    E. 我不知道

**34. suckle**

A. 摇摇摆摆地    B. 喂或吃奶    C. 抚摸    D. 低声说    E. 我不知道

**35. barneys**

A. 争执    B. 羚羊    C. 损伤    D. 小马驹    E. 我不知道

**36. combining**

A. 遇到    B. 干涉    C. 分享    D. 使结合    E. 我不知道

**37. serious**

A. 可共存的    B. 严重的    C. 不确定的    D. 极好的    E. 我不知道

**38. bunting**

A. 犹豫不决    B. 撞    C. 使震惊    D. 没收    E. 我不知道

**39. discovered**

A. 冲突    B. 发现    C. 使结合    D. 断定    E. 我不知道

**40. realise**

A. 分享 B. 意识到 C. 分开 D. 冲突 E. 我不知道

**41. sanctuary**

A. 保护区 B. 缠绕 C. 预防药 D. 兴奋感 E. 我不知道

**42. uncertain**

A. 明显的 B. 不确定的 C. 怪异的 D. 勇敢的 E. 我不知道

**43. foraging**

A. 侵扰 B. 用鼻子蹭 C. 觅食 D. 应急 E. 我不知道

**44. foal**

A. 助产士 B. 枢轴 C. 猛犸象 D. 小马驹 E. 我不知道

**45. fulfil**

A. 说明 B. 修剪 C. 实现 D. 攻击 E. 我不知道

**46. attacks**

A. 遇到 B. 攻击 C. 意识到 D. 想象 E. 我不知道

**47. fawn**

A. 治疗师 B. 保护区 C. 羚羊 D. 小鹿 E. 我不知道

**48. twirls**

A. 兴奋感 B. 争执 C. 缠绕 D. 幼鸟 E. 我不知道

**49. loyalty**

A. 忠诚 B. 对峙 C. 智力 D. 危机 E. 我不知道

**50. mature**

A. 绝妙的 B. 精确的 C. 未解决的 D. 成熟的 E. 我不知道

**51. understand**

A. 出现 B. 理解 C. 修剪 D. 协作 E. 我不知道

**52. dinky**

A. 愚蠢的 B. 使尴尬的 C. 受创伤的 D. 小巧的 E. 我不知道

**53. purring**

A. 跌跌撞撞 B. 喂或吃奶 C. 发出咕噜声 D. 快速旋转 E. 我不知道

**54. crisis**

A. 荣誉 B. 个体 C. 智力 D. 危机 E. 我不知道

**55. midwife**

A. 烦扰 B. 圣坛 C. 助产士 D. 皮质醇 E. 我不知道

**56. wildebeest**

A. 猛犸象 B. 圣坛 C. 兴奋感 D. 角马 E. 我不知道

**57. channel**

A. 生物 B. 喜爱 C. 对峙 D. 管道 E. 我不知道

## Appendix S6. Comprehension Test

阅读理解 请把答案写在答题卡上

### 第一段：熊，老虎和狮子

姓名对应：

|             |                    |                  |                |                         |               |
|-------------|--------------------|------------------|----------------|-------------------------|---------------|
| 主持人：<br>Liz | 保护区创始人：<br>Jama 贾玛 | 专家：<br>Clive 克莱夫 | 熊：<br>Baloo 巴鲁 | 老虎：<br>Shere Khan 谢利·可汗 | 狮子：<br>Leo 里奥 |
|-------------|--------------------|------------------|----------------|-------------------------|---------------|

#### 单选题：

- 熊 Baloo 是什么时候来到这个野生动物庇护所的？
  - 当它几个月大的时候
  - 在半年前
  - 在一年前
  - 当它两岁的时候
- 视频中提到关于亚洲黑熊与老虎的表述，下列哪一项是正确的？
  - 当它们相遇时大多数情况下可以和平共处
  - 它们在一般情况下很难相遇
  - 它们所占领地的面积大小相似
  - 它们在远东地区有着相同的领地
- 在视频中的三只动物（熊、老虎和狮子）里，谁是老大？
  - 老虎
  - 狮子
  - 熊
  - 没有提到
- 下列哪项是熊、老虎和狮子被送到这个野生动物庇护所的主要原因？
  - 它们被非法当作宠物饲养
  - 它们伤害了人类
  - 它们被人类虐待身体受伤
  - 它们的品种很珍稀
- 熊、老虎和狮子刚被送到底护所时，为什么整天哭喊？
  - 饲养员们试图把它们分开
  - 它们不适应新的环境
  - 饲养员们没有足够的食物喂它们
  - 它们身体有伤感到疼痛
- Jama 如何解释视频中老虎和熊要打架的行为？

- A.老虎侵占了熊的领地
- B.熊侵占了老虎的领地
- C.它们在打闹玩耍
- D.老虎脾气不好

7. Liz 如何看待视频中老虎和熊发生争执的行为?

- A.很开心看到了熊和老虎真实状态下的行为
- B.很开心看到熊和老虎很快恢复友爱关系
- C.担心熊和老虎被一定程度上地抑制了天性
- D.担心熊和老虎本性显现伤害彼此

8. Clive 如何解释三只动物之间的关系?

- A.不分彼此的亲密兄弟关系
- B.彼此爱护帮助的家人关系
- C.含有对抗试探的兄弟关系
- D.含有等级挑战的家人关系

9. Clive 提到陪伴对三只动物有什么主要的益处?

- A.充满活力
- B.减轻压力
- C.更加专注
- D.保持天性

## **第二段：犀牛和它的朋友**

### **单选题：**

10. 下列关于视频中对小犀牛的表述，哪一项是正确的?

- A.小犀牛没有陪伴依然可以存活
- B.小犀牛非常害怕孤独
- C.小犀牛的独立性会随着年龄增强
- D.小犀牛大约一岁后可以不再依赖妈妈

11. 视频中提到是什么导致了大量犀牛孤儿的产生?

- A.人类偷猎大量屠杀犀牛
- B.犀牛生存环境变化使得寿命缩短
- C.科研需求将成年犀牛隔离
- D.环境污染导致食物来源锐减

12. 视频中提到人类对犀牛的价值存在以下哪种误解?

- A.犀牛皮被认为有商业价值
- B.犀牛皮可以用来做新材料研发
- C.犀牛角可以用来做珍贵的工艺品

D.犀牛角被认为有药用价值

13. 为什么过多的压力会给犀牛带来严重的问题?

- A.因为会使处理压力的激素消耗殆尽
- B.因为会让犀牛过于紧张不愿进食
- C.因为会容易引发其他肠胃疾病
- D.因为会压迫犀牛的眼部神经

14. 据视频中所说, 与其他动物相比, 为什么人类不太适合照顾犀牛幼崽?

- A.因为人类生病时的病菌很容易杀死小犀牛
- B.因为和人类在一起小犀牛会害怕
- C.因为和人类在一起不利于它们适应野生生活
- D.因为人类不会永远和它们在一起

15. 为什么不建议把狗和小犀牛搭配起来养?

- A.因为狗容易伤害小犀牛
- B.因为狗不是食草动物
- C.因为狗无法一直陪在小犀牛身边
- D.因为狗吠声会吓到小犀牛

16. 在视频中所提到犀牛的性格更加倾向于下列哪一种?

- A.害羞胆小
- B.温柔安静
- C.活跃开朗
- D.独立强壮

### 第三段: 鹿和狗

|          |                   |                  |            |
|----------|-------------------|------------------|------------|
| 主持人: Liz | 狗的主人: Isobel 伊莎贝尔 | 鹿: Pip/Pippin 皮平 | 狗: Kate 凯特 |
|----------|-------------------|------------------|------------|

#### 单选题:

17. Liz 在视频中提到 Pip 和 Kate 的关系最为独特的一点是什么?

- A.它们的年龄差距比较大
- B.改变了它们各自的生活习惯
- C.情感建立的过程更加漫长
- D.纯粹源于它们自己的选择

18. 小鹿 Pip 在什么时间可以进入 Isobel 的家?

- A.随时都可以
- B. Isobel 允许的时候
- C. Kate 在门口等她的时候
- D.没有提到

19. 以下关于 Isobel 与 Pip 相识的过程描述准确的是：
- A. Isobel 在森林里发现了小鹿 Pip 后就立刻把她带回了家
  - B. Isobel 在森林里发现了小鹿 Pip 后并没有立刻把她带回家
  - C. Isobel 在家门口发现了小鹿 Pip 后就立刻把她接回了家
  - D. Isobel 在家门口发现了小鹿 Pip 后并没有立刻把她接回家
20. 根据 Isobel 所说的，为什么她把小鹿放在了狗床上？
- A. 因为她想让狗照顾小鹿
  - B. 因为她觉得狗床上比较温暖
  - C. 因为那是唯一可以安放的地方
  - D. 因为狗对小鹿产生了很强的好奇
21. 关于 Kate 给 Pip 喂奶，视频中发生的是哪一种情况？
- A. Kate 有奶，但它从未尝试过给 Pip 喂奶
  - B. Kate 有奶，并且尝试过给 Pip 喂奶
  - C. Kate 没有奶，但它尝试过给 Pip 喂奶
  - D. Kate 没有奶，并且从未尝试过给 Pip 喂奶
22. Pip 开始会撞 Kate 是因为什么？
- A. 在和 Kate 玩耍
  - B. 不喜欢 Kate
  - C. 想要引起注意
  - D. 想要吃奶
23. Pip 是什么时候开始回到野外的？
- A. 当它两周大时
  - B. 当它六周大时
  - C. 当它两个月大时
  - D. 当它六个月大时
24. Pip 多久会回来看 Kate 一次？
- A. 每一天
  - B. 每两天
  - C. 每一周
  - D. 每两周
25. Isobel 如何描述 Pip 与 Kate 玩耍时的样子？
- A. 小鹿 Pip 会像与其他鹿玩耍时一样
  - B. 小鹿 Pip 会更勇敢大胆一些
  - C. 小鹿 Pip 会更小心谨慎一些
  - D. 小鹿 Pip 会更活泼自在一些

26. 下列哪一种更适合描述 Pip 与 Kate 的关系?

- A.从养育关系到朋友关系
- B.从保持防备到互相照顾
- C.从养育到互相照顾的关系
- D.一直是亲密的朋友关系

27. Pip 为什么每年都会回到房子周围?

- A.来生下她的幼崽
- B.带回一些食物报恩
- C.来看 Kate 和它的主人
- D.没有提到

28. 为什么 Isobel 会说 Pip 和 Kate 后来的相处就像老朋友一样?

- A.它们见到彼此非常热情
- B.它们打招呼的方式很平静
- C.它们看彼此的眼神非常信任
- D.它们不需要太多交流

#### 第四段：猫和鸭子

|          |              |             |             |
|----------|--------------|-------------|-------------|
| 主持人: Liz | 丈夫: Ronan 罗南 | 妻子: Emma 艾玛 | 猫: Della 黛拉 |
|----------|--------------|-------------|-------------|

#### 单选题:

29. 最开始找不到小鸭们, 为什么 Ronan 会觉得猫吃了小鸭们?

- A.他听 Emma 说她看到猫吃了小鸭们
- B.他听到了猫的叫声
- C.他看到了地上的蛋壳和绒毛
- D.他看到有一只猫在谷仓里

30. 下列哪一项描述是准确的?

- A. Emma 和 Ronan 知道猫妈妈是先生下小猫然后才发现小鸭子的
- B. Emma 和 Ronan 不知道猫妈妈是先生下小猫然后才发现小鸭子的
- C. Emma 和 Ronan 知道猫妈妈是先发现小鸭子然后才生下小猫的
- D. Emma 和 Ronan 不知道猫妈妈是先发现小鸭子然后才生下小猫的

31. Ronan 觉得如果猫晚几个小时看到小鸭们会出现什么情况?

- A.会呵护小鸭们
- B.会把小鸭们当作食物
- C.会对小鸭们充满敌意
- D.会对小鸭们感到好奇

32. 让 Emma 和 Ronan 感到特别吃惊的是什么？
- A.小鸭子们长得特别快
  - B.小鸭子们会欺负小猫
  - C.小鸭子们依旧学会了游泳
  - D.小鸭子们在吃猫奶
33. Emma 为什么要说服 Ronan 不让他把鸭子和猫分开？
- A.她觉得和猫在一起小鸭子也依旧能学会游泳
  - B.她觉得它们在一起非常可爱
  - C.她担心分开了之后小鸭子就会死掉
  - D.她觉得它们深爱彼此
34. 为什么在视频最后提到猫妈妈有时会不太高兴？
- A.因为它渐渐不喜欢小鸭子们了
  - B.因为小鸭子有时会欺负小猫
  - C.因为小鸭子们太吵了
  - D.因为它发现很难管教小鸭子们

**Translated version (NOT provided to the participants):**

One: The bear, the tiger and the lion

Name correspondence:

|                  |              |             |           |            |           |
|------------------|--------------|-------------|-----------|------------|-----------|
| The interviewer: | The founder: | The expert: | The bear: | The tiger: | The lion: |
| Liz              | Jama         | Clive       | Baloo     | Shere Khan | Leo       |

**Multiple-choice questions:**

1. When did the bear Baloo come to this wildlife refuge?
  - A. When it was a few months old
  - B. Six months ago
  - C. One year ago
  - D. When it was two years old
  
2. According to the video, which following statement is true about Asian black bears and tigers?
  - A. When they meet, they can coexist peacefully on most occasions
  - B. Under normal circumstances, they seldom meet each other
  - C. The size of their territories is similar
  - D. They share the same territory in the Far East

3. According to the video, who was the boss among the three animals (the bear, tiger, and lion)?
- A. The tiger
  - B. The lion
  - C. The bear
  - D. Did not mention
4. Which of the following was the main reason that the bear, tiger and lion were sent to this wildlife refuge?
- A. They were illegally kept as pets
  - B. They hurt humans
  - C. They were abused by humans and injured
  - D. They were rare species
5. Why did the bear, tiger and lion cry all day when they were just sent to the shelter?
- A. The workers tried to separate them
  - B. They were not adapted to the new environment
  - C. The workers did not have enough food to feed them
  - D. They felt pain because of the injured bodies
6. What was Jama's explanation about the behaviour of the tiger and the bear?
- A. The tiger invaded the territory of the bear
  - B. The bear invaded the territory of the tiger
  - C. They were playing
  - D. The tiger had a bad temper
7. What did Liz think about the conflict between the tiger and the bear?
- A. She was very happy to see the natural behaviour of them
  - B. She was very happy to see that they re-established close relationship very soon
  - C. She was worried that their natural instincts were impeded
  - D. She was worried that their natural instincts led them to hurt each other
8. How did Clive explain the relationship between the three animals?
- A. Very intimate brotherhood
  - B. Supportive and loving family relationship
  - C. Brotherhood including rivalry and testing
  - D. Family relationship with status challenge
9. What was the main benefit of companionship for the three animals as mentioned by Clive?
- A. Keeping energetic
  - B. Reducing stress

- C. Being focused
- D. Keeping instinct

## **Two: Rhinos and their friends**

10. According to the video, which of the following statements about the young rhinos is correct?

- A. Young rhinos can survive without companionship
- B. Young rhinos are extremely afraid of loneliness
- C. Young rhinos' independence increases with age
- D. Young rhinos can stop relying on mom after about one year old

11. According to the video, what was the reason that caused many orphaned rhinos?

- A. Human poach and slaughter a great number of rhinos
- B. Rhinos' lives are shortened due to the changes of living environments
- C. Adult rhinos are isolated for scientific research purposes
- D. Environmental pollution causes a sharp drop in food sources of rhinos

12. Which of the following misunderstandings of humans to the value of rhinos was mentioned in the video?

- A. Rhino skin is considered to have commercial value
- B. Rhino skin can be used for new material development
- C. Rhino horns can be used to make precious crafts
- D. Rhino horns are considered to have medicine value

13. Why did excessive stress cause serious problems for the rhinos?

- A. Because the hormones handling the stress were exhausted
- B. Because the rhinos were too nervous to eat food
- C. Because it caused other gastrointestinal diseases
- D. Because it oppressed rhinos' ocular nerves

14. According to the video, why was human less suitable to take care of young rhinos than other animals?

- A. If humans were sick, the bacteria were likely to kill young rhinos
- B. Young rhinos would be afraid to live with humans
- C. Being with humans would constrain rhinos' ability to adapt to the wildlife
- D. Humans could not stay with rhinos forever

15. Why was it not recommended to pair young rhinos with dogs?

- A. Because dogs are likely to hurt young rhinos
- B. Because dogs do not graze

- C. Because dogs cannot always stay with young rhinos
- D. Because dog barking will scare young rhinos

16. According to the video, which of the followings is more suitable to describe rhinos' personality?

- A. Shy and timid
- B. Gentle and quiet
- C. Active and cheerful
- D. Independent and strong

### Three: The deep and the dog

|                      |                         |                      |               |
|----------------------|-------------------------|----------------------|---------------|
| The interviewer: Liz | The dog's owner: Isobel | The deer: Pip/Pippin | The dog: Kate |
|----------------------|-------------------------|----------------------|---------------|

17. According to Liz's explanation, what was the most unique feature of the relationship between Pip and Kate?

- A. Their age difference was relatively large
- B. Their living habits were affected and changed
- C. The process of the relationship establishment was quite long
- D. The relationship was purely from their own choices

18. When could the deer Pip enter Isobel's house?

- A. At any time
- B. When Isobel permitted
- C. When Kate waited at the door
- D. Did not mention

19. Which of the following descriptions about Isobel and Pip is accurate?

- A. Isobel took Pip back to her home immediately after she found it in the woods
- B. Isobel did not take Pip home immediately after she found it in the woods
- C. Isobel took Pip back to her home immediately after she found it at the door
- D. Isobel did not take Pip home immediately after she found it at the door

20. According to Isobel, why did she put the deer on the dog's bed?

- A. Because she wanted the dog to take care of the deer
- B. Because she thought the dog's bed was warmer
- C. Because that was the only place to put the deer
- D. Because the dog had a strong curiosity about the deer

21. What was happened in the video about Kate's feeding of Pip?

- A. Kate had milk, but it never tried to feed Pip
- B. Kate had milk, and tried to feed Pip

- C. Kate did not have milk, but it tried to feed Pip
- D. Kate did not have milk, and never tried to feed Pip

22. Why did Pip start to bump against Kate?

- A. It played with Kate
- B. It did not like Kate
- C. It wanted to attract attention
- D. It wanted to have milk

23. When did Pip get back to the wild?

- A. When Pip was 2 weeks old
- B. When Pip was 6 weeks old
- C. When Pip was 2 months old
- D. When Pip was 6 months old

24. How often did Pip come back to see Kate?

- A. Every day
- B. Every two days
- C. Every week
- D. Every two weeks

25. How did Isobel describe Pip's behaviour when playing with Kate?

- A. Pip behaved the same as playing with other deer
- B. Pip was bolder and braver
- C. Pip was more cautious and careful
- D. Pip was happier and more comfortable

26. Which of the following is more suitable to describe the relationship between Pip and Kate?

- A. From parenting to friend
- B. From being precautious to mutual care
- C. From parenting to mutual care
- D. Always as close friend

27. Why did Pip return to the house every year?

- A. Came and gave birth to its children
- B. Brought back some food to show gratitude
- C. Visited Kate and its owner
- D. Did not mention

28. Why did Isobel say that Pip and Kate were like old friends?

- A. They were very happy when seeing each other.

- B. They greeted each other in a very calm way
- C. They looked at each other in a reliable way
- D. They did not need much communication

**Four: The cat and ducks**

|                      |                    |                |                |
|----------------------|--------------------|----------------|----------------|
| The interviewer: Liz | The husband: Ronan | The wife: Emma | The cat: Della |
|----------------------|--------------------|----------------|----------------|

29. Why did Ronan think that the cat had eaten the ducklings when they could not find the ducklings?

- A. He heard from Emma who said she saw the cat was eating the ducklings
- B. He heard the cat's meow
- C. He saw the eggshell and fluff on the ground
- D. He saw a cat in the barn

30. Which of the following statement is true?

- A. The cat gave birth to the kittens before finding out the ducklings, and Emma and Ronan knew that
- B. The cat gave birth to the kittens before finding out the ducklings, but Emma and Ronan did not know that
- C. The cat found the ducklings before giving birth to the kittens, and Emma and Ronan knew that
- D. The cat found the ducklings before giving birth to the kittens, but Emma and Ronan did not know that

31. What did Ronan think would happen if the cat saw the ducklings a few hours later?

- A. It would take care of the little ducklings
- B. It would take the ducklings as food
- C. It would be hostile to the ducklings
- D. It would be curious about the ducklings

32. What surprised Emma and Ronan?

- A. The ducklings grew very fast
- B. The ducklings could bully the kittens
- C. The ducklings still acquired swimming skill
- D. The ducklings sucked the cat's milk

33. Why did Emma convince Ronan not to separate the ducklings from the cat?

- A. She thought that the ducklings could still learn how to swim even being with the cat
- B. She thought they love each other very deeply
- C. She worried that the ducklings would die after being separated
- D. She thought they were very cute being together

34. Why was the cat not very happy sometimes as mentioned at the end of the video?
- A. Because it gradually did not like the ducklings
  - B. Because the ducklings sometimes bullied the kittens
  - C. Because the ducklings were too noisy
  - D. Because it found it was difficult to control the ducklings

## Appendix S7. 3K Vocabulary Levels Test

### The 3000 word level

|                  |                       |              |                       |
|------------------|-----------------------|--------------|-----------------------|
| 1 bull           | _____ formal and      | 1 muscle     | _____ advice          |
| 2 champion       | _____ serious         | 2 counsel    | _____ a place         |
| 3 dignity        | _____ manner          | 3 factor     | _____ covered with    |
| 4 hell           | _____ winner of a     | 4 hen        | _____ grass           |
| 5 museum         | _____ sporting event  | 5 lawn       | _____ female          |
| 6 solution       | _____ building        | 6 atmosphere | _____ chicken         |
|                  | _____ where           | 5 lawn       |                       |
|                  | _____ valuable        | 6 atmosphere |                       |
|                  | _____ objects are     |              |                       |
|                  | _____ shown           |              |                       |
| 1 blanket        | _____ holiday         | 1 abandon    | _____ live in a place |
| 2 contest        | _____ good quality    | 2 dwell      | _____ follow in       |
| 3 generation     | _____ wool covering   | 3 oblige     | _____ order to catch  |
| 4 merit          | _____ used on         | 4 pursue     | _____ leave           |
| 5 plot           | _____ beds            | 5 quote      | _____ something       |
| 6 vacation       |                       | 6 resolve    | _____ permanently     |
| 1 comment        | _____ long formal     | 1 assemble   | _____ look closely    |
| 2 gown           | _____ dress           | 2 attach     | _____ stop doing      |
| 3 import         | _____ goods from a    | 3 peer       | _____ something       |
| 4 nerve          | _____ foreign         | 4 quit       | _____ cry out loudly  |
| 5 pasture        | _____ country         | 5 scream     | _____ in fear         |
| 6 tradition      | _____ part of the     | 6 toss       |                       |
|                  | _____ body which      |              |                       |
|                  | _____ carries feeling |              |                       |
| 1 pond           | _____ group of        | 1 drift      | _____ suffer          |
| 2 angel          | _____ animals         | 2 endure     | _____ patiently       |
| 3 frost          | _____ spirit who      | 3 grasp      | _____ join wool       |
| 4 herd           | _____ serves God      | 4 knit       | _____ threads         |
| 5 fort           | _____ managing        | 5 register   | _____ together        |
| 6 administration | _____ business and    | 6 tumble     | _____ hold firmly     |
|                  | _____ affairs         |              | _____ with your       |
|                  |                       |              | _____ hands           |
| 1 brilliant      | _____ thin            | 1 aware      | _____ usual           |
| 2 distinct       | _____ steady          | 2 blank      | _____ best or most    |
| 3 magic          | _____ without         | 3 desperate  | _____ important       |
| 4 naked          | _____ clothes         | 4 normal     | _____ knowing what    |
| 5 slender        |                       | 5 striking   | _____ is happening    |
| 6 stable         |                       | 6 supreme    |                       |

## Appendix S8. Background Questionnaire for Participants in the Main Study

### 被试填写 — 视频观看习惯调查问卷

#### Viewing Habits Questionnaire for Participants

本问卷包含 16 道题，预计总用时 3 分钟。本问卷旨在了解您的英文学习背景以及使用字幕观看视频的基本情况。本问卷采取匿名形式，所收集到的数据将只被用于我的博士论文研究。您的个人资料以及问卷答案将会被严格保密。非常感谢您的慷慨帮助! :-)

关于此问卷如您有任何疑问，可通过以下方式联系我：

研究者：王安荻

邮箱：xxxxx@ucl.ac.uk

#### 注意 Notice:

该项目的数据将由伦敦大学学院（UCL）进行管理。UCL 数据保护办公室负责监督涉及个人数据处理的活动，您可通过 data-protection@ucl.ac.uk 与其联系。有关 UCL 如何使用参与者信息的更多信息，请访问：

[www.ucl.ac.uk/legal-services/privacy/participants-health-and-care-research-privacy-notice](http://www.ucl.ac.uk/legal-services/privacy/participants-health-and-care-research-privacy-notice)。有关数据自主权的详细联系方式和详细信息，请访问信息专员办公室 ICO 网站：<https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/>

编号 Participant code: [填空题] \*

---

1. 你性别是? What is your gender? [单选题] \*

☐ A. 男 Male

☐ B. 女 Female

☐ C. 不愿透露 Prefer not to say

2. 你的年龄是? What is your age? [填空题] \*

---

3. 你认为自己的英语水平在哪个级别? What do you think your English level is? [单选题] \*

- A. 初学者 Beginner
- B. 低中等 Low intermediate
- C. 高中等 High intermediate
- D. 准高级 Low advanced
- E. 高级 High advanced

4. 你认为自己的四项英语水平为 Please select your proficiency in listening, reading, speaking, and writing English.[矩阵文本题] [输入 1 到 5 的数字] \*

注: 1 为“初学者水平 Beginner level”; 5 为“母语者水平 Native speaker level”

|              |       |
|--------------|-------|
| 听力 Listening | _____ |
| 阅读 Reading   | _____ |
| 口语 Speaking  | _____ |
| 写作 Writing   | _____ |

5. 你的英语成绩和考试年份是: Your overall score in the language proficiency examinations, and the year of taking it is:[矩阵文本题] \*

注: 没有参与某项考试的请填 0

|                                |       |
|--------------------------------|-------|
| 雅思成绩 IELTS:                    | _____ |
| 雅思考试年份 Year:                   | _____ |
| 托福成绩 TOEFL:                    | _____ |
| 托福考试年份 Year:                   | _____ |
| 其他, 请注明 Other, please specify: | _____ |

6. 你的四项雅思成绩分别是: Your four components scores in IELTS are:[矩阵文本题] \*

|               |       |
|---------------|-------|
| 听力 Listening: | _____ |
| 阅读 Reading:   | _____ |
| 口语 Speaking:  | _____ |
| 写作 Writing:   | _____ |

7. 在过去四年中, 你用英文进行读/写/说/听的频率大约为 How often do you read/write/speak/listen to English....[矩阵文本题] [输入 1 到 6 的数字] \*

注: 1 为“几乎不 hardly ever”; 6 为“非常频繁 very often”。

|                                                                  |       |
|------------------------------------------------------------------|-------|
| 阅读英语（以各种形式）的频率为 How often do you read in English (in any way)?   | _____ |
| 听英语（以各种形式）的频率为 How often do you listen to English (in any way)?  | _____ |
| 说英语（以各种形式）的频率为 How often do you speak in English (in any way)?   | _____ |
| 用英文写作（以各种形式）的频率为 How often do you write in English (in any way)? | _____ |

8. 如果你到英国的时间少于三个月, 你在英国用英文进行读/写/说/听的频率大约为 If you have just arrived in the UK within the past three months, how often do you read/write/speak/listen to English in this period of time....[矩阵文本题] [输入 1 到 6 的数字]

注: 1 为“几乎不 hardly ever”; 6 为“非常频繁 very often”。

|                                                                  |       |
|------------------------------------------------------------------|-------|
| 阅读英语（以各种形式）的频率为 How often do you read in English (in any way)?   | _____ |
| 听英语（以各种形式）的频率为 How often do you listen to English (in any way)?  | _____ |
| 说英语（以各种形式）的频率为 How often do you speak in English (in any way)?   | _____ |
| 用英文写作（以各种形式）的频率为 How often do you write in English (in any way)? | _____ |

9. 你喜欢在空闲时间看英文视频吗(美/英剧,电影,纪录片等)? Do you like watching English videos (films, series, documentaries etc.) as an entertainment? [单选题] \*

○A. 是 Yes

○B. 否 No

10. 你喜欢在英语课上看英文视频吗(美/英剧,电影,纪录片等)? Do you like watching English videos (films, series, documentaries etc.) in the classroom? [单选题]

\*

○A. 是 Yes

○B. 否 No

11. 在观看英文视频(美/英剧, 电影, 纪录片等)时 When watching English videos:  
[矩阵文本题] [输入 1 到 6 的数字] \*

注: 1 为“几乎不 hardly ever”; 6 为“非常频繁 very often”。

|                                                                                                                               |       |
|-------------------------------------------------------------------------------------------------------------------------------|-------|
| 你不使用任何字幕的频率为 How often do you watch them without subtitles?                                                                   | _____ |
| 你使用字幕(包括中、英和双语字幕)的频率为 How often do you watch them with any type of subtitles (including English, Chinese, or dual subtitles)? | _____ |
| 你使用双语字幕(中英文同时出现)的频率为 How often do you watch them with dual subtitles?                                                         | _____ |
| 你仅使用中文字幕的频率为 How often do you watch them with Chinese subtitles?                                                              | _____ |
| 你仅使用英文字幕的频率为 How often do you watch them with English subtitles?                                                              | _____ |

12. 你在下述两种情况下看英文视频的频繁程度大致为 How often do you watch English videos (film, series, documentaries, etc.) in the following situations?[矩阵文本题] [输入 1 到 6 的数字] \*

注: 1 为“几乎不 hardly ever”; 6 为“非常频繁 very often”。

|                          |       |
|--------------------------|-------|
| 在英语课上 In the classroom   | _____ |
| 在空闲时间 In your spare time | _____ |

13. 你以前是否听说过“双语字幕”(中英文字幕同时出现)? Have you ever heard of “dual/bilingual subtitles” (English and Chinese presented at the same time) before? [单选题] \*

○A. 是 Yes

○B. 否 No

14. 看英文视频时, 你对下列字幕的喜爱程度为 How much do you like the following types of subtitles when watching English videos:[矩阵文本题] [输入 1 到 6 的数字] \*

注: 1 为“一点也不喜欢 not at all”; 6 为“非常喜欢 very much”。

|                              |       |
|------------------------------|-------|
| 中文字幕 Chinese subtitles       | _____ |
| 英文字幕 English subtitles       | _____ |
| 双语字幕(中英文同时出现) Dual subtitles | _____ |

无字幕 No subtitles

15. 你在英国待了多长时间? How long have you lived in the UK? [填空题] \*

\_\_\_\_\_

16. 你的专业是 (请用英文填写): [填空题] \*

\_\_\_\_\_

## Appendix S9. Stimulated Recall Interview Oral Instructions

### 说明：

现在我们要回看一下几段视频内容。刚才在你看视频的过程中，你眼睛移动的轨迹被记录了下来。我们可以看到，屏幕上的这个红色的点就是你当时眼睛注意的地方。现在我想要知道的是，刚才当你遇到一些你可能不认识的单词时，你在想些什么。这些词会被一个橙色的方框圈住。我们知道在看视频的时候你能够听到英文的语音，能够看到动态的图像，还可以看到 XX 字幕。通过眼动仪的数据我们可以知道你关注了哪些区域，但我无法知道当你在看或听到这些生词时具体在想些什么，你又是如何处理关于这个词的不同信息的。

所以，现在我们要一起看一下一些生词出现时的片段，然后我需要你回想一下刚才看到这些词的时候，你在想些什么。不要告诉我你现在的想法，也不用告诉我你在词汇测试中的回答，你只需要告诉我当时你在想什么。想不起来了就告诉我你忘记了。这个过程我不会和你有太多的互动，所以请你尽可能多地告诉我你当时的想法。我会做一些笔记，在最后再来问你。

Now we will look back at some of the video clips. Just now when you were watching the video, your eye-movement trajectory was recorded. As we can see on the screen, this red dot on the screen is where your eyes were paying attention at the time. Now I want to know, during your viewing of the video, when you encountered some words that you did not know, what were you thinking? These words will be circled by a rectangular box. I know you can hear the English audio, see the dynamic images, and read the subtitles. Also, through the eye-tracking data, I know what you were looking at and for how long, but I cannot know when you heard or saw some words, what you were thinking, and how you processed the different information about this word.

So now we will review several video clips together, you should try your best to recall your thoughts when these words were presented on the video. Please do NOT tell me what you are thinking right now, or your answers to the vocabulary tests. Please only tell me about your thought at THAT time. If you cannot remember, or you did not notice the word, you can simply tell me you do not remember it. I will not interact with you too much during this recall process, so please tell me as much as you can about your thoughts at the time. I will make some notes and ask you at the end.

### Questions for each TW:

“The first word is “XXX”, did you notice this word at that time?”

(If they said yes) “What were you thinking at that time?”

## Appendix S10. Meaning Recall Test Scoring Scheme

| 正确的翻译（中英皆可）给 1 分，错误的给 0 分。有争议的请参照具体要求，谢谢！（按首字母排序） |      |                       |                                                                                             |                                                                      |                                                                   |                                        |
|---------------------------------------------------|------|-----------------------|---------------------------------------------------------------------------------------------|----------------------------------------------------------------------|-------------------------------------------------------------------|----------------------------------------|
| Item                                              | 词性   | 中文翻译                  | English translation                                                                         | Controversial Examples                                               | When to give 1                                                    | When to give 0                         |
| barneys                                           | n.   | 争执                    | A loud argument                                                                             |                                                                      | 与争吵，争执有关<br>Related to argument, confliction.                     |                                        |
| bizarre                                           | adj. | 古怪的；奇怪的；怪诞的；罕见的；异乎寻常的 | Very strange and unusual                                                                    | 荒诞 1 奇异 1 奇怪的事 1<br>奇葩 1 惊奇的事 1<br><br>使困惑 0 令人吃惊，疯狂 0<br>乱的 0 混乱的 0 | 与奇怪，古怪有关<br>Sth related to strange, weird, odd.                   | 与迷惑，乱，混乱有关<br>Sth only puzzled, messy. |
| buffering                                         | v.   | 缓解；缓冲物；起缓冲作用的人        | To provide protection against harm                                                          | 缓冲 1 减缓 1 缓冲区 1<br><br>解脱 0 释放 0                                     | 含有缓，缓冲<br>Including 缓，缓冲                                          | 解脱，挣脱类<br>Just about relieve           |
| bunting                                           | v.   | 撞；（棒球运动中）触击           | (In baseball) To deliberately hit the ball very gently so that the ball does not travel far | 打击 1 冲撞 1<br><br>对抗，打闹 0 攻击 0 跳 0<br>跳跃 0                            | 描述一种撞，击打的状态<br>Describe the related status of bumping and hitting | 攻击或跳跃类<br>Attacking or jumping         |
| confiscated                                       | v.   | 没收；把…                 | To take a possession away from                                                              | 密封的；关押的 0 非法滞                                                        | 含有带走，没收，充公                                                        | 关闭的，放入监                                |

|            |      |                   |                                                                                                                                           |                                                                    |                                                    |                                                                               |
|------------|------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|----------------------------------------------------|-------------------------------------------------------------------------------|
|            |      | 充公                | someone when you have the right to do so, usually as a punishment and often for a limited period, after which it is returned to the owner | 留 0                                                                | 的含义 Relate to taking away                          | 狱<br>Relate to closed, jailed                                                 |
| cortisol   | n.   | 皮质醇               | A hormone (= a chemical made in the body) that is used in medicine to treat parts of the body that are swollen and painful                | a kind of 激素 1<br>激素 1<br><br>a chemical 0 皮质层 0 肉质素 0             | 提到是一种激素<br>Related to the hormone                  | 只说是一种化学物质, 或给出错误的名称翻译<br>The name of another chemical, or only said chemical. |
| dinky      | adj. | 小巧的               | Very small or slight                                                                                                                      | 短小的 1                                                              | 和小有关<br>Related to small                           |                                                                               |
| endearing  | adj. | 可爱的; 使人喜爱的; 引人爱慕的 | Making someone like you; inspiring affection; lovable, adorable                                                                           | 亲密的 1 亲爱的 1 与亲密相关 1 喜爱的 1 喜爱 1 有爱的 1<br><br>珍贵的 0 美妙的 0 真挚的, 深厚的 0 | 与亲密, 喜爱, 可爱, 爱相关<br>Related to closed or love      | 珍贵美好的 Only about precious                                                     |
| fawn/fawns | n.   | 小鹿; 幼鹿            | A young deer                                                                                                                              | 鹿 0 幼崽 0 小动物 0 小崽 0                                                | 同时含有“鹿”和“小, 幼小”两个含义<br>Include both deer and young | 只有“鹿”或者只有“小”的含义<br>Only include deer                                          |

|          |    |                                |                                                                                                                                                                                                |                                                              |                                                                         |                                                           |
|----------|----|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|-------------------------------------------------------------------------|-----------------------------------------------------------|
|          |    |                                |                                                                                                                                                                                                |                                                              |                                                                         | or young                                                  |
| foal     | n. | 小马驹                            | A young horse                                                                                                                                                                                  | 一种生物 0 小鹿 0 幼崽 0<br>小动物 0                                    | 同时含有“马”和<br>“小，幼小”两个含义<br>Include both horse and<br>young               | 只有“马”或者<br>只有“小”的含<br>义<br>Only include<br>horse or young |
| foraging | v. | 觅食；搜寻                          | To go from place to place<br>searching for things that you can<br>eat or use                                                                                                                   | 捕食 0 捕获 0 养育 0                                               | 含有寻找，搜寻的动作<br>Searching                                                 | 捕，捕食的动作<br>Prey and actions<br>related                    |
| gland    | n. | 腺体                             | An organ of the body or of a plant<br>that secretes (= produces) liquid<br>chemicals that have various<br>purposes                                                                             | 腺 1 胰腺 1<br>胆 0                                              | 含有“腺”<br>Including 腺                                                    | 其他身体器官<br>Other parts of<br>body                          |
| hump     | v. | 弓起；<br>隆起，凸起；<br>驼峰；（人的）<br>驼背 | <u>As a noun:</u><br>A large, round raised area or part;<br>A round raised part on a person's<br>or animal's back;<br><u>as a verb:</u><br>To carry or lift something heavy<br>with difficulty | 山包，小包 1 小山丘 1 肿<br>块，胞 1 肿块 1 峰 1 弓身 1<br>弓 1<br><br>跳 0 背 0 | 与隆起的，凸起的形状，<br>或者弓，弓起相关<br>Anything related to the<br>round raised area | 跳或背部<br>Jump or the back                                  |
| midwife  | n. | 助产士，接生<br>员；产婆                 | a person, usually a woman, who is<br>trained to help women when they<br>are giving birth                                                                                                       | 助产的 1 接生 1<br><br>生育 0 护理人员 0 情妇，                            | 含有助产，接生的意思<br>Include helping to give<br>birth                          | 只有照顾，生养<br>的含义<br>Only about                              |

|           |    |                                         |                                                                                                                                                                                                                                                                |                                                |                                                                                                                                                          |                                                   |
|-----------|----|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|
|           |    |                                         |                                                                                                                                                                                                                                                                | 保姆 0 保姆 0                                      |                                                                                                                                                          | nursing                                           |
| nuzzle    | v. | 用鼻子蹭；<br>（尤指用头<br>或鼻）轻触，<br>轻擦，磨擦       | To touch, rub, or press something<br>or someone gently and/or in a way<br>that shows your love, especially<br>with the head or nose, usually with<br>small repeated movements                                                                                  | 蹭 1 蹭鼻子 1 碰鼻子 1<br><br>鼻腔的 0 闻 0 鼻 0           | 蹭，碰<br>Including 蹭，碰                                                                                                                                     | 只提到与鼻子有<br>关<br>Only related to<br>nose           |
| poaching  | v. | 偷猎；偷捕；<br>炖，水煮（鱼、<br>荷包蛋等）              | To catch and kill animals without<br>permission on someone else's land;<br>To cook something such as a fish,<br>or an egg with its shell removed,<br>by putting it in gently boiling<br>water or other liquid                                                  | 捕捉，捉到 0 捕猎 0 打猎 0<br>追捕 0                      | 含有“偷”，不被允许<br>的行为<br>Including action without<br>permission                                                                                              | 只含有捕，捕捉，<br>打猎，狩猎的含<br>义<br>Only about<br>capture |
| purring   | v. | 发出咕噜声                                   | (Of a cat) To make a soft, low,<br>continuous sound, or (of a<br>machine) to make a similar sound                                                                                                                                                              | 小猫叫 1 喵喵叫 1 打呼噜 1<br>呼噜 1 猫发出叫声 1              | 和猫的叫声或呼噜声有<br>关<br>The sounds of cat                                                                                                                     |                                                   |
| sanctuary | n. | 保护区；庇<br>护，保护；避<br>难所，庇护<br>所；<br>圣殿，圣堂 | Protection or a safe place,<br>especially for someone or<br>something being chased or hunted;<br>a place where birds or animals can<br>live and be protected, especially<br>from being hunted or dangerous<br>conditions;<br>the most holy part of a religious | 圣所 1 避难 1 保护基地 1<br>救所 1<br><br>安全 0 守卫 0 监视 0 | 与保护所，救助所，圣<br>所圣殿有关；或含有保<br>护，避难的含义<br>Related to shelters,<br>rescue shelters, sanctuary<br>temples; or contain the<br>meaning of protection,<br>refuge | 与监视有关<br>Related to<br>monitoring                 |

|             |      |                |                                                                                                                                                         |                                                                  |                                                                                       |                                                              |
|-------------|------|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------------------------|--------------------------------------------------------------|
|             |      |                | building                                                                                                                                                |                                                                  |                                                                                       |                                                              |
| sedated     | v.   | 安静的；平静的；被注射镇静剂 | Avoiding excitement or great activity and usually calm and relaxed;<br>to cause a person or animal to be very calm or go to sleep by giving them a drug | 静置下来的 1 镇静 1 静止的 1 冷静的 1<br><br>隔离 0 分离的 0 久坐，和坐相关 0 沉淀的 0 呆滞的 0 | 含有“静”<br>Including 静, quiet and calm                                                  | 与坐相关的，呆滞的<br>Related to sitting and slow                     |
| surrogate   | n.   | 替代；代理          | Replacing someone else or used instead of something else                                                                                                | 代替者 1<br><br>养母，养父母，领养 0                                         | 含代替，代理的含义<br>Related to substitution, and replacement                                 | 只有领养的含义<br>Only about adoption                               |
| traumatised | adj. | 受创伤的           | To shock and upset someone severely and for a long time                                                                                                 | 和灾难创伤相关 1 创伤 1 深受打击的 1<br><br>Trauma 0 毁坏 0 糟糕的 0 毁灭的 0          | 与受伤，受打击有关，显示出主体的受伤<br>Related to injuries, showing the main agency's injury and upset | 与坏，灾难，毁灭有关<br>Only related to bad, disaster, and destruction |
| twirls      | n.   | 缠绕；（使）旋转；（使）转动 | To (cause to) give a sudden quick turn or set of turns in a circle                                                                                      | 纠缠的 1 扭曲 1 旋梯 1 卷曲 1 扭 1 脖子卷在一起 1<br><br>摇摆 0 纠结 0 争扯 0          | 含旋，扭，缠，绕，卷，弯，搅等含义<br>Relate to turning in a circle                                    |                                                              |
| ulcers      | n.   | 溃疡             | A break in the skin, or on the surface of an organ inside the body, that does not heal naturally                                                        | 一种疾病 0 痤疮 0 肾上腺素 0                                               | 指出是溃疡<br>Mentioned ulcers                                                             | 其他疾病<br>Other disease                                        |

|         |    |             |                                                                                                                                    |                                                                   |                                                                     |                                                        |
|---------|----|-------------|------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------|--------------------------------------------------------|
| waddled | v. | 摇摇摆摆地走;蹒跚而行 | (Usually of a person or animal with short legs and a fat body)To walk with short steps, moving the body from one side to the other | 跋涉 1 摇摆的 1 蹒跚学步 1<br>走路不稳的 1 摇摆 1<br><br>蹠的 0 漫步 0 和走相关 0<br>散步 0 | 与蹒跚, 走路不稳, 摇摆有关<br>Related to staggering, unsteady walking, swaying | 只和走, 走路, 漫步有关<br>Only related to walking and strolling |
|---------|----|-------------|------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------|--------------------------------------------------------|

**Appendix S11. Mixed-Effects Models for Four Eye-Movement Measures for Level 1 Overall Subtitling area (Level 1)**

**Total Reading Time %**

| Fixed effects                                                                        | <i>b</i>  | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> |
|--------------------------------------------------------------------------------------|-----------|----------------|-----------|----------|----------|
| Intercept                                                                            | 1.47      | [0.94, 1.99]   | 0.27      | 5.42     | <.001    |
| Captions                                                                             | 0.03      | [-0.01, 0.07]  | 0.02      | 1.11     | .27      |
| L1 subtitles                                                                         | -0.11     | [-0.15, -0.07] | 0.02      | -4.46    | <.001    |
| No subtitles                                                                         | -0.42     | [-0.46, -0.38] | 0.02      | -17.69   | <.001    |
| log.Vsize                                                                            | -0.12     | [-0.18, -0.06] | 0.03      | -3.78    | <.001    |
| Random effects                                                                       |           | Variance       | <i>SD</i> |          |          |
| by IP                                                                                | Intercept | 0.003          | 0.06      |          |          |
| by participant                                                                       | Intercept | 0.007          | 0.08      |          |          |
| residual                                                                             |           | 0.02           | 0.15      |          |          |
| Best model: log.total.time.percentage ~ Group + log.Vsize + (1 Participant) + (1 IP) |           |                |           |          |          |
| Marginal $R^2 = .48$ ; Conditional $R^2 = .65$                                       |           |                |           |          |          |

**Fixation %**

| Fixed effects  | <i>b</i>  | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i>        |
|----------------|-----------|----------------|-----------|----------|-----------------|
| Intercept      | 1.31      | [0.82, 1.80]   | 0.25      | 5.21     | <b>&lt;.001</b> |
| Captions       | 0.03      | [-0.01, 0.07]  | 0.02      | 1.52     | .13             |
| L1 subtitles   | -0.08     | [-0.12, -0.04] | 0.02      | -3.74    | <b>&lt;.001</b> |
| No subtitles   | -0.44     | [-0.48, -0.40] | 0.02      | -20.01   | <b>&lt;.001</b> |
| log.Vsize      | -0.10     | [-0.16, -0.04] | 0.03      | -3.35    | <b>.001</b>     |
| Random effects |           | Variance       | <i>SD</i> |          |                 |
| by IP          | Intercept | 0.003          | 0.06      |          |                 |
| by participant | Intercept | 0.006          | 0.08      |          |                 |
| residual       |           | 0.02           | 0.14      |          |                 |

Best model: log.fixation.percentage ~ Group + log.Vsize + (1|Participant) + (1|IP)

Marginal  $R^2 = .54$ ; Conditional  $R^2 = .68$

### Run Count

| Fixed effects  | <i>b</i>  | 95% CI         | <i>SE</i> | <i>z</i> | <i>p</i> |
|----------------|-----------|----------------|-----------|----------|----------|
| Intercept      | 0.17      | [0.05, 0.29]   | 0.06      | 2.96     | <.001    |
| Captions       | -0.06     | [-0.22, 0.10]  | 0.08      | -0.73    | .47      |
| L1 subtitles   | -0.01     | [-0.17, 0.15]  | 0.08      | -0.10    | .92      |
| No subtitles   | -2.30     | [-2.48, -2.12] | 0.09      | -26.83   | <.001    |
| Random effects |           | Variance       | <i>SD</i> |          |          |
| by IP          | Intercept | 0.07           | 0.27      |          |          |
| by participant | Intercept | 0.09           | 0.29      |          |          |

Best model: run.count ~ Group + (1|Participant) + (1|IP)

Marginal  $R^2 = .40$ ; Conditional  $R^2 = .46$

### Skip Rate

| Fixed effects  | <i>b</i>  | 95% CI         | <i>SE</i> | <i>z</i> | <i>p</i> |
|----------------|-----------|----------------|-----------|----------|----------|
| Intercept      | -4.18     | [-4.85, -3.51] | 0.34      | -12.23   | <.001    |
| Captions       | -0.57     | [-1.26, 0.12]  | 0.35      | -1.62    | .12      |
| L1 subtitles   | 0.34      | [-0.37, 1.05]  | 0.36      | 0.95     | .34      |
| No subtitles   | 6.26      | [5.57, 6.95]   | 0.35      | 17.82    | <.001    |
| res.Vsize      | 1.50      | [0.44, 2.56]   | 0.54      | 2.78     | .01      |
| Random effects |           | Variance       | <i>SD</i> |          |          |
| by IP          | Intercept | 1.59           | 1.26      |          |          |
| by participant | Intercept | 1.56           | 1.25      |          |          |

Best model: skip ~ Group + res.Vsize + (1 | Participant) + (1 | IP)

Hosmer and Lemeshow's  $R^2 = .07$

**Appendix S12. Mixed-Effects Models for Four Eye-Movement Measures for Subtitling area Within the Bilingual Subtitles Group (Level 2)**

**Total Reading Time %**

| Fixed effects                                                                  | <i>b</i>     | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------------------------------------------------------------|--------------|----------------|-----------|----------|----------|----------|
| Intercept                                                                      | 0.33         | [0.31, 0.35]   | 0.01      | 31.79    | <.001    | 0.82     |
| Bilingual L2                                                                   | -0.16        | [-0.17, -0.15] | 0.005     | -33.45   | <.001    |          |
| Random effects                                                                 |              | Variance       | <i>SD</i> |          |          |          |
| by IP                                                                          | Intercept    | 0.006          | 0.07      |          |          |          |
|                                                                                | Bilingual L2 | 0.01           | 0.10      |          |          |          |
| by participant                                                                 | Intercept    | 0.003          | 0.05      |          |          |          |
| residual                                                                       |              | 0.03           | 0.17      |          |          |          |
| Best model: log.total.time.percentage ~ Group + (1 Participant) + (1+Group IP) |              |                |           |          |          |          |
| Marginal $R^2 = .15$ ; Conditional $R^2 = .32$                                 |              |                |           |          |          |          |

**Fixation %**

| Fixed effects                                                                | <i>b</i>     | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|------------------------------------------------------------------------------|--------------|----------------|-----------|----------|----------|----------|
| Intercept                                                                    | 0.35         | [0.33, 0.37]   | 0.01      | 35.16    | <.001    | 0.87     |
| Bilingual L2                                                                 | -0.17        | [-0.18, -0.16] | 0.005     | -34.47   | <.001    |          |
| Random effects                                                               |              | Variance       | <i>SD</i> |          |          |          |
| by IP                                                                        | Intercept    | 0.005          | 0.07      |          |          |          |
|                                                                              | Bilingual L2 | 0.01           | 0.10      |          |          |          |
| by participant                                                               | Intercept    | 0.002          | 0.05      |          |          |          |
| residual                                                                     |              | 0.03           | 0.17      |          |          |          |
| Best model: log.fixation.percentage ~ Group + (1 Participant) + (1+Group IP) |              |                |           |          |          |          |
| Marginal $R^2 = .17$ ; Conditional $R^2 = .32$                               |              |                |           |          |          |          |

### Run Count

| Fixed effects                                                  | <i>b</i>     | <i>SE</i> | <i>z</i> | <i>p</i>  | OR   | 95% CI       |
|----------------------------------------------------------------|--------------|-----------|----------|-----------|------|--------------|
| Intercept                                                      | 0.18         | 0.04      | 4.16     | <.001     | 1.20 | [1.10, 1.30] |
| Bilingual L2                                                   | -0.52        | 0.01      | -37.36   | <.001     | 0.59 | [0.58, 0.61] |
| Random effects                                                 |              | Variance  |          | <i>SD</i> |      |              |
| by IP                                                          | Intercept    | 0.08      |          | 0.28      |      |              |
|                                                                | Bilingual L2 | 0.02      |          | 0.15      |      |              |
| by participant                                                 | Intercept    | 0.05      |          | 0.21      |      |              |
| Best model: run.count ~ Group + (1 Participant) + (1+Group IP) |              |           |          |           |      |              |
| Marginal $R^2 = .05$ ; Conditional $R^2 = .13$                 |              |           |          |           |      |              |

### Skip Rate

| Fixed effects                                             | <i>b</i>     | <i>SE</i> | <i>z</i>  | <i>p</i> | OR   | 95% CI        |
|-----------------------------------------------------------|--------------|-----------|-----------|----------|------|---------------|
| Intercept                                                 | -2.63        | 0.16      | -16.71    | <.001    | 0.07 | [0.05, 0.10]  |
| Bilingual L2                                              | 2.27         | 0.05      | 42.81     | <.001    | 9.69 | [8.73, 10.75] |
| Random effects                                            |              | Variance  | <i>SD</i> |          |      |               |
| by IP                                                     | Intercept    | 0.62      | 0.79      |          |      |               |
|                                                           | Bilingual L2 | 0.77      | 0.88      |          |      |               |
| by participant                                            | Intercept    | 0.69      | 0.83      |          |      |               |
| Best model: skip ~ Group + (1 Participant) + (1+Group IP) |              |           |           |          |      |               |
| Hosmer and Lemeshow's $R^2 = .06$                         |              |           |           |          |      |               |

**Appendix S13. Mixed-Effects Models for Four Eye-Movement Measures for Subtitling area Between the Bilingual Subtitles and the Captions Groups (Level 2)**

**Total Reading Time %**

| Fixed effects                                                                        | <i>b</i>  | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------------------------------------------------------------------|-----------|----------------|-----------|----------|----------|----------|
| Intercept                                                                            | 0.17      | [0.09, 0.25]   | 0.04      | 4.43     | <.001    | 1.24     |
| Captions                                                                             | 0.39      | [0.27, 0.51]   | 0.06      | 6.98     | <.001    |          |
| log.Vsize                                                                            | -0.002    | [-0.14, 0.14]  | 0.07      | -0.03    | .98      |          |
| GroupA:log.Vsize                                                                     | -0.29     | [-0.49, -0.09] | 0.10      | -2.78    | .01      |          |
| Random effects                                                                       |           | Variance       | <i>SD</i> |          |          |          |
| by IP                                                                                | Intercept | 0.003          | 0.05      |          |          |          |
| by participant                                                                       | Intercept | 0.01           | 0.10      |          |          |          |
| residual                                                                             |           | 0.03           | 0.17      |          |          |          |
| Best model: log.dwell.time.percentage ~ Group * log.Vsize + (1 Participant) + (1 IP) |           |                |           |          |          |          |
| Marginal $R^2 = .32$ ; Conditional $R^2 = .52$                                       |           |                |           |          |          |          |

**Fixation %**

| Fixed effects    | <i>b</i>  | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|------------------|-----------|----------------|-----------|----------|-----------------|----------|
| Intercept        | 0.18      | [0.10, 0.26]   | 0.04      | 4.76     | <b>&lt;.001</b> | 1.36     |
| Captions         | 0.39      | [0.29, 0.49]   | 0.05      | 7.15     | <b>&lt;.001</b> |          |
| log.Vsize        | 0.004     | [-0.13, 0.14]  | 0.07      | 0.06     | .95             |          |
| GroupA:log.Vsize | -0.26     | [-0.46, -0.06] | 0.10      | -2.58    | <b>.01</b>      |          |
| Random effects   | Variance  |                | <i>SD</i> |          |                 |          |
| by IP            | Intercept | 0.003          | 0.06      |          |                 |          |
|                  | Captions  | 0.005          | 0.07      |          |                 |          |
| by participant   | Intercept | 0.01           | 0.09      |          |                 |          |
| residual         |           | 0.03           | 0.16      |          |                 |          |

Best model:  $\log.\text{fixation.percentage} \sim \text{Group} * \log.\text{Vsize} + (1|\text{Participant}) + (1+\text{Group}|\text{IP})$

Marginal  $R^2 = .34$ ; Conditional  $R^2 = .55$

### Run Count

| Fixed effects  | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i> | OR   | 95% CI       |
|----------------|-----------|-----------|-----------|----------|------|--------------|
| Intercept      | -0.39     | 0.06      | -6.25     | <.001    | 0.67 | [0.60, 0.76] |
| Captions       | 0.50      | 0.09      | 5.58      | <.001    | 1.65 | [1.38, 1.96] |
| Random effects |           | Variance  | <i>SD</i> |          |      |              |
| by IP          | Intercept | 0.10      | 0.32      |          |      |              |
|                | Captions  | 0.01      | 0.12      |          |      |              |
| by participant | Intercept | 0.10      | 0.32      |          |      |              |

Best model:  $\text{run.count} \sim \text{Group} + (1|\text{Participant}) + (1+\text{Group}|\text{IP})$

Marginal  $R^2 = .05$ ; Conditional  $R^2 = .20$

### Skip Rate

| Fixed effects    | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i> | OR    | 95% CI         |
|------------------|-----------|-----------|-----------|----------|-------|----------------|
| Intercept        | -0.29     | 0.46      | -0.64     | .52      | 0.75  | [0.30, 1.84]   |
| Captions         | -4.70     | 0.48      | -6.97     | <.001    | 0.01  | [0.003, 0.02]  |
| res.Vsize        | -0.20     | 0.84      | -0.24     | .81      | 0.82  | [0.16, 4.25]   |
| GroupA:res.Vsize | 3.82      | 1.24      | 3.08      | .002     | 45.54 | [4.01, 518.22] |
| Random effects   |           | Variance  | <i>SD</i> |          |       |                |
| by IP            | Intercept | 0.48      | 0.69      |          |       |                |
|                  | Captions  | 0.75      | 0.87      |          |       |                |
| by participant   | Intercept | 1.31      | 1.15      |          |       |                |

Best model:  $\text{skip} \sim \text{Group} * \text{res.Vsize} + (1|\text{Participant}) + (1+\text{Group}|\text{IP})$

Hosmer and Lemeshow's  $R^2 = .07$

**Appendix S14. Mixed-Effects Models for Four Eye-Movement Measures for Subtitling area Between the Bilingual Subtitles and the L1 Subtitles Groups (Level 2)**

**Total Reading Time %**

| Fixed effects                                                                  | <i>b</i>     | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------------------------------------------------------------|--------------|----------------|-----------|----------|----------|----------|
| Intercept                                                                      | 0.33         | [0.29, 0.37]   | 0.02      | 20.83    | <.001    | 0.28     |
| L1 subtitles                                                                   | -0.06        | [-0.10, -0.02] | 0.02      | -2.65    | .01      |          |
| Random effects                                                                 |              | Variance       | <i>SD</i> |          |          |          |
| by IP                                                                          | Intercept    | 0.01           | 0.08      |          |          |          |
|                                                                                | L1 subtitles | <0.001         | 0.03      |          |          |          |
| by participant                                                                 | Intercept    | 0.007          | 0.08      |          |          |          |
| residual                                                                       |              | 0.02           | 0.15      |          |          |          |
| Best model: log.total.time.percentage ~ Group + (1 Participant) + (1+Group IP) |              |                |           |          |          |          |
| Marginal $R^2 = .02$ ; Conditional $R^2 = .32$                                 |              |                |           |          |          |          |

**Fixation %:**  $\chi^2(1) = 2.63, p = .10, R^2 < .001$

**Run Count**

| Fixed effects                                            | <i>b</i>  | <i>SE</i> | <i>z</i> | <i>p</i>        | OR   | 95% CI       |
|----------------------------------------------------------|-----------|-----------|----------|-----------------|------|--------------|
| Intercept                                                | 0.19      | 0.04      | 5.16     | <b>&lt;.001</b> | 1.20 | [1.12, 1.29] |
| L1 subtitles                                             | -0.13     | 0.05      | -2.61    | <b>.01</b>      | 0.88 | [0.79, 0.97] |
| Random effects                                           |           | Variance  |          | <i>SD</i>       |      |              |
| by IP                                                    | Intercept | 0.08      |          | 0.29            |      |              |
| by participant                                           | Intercept | 0.03      |          | 0.17            |      |              |
| Best model: run.count ~ Group + (1 Participant) + (1 IP) |           |           |          |                 |      |              |
| Marginal $R^2 = .05$ ; Conditional $R^2 = .08$           |           |           |          |                 |      |              |

### Skip Rate

| Fixed effects                                             | <i>b</i>     | <i>SE</i> | <i>z</i> | <i>p</i>  | OR   | 95% CI       |
|-----------------------------------------------------------|--------------|-----------|----------|-----------|------|--------------|
| Intercept                                                 | -2.79        | 0.20      | -13.60   | <.001     | 0.06 | [0.04, 0.09] |
| L1 subtitles                                              | 0.76         | 0.30      | 2.54     | .01       | 2.13 | [1.19, 3.81] |
| Random effects                                            |              | Variance  |          | <i>SD</i> |      |              |
| by IP                                                     | Intercept    | 0.86      |          | 0.93      |      |              |
|                                                           | L1 subtitles | 0.02      |          | 0.15      |      |              |
| by participant                                            | Intercept    | 1.08      |          | 1.04      |      |              |
| Best model: skip ~ Group + (1 Participant) + (1+Group IP) |              |           |          |           |      |              |
| Marginal <i>R</i> <sup>2</sup> = .06                      |              |           |          |           |      |              |

**Appendix S15. Mixed-Effects Models for Seven Eye-Movement Measures Within the Bilingual Subtitles Group (Level 3)**

**Total Reading Time**

| Fixed effects                                                                                    | <i>b</i>     | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------------------------------------------------------------------------------------------|--------------|----------------|-----------|----------|----------|----------|
| Intercept                                                                                        | -8.10        | [-16.21, 0.01] | 4.14      | -1.96    | .06      | 0.34     |
| Bilingual L2                                                                                     | -1.84        | [-2.45, -1.23] | 0.31      | -5.84    | <.001    |          |
| log.area                                                                                         | 1.12         | [0.30, 1.94]   | 0.42      | 2.69     | .01      |          |
| log.FoO                                                                                          | 2.09         | [1.09, 3.09]   | 0.51      | 4.09     | <.001    |          |
| Random effects                                                                                   |              | Variance       | <i>SD</i> |          |          |          |
| by participant                                                                                   | Intercept    | 0.88           | 0.94      |          |          |          |
| by item                                                                                          | Intercept    | 0.46           | 0.67      |          |          |          |
|                                                                                                  | Bilingual L2 | 1.70           | 1.30      |          |          |          |
| residual                                                                                         |              | 4.97           | 2.23      |          |          |          |
| Best model: log.total.time ~ Group + log.area + log.FoO + (1 Participant) + (1+Group IA_24LABEL) |              |                |           |          |          |          |
| Marginal $R^2 = .15$ ; Conditional $R^2 = .39$                                                   |              |                |           |          |          |          |

**1<sup>st</sup>-Pass Reading Time**

| Fixed effects  | <i>b</i>     | 95% CI          | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|----------------|--------------|-----------------|-----------|----------|-----------------|----------|
| Intercept      | -7.16        | [-14.12, -0.20] | 3.55      | -2.02    | <b>.05</b>      | 0.28     |
| Bilingual L2   | -1.69        | [-2.26, -1.12]  | 0.29      | -5.76    | <b>&lt;.001</b> |          |
| log.area       | 0.10         | [-0.61, 0.81]   | 0.36      | 2.80     | <b>.01</b>      |          |
| log.FoO        | 2.10         | [1.24, 2.96]    | 0.44      | 4.82     | <b>.01</b>      |          |
| Random effects |              | Variance        | <i>SD</i> |          |                 |          |
| by participant | Intercept    | 0.78            | 0.88      |          |                 |          |
| by item        | Intercept    | 0.27            | 0.52      |          |                 |          |
|                | Bilingual L2 | 1.49            | 1.22      |          |                 |          |

|          |      |      |
|----------|------|------|
| residual | 4.56 | 2.14 |
|----------|------|------|

---

Best model:  $\log.X1st.pass.time \sim \text{Group} + \log.area + \log.FoO + (1|\text{Participant}) + (1+\text{Group}|\text{IA\_24LABEL})$

Marginal  $R^2 = .15$ ; Conditional  $R^2 = .37$

---

### 1<sup>st</sup> Fixation Duration

| Fixed effects  | <i>b</i>     | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|----------------|--------------|----------------|-----------|----------|----------|----------|
| Intercept      | 2.50         | [1.76, 3.24]   | 0.38      | 6.55     | <.001    | 0.35     |
| Bilingual L2   | -1.40        | [-1.97, -0.83] | 0.29      | -4.91    | <.001    |          |
| log.FoO        | 2.02         | [1.28, 2.76]   | 0.38      | 5.30     | <.001    |          |
| Random effects |              | Variance       | <i>SD</i> |          |          |          |
| by participant | Intercept    | 0.68           | 0.83      |          |          |          |
| by item        | Intercept    | 0.22           | 0.47      |          |          |          |
|                | Bilingual L2 | 1.56           | 1.25      |          |          |          |
| residual       |              | 4.14           | 2.03      |          |          |          |

---

Best model:  $\log.X1st.fixation.duration \sim \text{Group} + \log.FoO + (1|\text{Participant}) + (1+\text{Group}|\text{IA\_24LABEL})$

Marginal  $R^2 = .14$ ; Conditional  $R^2 = .36$

---

### 2<sup>nd</sup>-Pass Reading Time

| Fixed effects  | <i>b</i>  | 95% CI          | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|----------------|-----------|-----------------|-----------|----------|----------|----------|
| Intercept      | -15.04    | [-23.33, -6.75] | 4.23      | -3.56    | <.001    | 0.24     |
| Bilingual L2   | -1.01     | [-1.54, -0.48]  | 0.27      | -3.81    | <.001    |          |
| log.time       | 0.95      | [0.54, 1.36]    | 0.21      | 4.46     | <.001    |          |
| log.area       | 0.93      | [0.17, 1.69]    | 0.39      | 2.37     | .02      |          |
| Random effects |           | Variance        | <i>SD</i> |          |          |          |
| by participant | Intercept | 0.15            | 0.39      |          |          |          |

|          |              |      |      |
|----------|--------------|------|------|
| by item  | Intercept    | 0.79 | 0.89 |
|          | Bilingual L2 | 1.11 | 1.06 |
| residual |              | 4.15 | 2.04 |

---

Best model:  $\log.X2nd.pass.time \sim \text{Group} + \log.time + \log.area + (1|\text{Participant}) + (1+\text{Group}|\text{IA\_24LABEL})$

Marginal  $R^2 = .09$ ; Conditional  $R^2 = .22$

---

## 2<sup>nd</sup> Fixation Duration

| Fixed effects | <i>b</i> | 95% CI           | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|---------------|----------|------------------|-----------|----------|-----------------|----------|
| Intercept     | -22.04   | [-32.57, -11.51] | 5.37      | -4.11    | <b>&lt;.001</b> | 0.25     |
| Bilingual L2  | -1.22    | [-1.83, -0.61]   | 0.31      | -3.93    | <b>&lt;.001</b> |          |
| log.time      | 1.05     | [0.48, 1.62]     | 0.29      | 3.60     | <b>.002</b>     |          |
| log.area      | 1.63     | [0.69, 2.57]     | 0.48      | 3.39     | <b>.002</b>     |          |

---

| Random effects |              | Variance | <i>SD</i> |
|----------------|--------------|----------|-----------|
| by participant | Intercept    | 0.37     | 0.61      |
| by item        | Intercept    | 0.91     | 0.96      |
|                | Bilingual L2 | 1.59     | 1.26      |
| residual       |              | 4.67     | 2.16      |

---

Best model:  $\log.X2nd.fixation.duration \sim \text{Group} + \log.time + \log.area + (1|\text{Participant}) + (1+\text{Group}|\text{IA\_24LABEL})$

Marginal  $R^2 = .11$ ; Conditional  $R^2 = .30$

---

## Fixation count

| Fixed effects | <i>b</i> | <i>SE</i> | <i>z</i> | <i>p</i>        | OR     | 95% CI       |
|---------------|----------|-----------|----------|-----------------|--------|--------------|
| Intercept     | -10.85   | 2.37      | -4.58    | <b>&lt;.001</b> | <0.001 | [0.00, 0.00] |
| Bilingual L2  | -0.78    | 0.17      | -4.51    | <b>&lt;.001</b> | 0.46   | [0.33, 0.65] |
| log.time      | 0.46     | 0.18      | 2.54     | <b>.01</b>      | 1.58   | [1.11, 2.25] |

|                                                                                                             |           |          |      |       |      |              |
|-------------------------------------------------------------------------------------------------------------|-----------|----------|------|-------|------|--------------|
| log.area                                                                                                    | 0.71      | 0.19     | 3.75 | <.001 | 2.04 | [1.41, 2.97] |
| log.FoO                                                                                                     | 0.68      | 0.31     | 2.23 | .03   | 1.98 | [1.08, 3.60] |
| Random effects                                                                                              |           | Variance | SD   |       |      |              |
| by participant                                                                                              | Intercept | 0.20     | 0.45 |       |      |              |
| by item                                                                                                     | Intercept | 0.10     | 0.31 |       |      |              |
|                                                                                                             | Bilingual | 0.56     | 0.75 |       |      |              |
|                                                                                                             | L2        |          |      |       |      |              |
| Best model: fixation.count ~ Group + log.time + log.area + log.FoO + (1 Participant) + (1+Group IA_24LABEL) |           |          |      |       |      |              |
| Marginal $R^2 = .08$ ; Conditional $R^2 = .20$                                                              |           |          |      |       |      |              |

### Skip Rate

| Fixed effects                                                                | <i>b</i>     | <i>SE</i> | <i>z</i>  | <i>p</i>        | OR   | 95% CI        |
|------------------------------------------------------------------------------|--------------|-----------|-----------|-----------------|------|---------------|
| Intercept                                                                    | -1.12        | 0.35      | -3.17     | <b>.002</b>     | 0.33 | [0.16, 0.65]  |
| Bilingual L2                                                                 | 1.82         | 0.33      | 5.51      | <b>&lt;.001</b> | 6.17 | [3.23, 11.79] |
| res.time                                                                     | -2.41        | 0.76      | -3.19     | <b>.001</b>     | 0.09 | [0.02, 0.40]  |
| Random effects                                                               |              | Variance  | <i>SD</i> |                 |      |               |
| by participant                                                               | Intercept    | 1.07      | 1.04      |                 |      |               |
| by item                                                                      | Intercept    | 0.53      | 0.73      |                 |      |               |
|                                                                              | Bilingual L2 | 1.77      | 1.33      |                 |      |               |
| Best model: skip ~ Group + res.time + (1 Participant) + (1+Group IA_24LABEL) |              |           |           |                 |      |               |
| Hosmer and Lemeshow's $R^2 = .14$                                            |              |           |           |                 |      |               |

**Appendix S16. Mixed-Effects Models for Seven Eye-Movement Measures Between the Bilingual Subtitles and the Captions Groups (RQ3)**

**Total Reading Time**

| Fixed effects                                                                        | <i>b</i>  | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|--------------------------------------------------------------------------------------|-----------|----------------|-----------|----------|-----------------|----------|
| Intercept                                                                            | -9.55     | [-16.53, -2.6] | 3.56      | -2.68    | <b>.01</b>      | 0.75     |
| Captions                                                                             | 2.18      | [1.53, 2.83]   | 0.33      | 6.50     | <b>&lt;.001</b> |          |
| log.time                                                                             | 1.59      | [0.71, 2.47]   | 0.45      | 3.56     | <b>.001</b>     |          |
| Random effects                                                                       |           | Variance       | <i>SD</i> |          |                 |          |
| by participant                                                                       | Intercept | 0.90           | 0.95      |          |                 |          |
| by item                                                                              | Intercept | 1.83           | 1.35      |          |                 |          |
|                                                                                      | Captions  | 0.56           | 0.75      |          |                 |          |
| residual                                                                             |           | 4.48           | 2.12      |          |                 |          |
| Best model: log.total.time ~ Group + log.time + (1 Participant) + (1+Group IA_LABEL) |           |                |           |          |                 |          |
| Marginal $R^2 = .21$ ; Conditional $R^2 = .49$                                       |           |                |           |          |                 |          |

**1<sup>st</sup>-Pass Reading Time**

| Fixed effects                                                         | <i>b</i>  | 95% CI          | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|-----------------------------------------------------------------------|-----------|-----------------|-----------|----------|-----------------|----------|
| Intercept                                                             | -9.57     | [-15.67, -3.47] | 3.11      | -3.08    | <b>.01</b>      | 0.54     |
| Captions                                                              | 1.87      | [1.24, 2.50]    | 0.32      | 5.92     | <b>&lt;.001</b> |          |
| log.time                                                              | 1.57      | [0.81, 2.33]    | 0.39      | 4.05     | <b>&lt;.001</b> |          |
| Random effects                                                        |           | Variance        | <i>SD</i> |          |                 |          |
| by participant                                                        | Intercept | 0.72            | 0.85      |          |                 |          |
| by item                                                               | Intercept | 1.70            | 1.30      |          |                 |          |
|                                                                       | Captions  | 0.64            | 0.80      |          |                 |          |
| residual                                                              |           | 4.23            | 2.06      |          |                 |          |
| Best model: log.X1st.pass.time ~ Group + log.time + (1 Participant) + |           |                 |           |          |                 |          |

---

(1+Group|IA\_LABEL)

Marginal  $R^2 = .20$ ; Conditional  $R^2 = .46$

---

**1<sup>st</sup> Fixation Duration**

| Fixed effects  | <i>b</i>  | 95% CI          | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|----------------|-----------|-----------------|-----------|----------|-----------------|----------|
| Intercept      | -8.51     | [-14.10, -2.92] | 2.85      | -2.98    | <b>.007</b>     | 0.59     |
| Captions       | 1.75      | [1.18, 2.32]    | 0.29      | 5.95     | <b>&lt;.001</b> |          |
| log.time       | 1.42      | [0.71, 2.13]    | 0.36      | 3.98     | <b>.0006</b>    |          |
| Random effects |           | Variance        | <i>SD</i> |          |                 |          |
| by participant | Intercept | 0.63            | 0.79      |          |                 |          |
| by item        | Intercept | 1.40            | 1.18      |          |                 |          |
|                | Captions  | 0.52            | 0.72      |          |                 |          |
| residual       |           | 3.81            | 1.95      |          |                 |          |

Best model: log.X1st.fixation.duration ~ Group + log.time + (1|Participant) +

(1+Group|IA\_LABEL)

Marginal  $R^2 = .19$ ; Conditional  $R^2 = .44$

---

**2<sup>nd</sup>-Pass Reading Time**

| Fixed effects  | <i>b</i>  | 95% CI         | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|----------------|-----------|----------------|-----------|----------|-----------------|----------|
| Intercept      | -9.56     | [-23.01, 3.89] | 6.86      | -1.39    | .17             | 0.55     |
| Captions       | 1.71      | [1.18, 2.24]   | 0.27      | 6.27     | <b>&lt;.001</b> |          |
| log.time       | 0.57      | [0.06, 1.08]   | 0.26      | 2.24     | <b>.04</b>      |          |
| log.area       | 1.34      | [0.24, 2.44]   | 0.56      | 2.37     | <b>.03</b>      |          |
| log.Vsize      | -0.88     | [-1.68, -0.08] | 0.41      | -2.14    | <b>.04</b>      |          |
| Random effects |           | Variance       | <i>SD</i> |          |                 |          |
| by participant | Intercept | 0.43           | 0.65      |          |                 |          |
| by item        | Intercept | 0.24           | 0.49      |          |                 |          |

|          |          |      |      |
|----------|----------|------|------|
|          | Captions | 0.56 | 0.75 |
| residual |          | 4.29 | 2.07 |

---

Best model: log.X2nd.pass.time ~ Group + log.time + log.area + log.Vsize +  
(1|Participant) + (1+Group|IA\_LABEL)

Marginal  $R^2 = .15$ ; Conditional  $R^2 = .33$

---

## 2<sup>nd</sup> Fixation Duration

| Fixed effects  | <i>b</i>  | 95% CI           | <i>SE</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|----------------|-----------|------------------|-----------|----------|----------|----------|
| Intercept      | -31.44    | [-47.53, -15.35] | 8.21      | -3.83    | <.001    | 0.54     |
| Captions       | 1.81      | [1.20, 2.42]     | 0.31      | 5.75     | <.001    |          |
| log.time       | 1.21      | [0.52, 1.90]     | 0.35      | 3.41     | .003     |          |
| log.area       | 2.32      | [0.77, 3.87]     | 0.79      | 2.93     | .008     |          |
| Random effects |           | Variance         | <i>SD</i> |          |          |          |
| by participant | Intercept | 0.62             | 0.78      |          |          |          |
| by item        | Intercept | 1.01             | 1.01      |          |          |          |
|                | Captions  | 0.80             | 0.89      |          |          |          |
| residual       |           | 4.12             | 2.03      |          |          |          |

---

Best model: log.X2nd.fixation.duration ~ Group + log.time + log.area + (1|Participant)  
+ (1+Group|IA\_LABEL)

Marginal  $R^2 = .22$ ; Conditional  $R^2 = .44$

---

## Fixation count

| Fixed effects | <i>b</i> | <i>SE</i> | <i>z</i> | <i>p</i> | OR     | 95% CI       |
|---------------|----------|-----------|----------|----------|--------|--------------|
| Intercept     | -17.49   | 3.28      | -5.34    | <.001    | <0.001 | [0.00, 0.00] |
| Captions      | 1.00     | 0.16      | 6.10     | <.001    | 2.70   | [1.98, 3.82] |
| log.time      | 0.88     | 0.14      | 6.19     | <.001    | 2.41   | [1.81, 3.29] |
| log.area      | 1.01     | 0.32      | 3.20     | .001     | 2.76   | [1.43, 5.37] |

---

| Random effects                                                                                  |           | Variance | <i>SD</i> |
|-------------------------------------------------------------------------------------------------|-----------|----------|-----------|
| by participant                                                                                  | Intercept | 0.17     | 0.41      |
| by item                                                                                         | Intercept | 0.68     | 0.82      |
|                                                                                                 | Captions  | 0.21     | 0.46      |
| Best model: fixation.count ~ Group + log.time + log.area + (1 Participant) + (1+Group IA_LABEL) |           |          |           |
| Marginal $R^2 = .40$ ; Conditional $R^2 = .82$                                                  |           |          |           |

### Skip Rate

| Fixed effects                                                        | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>        | OR   | 95% CI        |
|----------------------------------------------------------------------|-----------|-----------|-----------|-----------------|------|---------------|
| Intercept                                                            | 0.85      | 0.49      | 1.72      | .09             | 2.33 | [0.89, 6.14]  |
| Captions                                                             | -2.48     | 0.38      | -6.61     | <b>&lt;.001</b> | 0.08 | [0.04, 0.17]  |
| res.time                                                             | -2.97     | 1.16      | -2.55     | <b>.01</b>      | 0.05 | [0.005, 0.50] |
| Random effects                                                       |           | Variance  | <i>SD</i> |                 |      |               |
| by participant                                                       | Intercept | 1.17      | 1.08      |                 |      |               |
| by item                                                              | Intercept | 1.54      | 1.24      |                 |      |               |
| Best model: skip ~ Group + res.time + (1 Participant) + (1 IA_LABEL) |           |           |           |                 |      |               |
| Hosmer and Lemeshow's $R^2 = .06$                                    |           |           |           |                 |      |               |

**Appendix S17. Mixed-Effects Models for Seven Eye-Movement Measures Between the Bilingual Subtitles and the L1 Subtitles Groups (RQ3)**

**Total Reading Time**

| Fixed effects                                                                             | <i>b</i> | 95% CI          | <i>SE</i> | <i>t</i> | <i>p</i>    | <i>d</i> |
|-------------------------------------------------------------------------------------------|----------|-----------------|-----------|----------|-------------|----------|
| Intercept                                                                                 | -17.52   | [-28.99, -6.05] | 5.85      | -2.99    | <b>.006</b> | 0.12     |
| L1 subtitles                                                                              | -0.63    | [-1.22, -0.04]  | 0.30      | -2.10    | <b>.04</b>  |          |
| log.time                                                                                  | 1.34     | [0.67, 2.01]    | 0.34      | 3.91     | <b>.001</b> |          |
| log.area                                                                                  | 1.18     | [0.22, 2.14]    | 0.49      | 2.42     | <b>.02</b>  |          |
| Random effects                                                                            | Variance |                 | <i>SD</i> |          |             |          |
| by participant (intercept)                                                                | 0.90     |                 | 0.95      |          |             |          |
| by item (intercept)                                                                       | 0.70     |                 | 0.84      |          |             |          |
| residual                                                                                  | 4.18     |                 | 2.04      |          |             |          |
| Best model: log.total.time ~ Group + log.time + log.area + (1 Participant) + (1 IA_LABEL) |          |                 |           |          |             |          |
| Marginal $R^2 = .11$ ; Conditional $R^2 = .36$                                            |          |                 |           |          |             |          |

**1<sup>st</sup>-Pass Reading Time**

| Fixed effects                                                                                      | <i>b</i> | 95% CI          | <i>SE</i> | <i>t</i> | <i>p</i>    | <i>d</i> |
|----------------------------------------------------------------------------------------------------|----------|-----------------|-----------|----------|-------------|----------|
| Intercept                                                                                          | -14.21   | [-25.34, -3.08] | 5.68      | -2.50    | <b>.02</b>  | 0.07     |
| L1 subtitles                                                                                       | -0.56    | [-1.11, -0.01]  | 0.28      | -1.99    | <b>.05</b>  |          |
| log.time                                                                                           | 1.09     | [0.44, 1.74]    | 0.33      | 3.27     | <b>.003</b> |          |
| log.area                                                                                           | 1.02     | [0.10, 1.94]    | 0.47      | 2.16     | <b>.04</b>  |          |
| Random effects                                                                                     | Variance |                 | <i>SD</i> |          |             |          |
| by participant (intercept)                                                                         | 0.76     |                 | 0.87      |          |             |          |
| by item (intercept)                                                                                | 0.66     |                 | 0.81      |          |             |          |
| residual                                                                                           | 3.97     |                 | 1.99      |          |             |          |
| Best model: log.X1st.run.total.time ~ Group + log.time + log.area + (1 Participant) + (1 IA_LABEL) |          |                 |           |          |             |          |

---

Marginal  $R^2 = .08$ ; Conditional  $R^2 = .32$

---

**1<sup>st</sup> fixation duration:**  $\chi^2(1) = 3.33, p = .07, R^2 = .001$

**2<sup>nd</sup>-Pass Reading Time**

| Fixed effects                                                                      | <i>b</i> | 95% CI          | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|------------------------------------------------------------------------------------|----------|-----------------|-----------|----------|-----------------|----------|
| Intercept                                                                          | -10.96   | [-15.96, -5.96] | 2.55      | -4.29    | <b>&lt;.001</b> | 0.17     |
| L1 subtitles                                                                       | -0.40    | [-0.79, -0.01]  | 0.20      | -2.01    | <b>.05</b>      |          |
| log.time                                                                           | 1.58     | [0.95, 2.21]    | 0.32      | 4.95     | <b>&lt;.001</b> |          |
| Random effects                                                                     | Variance |                 | <i>SD</i> |          |                 |          |
| by participant (intercept)                                                         | 0.26     |                 | 0.51      |          |                 |          |
| by item (intercept)                                                                | 0.61     |                 | 0.78      |          |                 |          |
| residual                                                                           | 4.26     |                 | 2.07      |          |                 |          |
| Best model: log.X2nd.pass.time ~ Group + log.time + (1 Participant) + (1 IA_LABEL) |          |                 |           |          |                 |          |
| Marginal $R^2 = .13$ ; Conditional $R^2 = .27$                                     |          |                 |           |          |                 |          |

---

**2<sup>nd</sup> Fixation Duration**

| Fixed effects       | <i>b</i> | 95% CI           | <i>SE</i> | <i>t</i> | <i>p</i>        | <i>d</i> |
|---------------------|----------|------------------|-----------|----------|-----------------|----------|
| Intercept           | -27.37   | [-39.84, -14.90] | 6.36      | -4.30    | <b>&lt;.001</b> | 0.15     |
| L1 subtitles        | -0.58    | [-1.03, -0.13]   | 0.23      | -2.50    | <b>.02</b>      |          |
| log.time            | 1.60     | [0.87, 2.33]     | 0.37      | 4.29     | <b>&lt;.001</b> |          |
| log.area            | 1.73     | [0.69, 2.77]     | 0.53      | 3.25     | <b>.003</b>     |          |
| Random effects      | Variance |                  | <i>SD</i> |          |                 |          |
| by participant      | 0.40     |                  | 0.63      |          |                 |          |
| (intercept)         |          |                  |           |          |                 |          |
| by item (intercept) | 0.84     |                  | 0.92      |          |                 |          |
| residual            | 4.58     |                  | 2.14      |          |                 |          |

---

Best model: log.X2nd.fixation.duration ~ Group + log.time + log.area + (1|Participant)  
+ (1|IA\_LABEL)

Marginal  $R^2 = .15$ ; Conditional  $R^2 = .33$

### Fixation count

| Fixed effects  | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>        | OR     | 95% CI          |
|----------------|-----------|-----------|-----------|-----------------|--------|-----------------|
| Intercept      | -12.65    | 2.55      | -4.96     | <b>&lt;.001</b> | <0.001 | [<0.001, 0.001] |
| L1 subtitles   | -0.21     | 0.11      | -1.94     | <b>.05</b>      | 0.81   | [0.65, 1.01]    |
| log.time       | 0.79      | 0.15      | 5.34      | <b>.01</b>      | 2.20   | [0.64, 2.96]    |
| log.area       | 0.69      | 0.21      | 3.21      | <b>.001</b>     | 1.99   | [1.32, 3.01]    |
| Random effects |           | Variance  | <i>SD</i> |                 |        |                 |
| by participant | Intercept | 0.11      | 0.32      |                 |        |                 |
| by item        | Intercept | 0.13      | 0.37      |                 |        |                 |

Best model: fixation.count ~ Group + log.time + log.area + (1|Participant) +  
(1|IA\_LABEL)

Marginal  $R^2 = .18$ , Conditional  $R^2 = .38$

### Skip Rate

| Fixed effects  | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>        | OR   | 95% CI       |
|----------------|-----------|-----------|-----------|-----------------|------|--------------|
| Intercept      | -1.09     | 0.34      | -3.21     | <b>.001</b>     | 0.34 | [0.17, 0.65] |
| L1 subtitles   | 0.68      | 0.33      | 2.07      | <b>.04</b>      | 1.97 | [1.03, 3.77] |
| res.time       | -2.62     | 0.76      | -3.43     | <b>&lt;.001</b> | 0.07 | [0.02, 0.32] |
| Random effects |           | Variance  | <i>SD</i> |                 |      |              |
| by participant | Intercept | 0.92      | 0.96      |                 |      |              |
| by item        | Intercept | 0.44      | 0.66      |                 |      |              |

Best model: skip ~ Group +res.time+ (1|Participant) + (1|IA\_LABEL)

Hosmer and Lemeshow's  $R^2 = .05$

**Appendix S18. Mixed-Effects Models with Significant Results for the Relationship Between Eye-Movement Measures on the L2 Unknown Target Words and Vocabulary Tests in the Bilingual Subtitles Group (RQ4)**

**Total Reading Time – Form Recognition**

| Fixed effects                                                                 | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>        | OR    | 95% CI        |
|-------------------------------------------------------------------------------|-----------|-----------|-----------|-----------------|-------|---------------|
| Intercept                                                                     | -1.79     | 0.33      | -5.49     | <b>.001</b>     | 0.17  | [0.08, 0.31]  |
| Total reading time                                                            | 1.10      | 0.32      | 3.45      | <b>.001</b>     | 3.01  | [1.63, 5.76]  |
| res.Vsize                                                                     | 2.73      | 0.61      | 4.52      | <b>&lt;.001</b> | 15.38 | [4.68, 56.23] |
| Random effects                                                                |           | Variance  | <i>SD</i> |                 |       |               |
| by participant                                                                | Intercept | 0.43      | 0.65      |                 |       |               |
| by item                                                                       | Intercept | 0.10      | 0.32      |                 |       |               |
| Best model: FR.Post ~ total.time + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |           |           |                 |       |               |
| Hosmer and Lemeshow's $R^2 = .53$                                             |           |           |           |                 |       |               |

**Total Reading Time – Meaning Recall**

| Fixed effects                                                                      | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>    | OR   | 95% CI        |
|------------------------------------------------------------------------------------|-----------|-----------|-----------|-------------|------|---------------|
| Intercept                                                                          | -3.21     | 0.45      | -7.12     | <b>.001</b> | 0.04 | [0.01, 0.09]  |
| Total reading time                                                                 | 1.13      | 0.39      | 2.86      | <b>.004</b> | 3.09 | [1.43, 6.89]  |
| res.Vsize                                                                          | 1.62      | 0.77      | 2.11      | <b>.04</b>  | 5.04 | [0.99, 24.26] |
| Random effects                                                                     |           | Variance  | <i>SD</i> |             |      |               |
| by participant                                                                     | Intercept | 0.51      | 0.72      |             |      |               |
| by item                                                                            | Intercept | 0.06      | 0.24      |             |      |               |
| Best model: Mrecall.Post ~ total.time + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |           |           |             |      |               |
| Hosmer and Lemeshow's $R^2 = .59$                                                  |           |           |           |             |      |               |

**Total Reading Time – Meaning Recognition:**  $\chi^2(1) = 2.46, p = .11, R^2 = .004$

**1<sup>st</sup>-Pass Reading Time – Form Recognition**

| Fixed effects                                                                       | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i> | OR    | 95% CI        |
|-------------------------------------------------------------------------------------|-----------|-----------|-----------|----------|-------|---------------|
| Intercept                                                                           | -1.80     | 0.32      | -5.6      | <.001    | 0.16  | [0.08, 0.31]  |
| 1st pass reading                                                                    | 1.70      | 0.43      | 3.95      | <.001    | 5.45  | [2.42, 13.14] |
| res.Vsize                                                                           | 2.65      | 0.60      | 4.39      | <.001    | 14.13 | [4.31, 51.41] |
| Random effects                                                                      |           | Variance  | <i>SD</i> |          |       |               |
| by participant                                                                      | Intercept | 0.43      | 0.65      |          |       |               |
| by item                                                                             | Intercept | 0.07      | 0.27      |          |       |               |
| Best model: FR.Post ~ 1st.pass.reading + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |           |           |          |       |               |
| Hosmer and Lemeshow's $R^2 = .55$                                                   |           |           |           |          |       |               |

**1<sup>st</sup>-Pass Reading Time – Meaning Recall**

| Fixed effects                                                                            | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>        | OR   | 95% CI        |
|------------------------------------------------------------------------------------------|-----------|-----------|-----------|-----------------|------|---------------|
| Intercept                                                                                | -3.10     | 0.44      | -7.05     | <b>&lt;.001</b> | 0.04 | [0.02, 0.10]  |
| 1st pass reading                                                                         | 1.22      | 0.49      | 2.47      | <b>.01</b>      | 3.38 | [1.29, 9.26]  |
| res.Vsize                                                                                | 1.51      | 0.77      | 1.98      | <b>.05</b>      | 4.54 | [0.89, 21.65] |
| Random effects                                                                           |           | Variance  | <i>SD</i> |                 |      |               |
| by participant                                                                           | Intercept | 0.51      | 0.72      |                 |      |               |
| by item                                                                                  | Intercept | 0.06      | 0.24      |                 |      |               |
| Best model: Mrecall.Post ~ 1st.pass.reading + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |           |           |                 |      |               |
| Hosmer and Lemeshow's $R^2 = .12$                                                        |           |           |           |                 |      |               |

**1<sup>st</sup>-Pass Reading Time – Meaning Recognition:**  $\chi^2(1) = 2.71, p = .10, R^2 = .004$

**2<sup>nd</sup>-Pass Reading Time – Form Recognition:**  $\chi^2(1) = 0.12, p = .73, R^2 < .001$

**2<sup>nd</sup>-Pass Reading Time – Meaning Recall:**  $\chi^2(1) = 0.82, p = .37, R^2 = .002$

**2<sup>nd</sup>-Pass Reading Time – Meaning Recognition:**  $\chi^2(1) = 0.52, p = .47, R^2 < .001$

**Appendix S19. Mixed-Effects Models with Significant Results for the Relationship Between Eye-Movement Measures on the L2 Unknown Target Words and Vocabulary Tests in the Captions Group (RQ4)**

**Total Reading Time – Form Recognition:**  $\chi^2(1) = 1.46, p = .23, R^2 = .01$

**Total Reading Time – Meaning Recall:**  $\chi^2(1) = 0.10, p = .75, R^2 < .001$

**Total Reading Time – Meaning Recognition:**

| Fixed effects                                                                    | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>        | OR    | 95% CI        |
|----------------------------------------------------------------------------------|-----------|-----------|-----------|-----------------|-------|---------------|
| Intercept                                                                        | -3.33     | 0.52      | -5.84     | <b>.001</b>     | 0.05  | [0.02, 0.13]  |
| Total reading time                                                               | 0.68      | 0.33      | 2.06      | <b>.04</b>      | 1.97  | [1.03, 3.83]  |
| res.Vsize                                                                        | 2.71      | 0.68      | 5.47      | <b>&lt;.001</b> | 15.03 | [3.96, 56.99] |
| Random effects                                                                   |           | Variance  | <i>SD</i> |                 |       |               |
| by participant                                                                   | Intercept | 0.29      | 0.54      |                 |       |               |
| by item                                                                          | Intercept | 1.23      | 1.11      |                 |       |               |
| Best model: MReco.Post ~ total.time + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |           |           |                 |       |               |
| Hosmer and Lemeshow's $R^2 = .05$                                                |           |           |           |                 |       |               |

**1<sup>st</sup>-Pass Reading Time – Form Recognition**

| Fixed effects                                                                       | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i>    | OR    | 95% CI        |
|-------------------------------------------------------------------------------------|-----------|-----------|-----------|-------------|-------|---------------|
| Intercept                                                                           | 0.96      | 0.45      | -2.13     | <b>.03</b>  | 0.39  | [0.15, 0.95]  |
| 1st pass reading                                                                    | 0.90      | 0.39      | 2.28      | <b>.02</b>  | 2.45  | [1.14, 5.47]  |
| res.Vsize                                                                           | 2.33      | 0.72      | 3.24      | <b>.001</b> | 10.91 | [2.46, 49.08] |
| Random effects                                                                      |           | Variance  | <i>SD</i> |             |       |               |
| by participant                                                                      | Intercept | 0.55      | 0.74      |             |       |               |
| by item                                                                             | Intercept | 0.60      | 0.78      |             |       |               |
| Best model: FR.Post ~ 1st.pass.reading + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |           |           |             |       |               |
| Hosmer and Lemeshow's $R^2 = .05$                                                   |           |           |           |             |       |               |

**1<sup>st</sup>-Pass Reading Time – Meaning Recall:**  $\chi^2(1) = 0.10, p = .75, R^2 < .001$

### 1<sup>st</sup>-Pass Reading Time – Meaning Recognition

| Fixed effects                                                                          | <i>b</i>  | <i>SE</i> | <i>z</i>  | <i>p</i> | OR    | 95% CI        |
|----------------------------------------------------------------------------------------|-----------|-----------|-----------|----------|-------|---------------|
| Intercept                                                                              | -2.87     | 0.49      | -5.85     | <.001    | 0.06  | [0.02, 0.14]  |
| 1st pass reading                                                                       | 0.79      | 0.41      | 1.96      | .05      | 2.21  | [0.99, 5.01]  |
| res.Vsize                                                                              | 2.58      | 0.69      | 5.23      | <.001    | 13.20 | [3.41, 51.03] |
| Random effects                                                                         |           | Variance  | <i>SD</i> |          |       |               |
| by participant                                                                         | Intercept | 0.34      | 0.58      |          |       |               |
| by item                                                                                | Intercept | 1.13      | 1.06      |          |       |               |
| Best model: MReco.Post ~ 1st.pass.reading + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |           |           |          |       |               |
| Hosmer and Lemeshow's $R^2 = .12$                                                      |           |           |           |          |       |               |

**2<sup>nd</sup>-Pass Reading Time – Form Recognition:**  $\chi^2(1) = 0.02, p = .89, R^2 < .001$

**2<sup>nd</sup>-Pass Reading Time – Meaning Recall:**  $\chi^2(1) = 0.19, p = .66, R^2 < .001$

**2<sup>nd</sup>-Pass Reading Time – Meaning Recognition:**  $\chi^2(1) = 0.001, p = .97, R^2 < .001$

## Appendix S20. Mixed-Effects Models for the Relationship Between Eye-Movement Measures on the L1 Translations of Unknown Target Words and Vocabulary Tests (RQ4)

### In the bilingual subtitles group:

**Total Reading Time – Form Recognition:**  $\chi^2(1) = 0.90, p = .34, R^2 = .001$

**Total Reading Time – Meaning Recall:**  $\chi^2(1) = 0.24, p = .63, R^2 < .001$

**Total Reading Time – Meaning Recognition:**  $\chi^2(1) = 0.10, p = .74, R^2 = .004$

**1<sup>st</sup>-Pass Reading Time – Form Recognition:**  $\chi^2(1) = 0.18, p = .67, R^2 < .001$

**1<sup>st</sup>-Pass Reading Time – Meaning Recall:**  $\chi^2(1) = 1.73, p = .19, R^2 = .005$

**1<sup>st</sup>-Pass Reading Time – Meaning Recognition:**  $\chi^2(1) = 2.66, p = .10, R^2 = .004$

**2<sup>nd</sup>-Pass Reading Time – Form Recognition:**  $\chi^2(1) = 1.63, p = .20, R^2 = .003$

**2<sup>nd</sup>-Pass Reading Time – Meaning Recall:**  $\chi^2(1) = 0.56, p = .45, R^2 = .002$

**2<sup>nd</sup>-Pass Reading Time – Meaning Recognition:**  $\chi^2(1) = 0.43, p = .51, R^2 < .001$

### In the L1 subtitles group:

**Total Reading Time – Form Recognition:**  $\chi^2(1) = 1.61, p = .20, R^2 = .003$

**Total Reading Time – Meaning Recall:**  $\chi^2(1) = 0.09, p = .77, R^2 < .001$

**Total Reading Time – Meaning Recognition:**  $\chi^2(1) = 0.33, p = .56, R^2 < .001$

**1<sup>st</sup>-Pass Reading Time – Form Recognition:**  $\chi^2(1) = 0.005, p = .94, R^2 < .001$

**1<sup>st</sup>-Pass Reading Time – Meaning Recall:**  $\chi^2(1) = 0.47, p = .49, R^2 = .002$

**1<sup>st</sup>-Pass Reading Time – Meaning Recognition:**  $\chi^2(1) = 0.03, p = .86, R^2 < .001$

### **2<sup>nd</sup>-Pass Reading Time – Form Recognition**

| Fixed effects    | <i>b</i> | <i>SE</i> | <i>z</i> | <i>p</i>   | OR   | 95% CI        |
|------------------|----------|-----------|----------|------------|------|---------------|
| Intercept        | -1.08    | 0.47      | -2.3     | <b>.03</b> | 0.34 | [0.12, 0.87]  |
| 2nd pass reading | -1.82    | 0.96      | 0.96     | .06        | 0.16 | [0.02, 0.98]  |
| res.Vsize        | 1.96     | 0.92      | 2.13     | <b>.03</b> | 7.12 | [1.13, 52.01] |

| Random effects                                                                      |           | Variance | <i>SD</i> |
|-------------------------------------------------------------------------------------|-----------|----------|-----------|
| by participant                                                                      | Intercept | 0.37     | 0.61      |
| by item                                                                             | Intercept | 0.79     | 0.89      |
| Best model: FR.Post ~ 2nd.pass.reading + res.Vsize + (1 Participant) + (1 IA_LABEL) |           |          |           |
| Hosmer and Lemeshow's $R^2 = .02$                                                   |           |          |           |

**2<sup>nd</sup>-Pass Reading Time – Meaning Recall:**  $\chi^2(1) = 0.07, p = .79, R^2 < .001$

**2<sup>nd</sup>-Pass Reading Time – Meaning Recognition:**  $\chi^2(1) = 0.93, p = .33, R^2 = .002$

