READING WITHOUT SPACES

BETWEEN WORDS:

EYE MOVEMENTS IN READING THAI

Benjawan Kasisopa

Thesis submitted for the degree of Doctor of Philosophy

MARCS Auditory Laboratories

University of Western Sydney

Revision submission January, 2011
STATEMENT OF AUTHENTICATION

I hereby declare that this submission is my own work and, to the best of my knowledge, it contains no material previously published or written by any other person, nor material which has been accepted for the award of any other degree at the University of Western Sydney, or any other educational institution, except where due acknowledgement is made in the thesis.
ACKNOWLEDGEMENTS

There are so many nice and generous people that I would like to thank.

First of all, I would like to thank my principal supervisor, Professor Denis Burnham, for all the opportunities, support and understanding he gave me over the years. I couldn’t imagine how I could complete this thesis without his great support as well as his patience. I really appreciate his knowledge and skills in many areas, psychology, statistical analysis and also his patience with my writing skills. Thank you so much for being such a great supervisor.

A million thanks to my co-supervisor, Professor Ronan Reilly, for all the technical help and all the hard work that he has put in for me, for his tolerance of my lack of technical skills, and for all his brilliant tools that made my life a lot easier in dealing with the complicated eye movement data and the analysis models. Thank you also for allowing me to take some time to work with him and his students at Department of Computer Science, National University of Ireland, Maynooth. I’m grateful that he was always there to help regardless of the 10-hour time difference between us! Also many thanks to all Professor Reilly’s PhD students for the great time during my stay in Maynooth, and especially to Mr. Siliang Tang, for creating so many great programs and kindly allowing me to use them.

I would also like to thank Dr. Sudaporn Luksaneeyanawin for allowing me to take over the Centre for Speech and Language Processing (CRSLP) lab at Chulalongkorn University, Bangkok to use as the hub for my data collection from time to time. I always had a good and productive time when visiting and working at
None of the studies in this thesis would have been possible without her support.

Thanks to Dr. Heather Winskel and Professor Ralph Radach who were on my supervisory panel and also to Professor Chris Davis, Associate Professor Jeessun Kim, Associate Professor Kate Stevens and Dr. Christian Kroos for all their help and advice.

I would also like to thank all the academic and support staff at MARCS Auditory Laboratories, especially Gail Charlton for all the administrative support as well as for her help regarding travel arrangements for field work and conferences; Darlene Williams and Karen McConachie for all the administrative and funding arrangements; Colin Schoknecht, Johnson Chen and Steven Fazio for all their technical support – I would have lost data without all your help; and last but not least, Kym Buckley, without whom I would not have been able to finish this thesis; thank you Kym for all your help in formatting the raw material into a presentable thesis.

I would like to thank all the students at MARCS; thank you so much for all the help. You have all made my life so far away from my home enjoyable and not as lonely as I expected, with special thanks to Nan Xu and Lidija Krebs-Lazendic for being more like family than friends.

Finally, I would like to thank my family – I could not have done this without their support – and to all my friends back home who helped find information and even assisted in recruiting participants for the studies.

Thank you so much to you all [ขอบคุณมาก!]
# TABLE OF CONTENTS

1. READING ACQUISITION, WRITING SYSTEM AND THAI LANGUAGE 6
   1.1. Reading Acquisition ................................................................. 8
   1.2. Writing Systems ................................................................. 13
       1.2.1. Ideographic Writing Systems ........................................ 13
       1.2.2. Alphabetic Writing Systems ............................................ 14
   1.3. Thai Language ........................................................................ 15
       1.3.1. Thai Script ................................................................. 16
       1.3.2. Thai Orthography .......................................................... 19
   1.4. Learning to Read in Thai ......................................................... 29
       1.4.1. Literacy Instruction in Thailand ....................................... 29
       1.4.2. Issues in Literacy Development and Reading .................... 31

2. EYE MOVEMENTS IN READING ......................................................... 33
   2.1. Introduction ........................................................................... 33
   2.2. Measurements of Eye Movements in Reading ......................... 34
   2.3. Approaches in Eye Movement Studies ..................................... 36
   2.4. Models of Eye Movement Control in Reading ......................... 39
       2.4.1. E-Z Reader Model .......................................................... 40
       2.4.2. Strategy-tactics/S-T ......................................................... 44
       2.4.3. Glenmore Model .............................................................. 46
   2.5. Eye Movements in Reading Spaced and Unspaced Texts ............ 49
       2.5.1. Reading With and Without Spaces in Spaced Languages .... 50
       2.5.2. Reading With and Without Spaces in Scriptio Continua Languages .. 57
   2.6. Eye Movements in Reading of Children .................................... 67
   2.7. Issues in Eye Movements in Reading Thai Text ......................... 69

3. EXPERIMENT 1: READING TIME STUDIES ........................................ 72
   3.1. Participants ............................................................................. 73
       3.1.1. Reading Ability Screening Test ...................................... 74
       3.1.2. Speech Production Screening Test .................................... 75
   3.2. Experiment 1: Relative Frequency of Initial and Final Characters Effects .. 75
       3.2.1. Materials ........................................................................ 76
6. CONCLUSION, DISCUSSION AND SUGGESTIONS FOR FUTURE STUDIES ................................................................. 179

   6.1. Conclusions for Experiments 1, 2 and 3 ................................................................. 179

       6.1.1. Experiment 1: Reading time studies ........................................................ 179

       6.1.2. Experiment 2: Eye Movements of Native Thai Adults Reading Silently
               and Aloud ........................................................................................................ 181

       6.1.3. Experiment 3: Eye movements of Thai children and adults reading
               silently 184

   6.2. Discussion ............................................................................................................ 186

   6.3. Suggestion for Future Studies ............................................................................ 190

7. REFERENCES ......................................................................................................... 194
LIST OF FIGURES

Figure 1. Diagram showing orthographic positions in Thai .............................................. 16
Figure 2. A schematic diagram of the E-Z Reader model (Reichle et al., 1998) ...... 42
Figure 3. A schematic diagram of strategy-tactics model (Reilly, 2010) .............. 45
Figure 4. An overview of Glenmore Model (Reilly et al., 2003a) ......................... 47
Figure 5. Probability of landing position of subject ME reading spaced vs. unspaced texts (from Epelboim et al., 1994) .......................................................... 51
Figure 6. Landing position of subjects when reading spaced and unspaced texts (Rayner et al., 1998) ........................................................................ 53
Figure 7. Landing position of subjects in spaced vs. unspaced vs. flanker conditions (Rayner et al., 1998) ........................................................................ 55
Figure 8. Examples of the Japanese word meaning ‘Japanese’ in three different writing systems: Kanji, Hiragana and Katakana (Kajii et al., 2001) ........ 58
Figure 9. Sample of stimulus configuration [a line of Japanese text flanked by two asterisks] (Kajii, et al. 2001) ....................................................... 58
Figure 10. Mean proportion of initial fixations (Kajii et al., 2001) ....................... 59
Figure 11. Sample of stimulus configuration (Sainio et al., 2007) ....................... 61
Figure 12. Mean proportion of initial fixations (Sainio et al., 2007) .................... 61
Figure 13. Example of Chinese stimuli from four spacing conditions (Bai et al., 2008) ......................................................................................... 62
Figure 14. Examples of Thai texts used in the experiment (Kohsom et al., 1997) .... 63
Figure 15. Excerpt of the Thai short story: ‘Deaf Sim’ used by Reilly et al. (2003) .. 64
Figure 16. PVL of Thai readers compared to German reader (Reilly et al., 2003) .... 65
Figure 17. Effects of final script frequency (Reilly et al., 2003) ......................... 66
Figure 18. Reading time extraction using PRAAT version 5.1.44 .......................... 81
Figure 19. Schematic representation of Hypotheses 3 and 4 ............................ 84
Figure 20. Reading time as a function of grade and number of syllables .......... 86
Figure 21. Reading time as a function of age group .............................................. 90
Figure 22. Reading time as a function of age group and reading ability ............ 91
Figure 23. Reading time as a function of age group and spacing ..................... 92
Figure 24. Reading time as a function of spacing and reading ability .............. 92
Figure 50. First fixation duration as a function of age and ability ..................... 157
Figure 51. Gaze duration as a function of reading ability and spacing ............... 159
Figure 52. Gaze duration as a function of start and end character frequencies .... 159
Figure 53. Gaze duration as a function of age group ........................................ 160
Figure 54. Gaze duration as a function of spacing and end character frequency .. 161
Figure 55. Landing site as a function of age group and first character frequency ... 164
Figure 56. Landing site as a function of age group and word length .................... 164
Figure 57. Landing site as a function of reading ability and first character frequency ................................................................................................................. 165
Figure 58. Landing site as a function of middle character frequency and spacing ... 165
Figure 59. Landing site as a function of middle character frequency and word length ................................................................................................................. 166
Figure 60. Landing site as a function of word length and age .............................. 167
Figure 61. Landing site as a function of reading ability and start character frequency ................................................................................................................. 168
Figure 62. First fixation duration as a function of word frequency and age ......... 169
Figure 63. Gaze duration as a function of age group and word frequency .......... 171
Figure 64. Gaze duration as a function of reading ability and word frequency ..... 171
Figure 65. Gaze duration as a function of spacing and age group ....................... 173
Figure 66. Gaze duration as a function of word frequency and age group .......... 174
LIST OF TABLES

Table 1. Hypothetical classification of participating languages relative to the dimensions of syllabic complexity (simple, complex) and orthographic depth (shallow to deep)........................................................................................................ 9
Table 2. Data for reading of familiar words and non-words by children learning different languages during their first year of instruction ......................... 10
Table 3. Initial consonant grapheme-phoneme correspondences .......................... 19
Table 4. Initial consonant classes ........................................................................ 21
Table 5. Paired consonants .................................................................................. 21
Table 6. Two syllable pseudo-consonant clusters ................................................. 22
Table 7. Consonant grapheme-phoneme correspondence changes in final position 23
Table 8. Final consonant grapheme-phoneme correspondences ............................ 24
Table 9. Short vowel grapheme-phoneme correspondences .................................. 26
Table 10. Long vowel grapheme-phoneme correspondences ................................ 26
Table 11. Diphthong grapheme-phoneme correspondences .................................. 26
Table 12. Vowel characters represented vowel plus final consonant ....................... 26
Table 13. Vowel characters change in closed syllable .......................................... 27
Table 14. Tone markers in Thai ........................................................................... 28
Table 15. Illustration of the spacing conditions in the Rayner et al. (1998) experiment .................................................................................................................. 54
Table 16. Frequency count of scripts at word initial and final positions ................. 65
Table 17. Relative frequency of characters in the initial position ............................ 77
Table 18. Relative frequency of characters in the final position .............................. 78
Table 19. Examples of sentences with high, medium or low frequency initial and final character using in Experiment 1 (Unspaced condition) .................. 79
Table 20. Examples of sentences with high, medium or low frequency initial and final character using in Experiment 1 (Spaced condition) ....................... 80
Table 21. Overall mean reading time of participants in Experiment 1 .................... 85
Table 22. Mean reading time of participants with different reading ability .......... 85
Table 23. Mean reading time of participants with different reading ability .......... 85
Table 24. Mixed – effect models results of participants’ reading time ................. 89
Table 25. Error types and mean numbers of errors of each participant group ....... 97
Table 26. Relative frequency of characters in the initial position ....................... 105
Table 27. Relative frequency of characters in the final position .......................... 106
Table 28. Examples of sentences with high, medium or low frequency start and end characters used in Experiment 2 (Unspaced condition) ...................... 108
Table 29. Examples of sentences with high, medium or low frequency start and end characters used in Experiment 2 (Spaced condition) ......................... 109
Table 30. Mean landing site location, first fixation duration and gaze duration of Analysis Set I ............................................................................................................. 116
Table 31. Mixed-effect models results of Analysis Set I on landing site location .. 117
Table 32. Mixed-effect models results of Analysis I on first fixation duration .... 122
Table 33. Mixed-effect Models results of Analysis I on Gaze Duration .......... 124
Table 34. Mixed-effect Models results of Analysis Set II on Landing Site Location ................................................................................................................. 126
Table 35. Mixed-effect Models results of Analysis Set II on First Fixation Duration ................................................................................................................. 127
Table 36. Mixed-effect Models results of Analysis Set II on Gaze Duration ...... 128
Table 37. Mean landing site location, first fixation duration and gaze duration of Analysis Set III ............................................................................................................. 129
Table 38. Mixed-effect models results of Analysis Set III on landing site location 130
Table 39. Mixed-effect models results of Analysis Set III on first fixation duration ................................................................................................................. 132
Table 40. Mixed-effect Models results of Analysis Set III on Gaze Duration .... 133
Table 41. Examples of sentences with high, medium or low frequency start and end characters used in this Experiment 3 (Unspaced condition) .............. 144
Table 42. Examples of sentences with high, medium or low frequency start and end characters used in this Experiment 3 (Spaced condition) ................. 144
Table 43. Mean landing site location, first fixation duration and gaze duration of participants in Experiment 3 ............................................................................................................. 148
Table 44. Mixed-effect models results of children participants’ landing site location on target word (Analysis Set IA) .......................................................................................... 150
Table 45. Mixed-effect models results of child and adult participants’ landing site location on target word (Analysis Set IB) .......................................................................................... 152
Table 46. Mixed-effect models results of child and adult participants’ landing site location on target word (Analysis Set IIA) .............................................. 153
Table 47. Mixed-effect models results of child and adult participants’ landing site location on target word (Analysis Set IIB) .............................................. 154
Table 48. Mixed-effects models results of child participants’ first fixation duration on target word (Analysis IA) ................................................................. 155
Table 49. Mixed-effects models results of child and adult participants’ first fixation duration on target word (Analysis IB) ..................................................... 155
Table 50. Mixed-effects models results of child and adult participants’ first fixation duration on target word (Analysis IIB) ..................................................... 157
Table 51. Mixed-effects models results of child participants’ gaze duration on target word ......................................................................................................... 158
Table 52. Mixed-effects models results of child and adult participants’ gaze duration on target word .......................................................................................... 158
Table 53. Mixed-effects models results of child and adult participants’ gaze duration on target word .......................................................................................... 158
Table 54. Mean landing site location, first fixation duration and gaze duration for words in sentences of participants in Experiment 3 ................................ 161
Table 55. Mixed-effect models results of child participants’ landing site location (sentences) ............................................................................................... 163
Table 56. Mixed-effect models results of child and adult participants’ landing site location .................................................................................................... 167
Table 57. Mixed-effect models results of child and adult participants’ on first fixation duration ....................................................................................... 169
Table 58. Mixed-effect models results of child and adult participants’ on gaze duration ................................................................................................. 170
Table 59. Mixed-effect models results of child and adult participants’ on gaze duration ................................................................................................. 172
ABSTRACT

Studies of eye movements in reading alphabetic writing system languages, such as English, suggest that the optimal viewing position (OVP), the most effective target in each word that allows fastest word processing, is the word centre. In alphabetic languages with spaces between the words, research has shown that readers’ preferred viewing location (PVL) is to the left of the OVP. It appears that spaces between words is the most salient low-level visual processing cue to guide the eyes during reading, for when the spaces are removed from the text of alphabetic spaced languages, reading rate decreases by approximately 35% and the PVL shifts dramatically from the word-centre towards the word-beginning. Although these finding are widely accepted, it is unclear how well such results generalize to languages without spaces between words – scriptio continua alphabetic languages.

Thai is a good model of a scriptio continua language in which to investigate eye movements during reading, not only because it is written without spaces between words but because its orthography is also quite complicated in terms of visual information. In brief, characters such as vowels, tones, and other diacritics or even some parts of the consonants can be written above or below the main horizontal line. Reilly et al. (2003) surprisingly found that Thai adult readers also target the word centre during saccadic eye movements, even though there is no spacing to help indicate word boundaries. They suggested that the relative position-specific frequency of occurrence of final and initial characters serve as visual cues to guide eye movements of Thai readers. Analysis of the frequency of initial and final characters calculated for the texts used in their experiment confirm their suggestion that
participants’ landing site location tended to be closer to the word centre if the position-specific frequency of the initial and especially final characters are relatively high.

The aim of this thesis was to test the effects of low-level visual distinctive features in Thai orthography – namely i) the relative frequency of occurrence of characters in the initial and final positions, and ii) spaces between words – on reading time, eye movement patterns and control, and fixation patterns of native Thai readers, both children and adults.

Experiment 1 involved studies of reading time with a group of adults and four groups of children (1st, 2nd, 5th and 6th Graders, with half in each group being good and half being poor readers). This experiment focussed particularly on the start and end characters of words. It was found that relative frequency of occurrence of characters in the word-start and word-end positions had significant effects on reading time and reading accuracy of Thai participants across all ages. Higher frequency characters, especially word-start characters helped reduce reading time, and spacing between words facilitated reading in general as indicated by shorter reading times, especially for poor reader young children. Differences due to groups and spacing decreased as the age of participants increased and their reading skills improved.

Experiments 2 and 3 involved precise measurements of eye movements using the EyeLink II apparatus and followed up on the effects found in Experiment 1, those of position-specific frequency of word-start and word-end characters. In Experiment 2, adults of lower and higher education levels were tested on unspaced and spaced text reading either silently or aloud. It was found that Thai readers’ PVL was at or near the word centre in all conditions. The presence or absence of spaces between words did
not cause any dramatic changes to this PVL, like those found in the eye movements of English language readers presented with the unusual unspaced text. That is, for Thai readers reading normal unspaced text and unusual spaced text, the oculomotor patterns were the same in contrast to the dramatic change in English readers’ eye movements when faced with the unusual unspaced text (Rayner et al., 1998). Nevertheless, in concert with the Experiment 1 reading time studies, spaces between words did allow faster reading in Thai readers; there were shorter first fixations and gaze durations for spaced than unspaced text. In addition, for reading aloud the PVL was closer to word start than for silent reading, and first fixation and gaze duration were longer for reading aloud. Relative frequency of characters especially at word-start position had significant effects on landing site location of Thai adults (skilled readers), i.e., higher start character frequency allowed participants to land their eyes at the PVL. Word-end character frequency had less effect on landing site but stronger effects on fixation time of the participants; both first fixation and gaze durations on words with higher end character frequency were shorter than those with the lower frequency.

In Experiment 3, with two groups of children the results were similar although the landing site of the younger child participants was a bit further to the left of the word centre than it was for older children and adults, but still too far into the word to be designated as the word-initial area. Unlike the results of the adults’ eye movements, spaces between words had significant effects on landing site location especially in younger children. Spaces facilitated young children’s oculomotor controls and assisted in lading their eyes closer to the OVP. There were no significant main effects of character frequency on landing site location of Thai children on the target words. However, frequency of characters in both word-start and word-end positions were used
by younger child readers if there were no visible visual cue, i.e., spaces, available. Thus younger children relied on low-level visual information such as spaces between words more when reading. This may be because literacy teaching in Thai starts with spaced texts therefore younger children were more familiar with spaced text and try to use this information to locate word boundaries first before moving to the next resource such as the characters at word-start and word-end position.

Generally the results of this thesis show that Thai readers, child and adult, use the same oculomotor controls when reading spaced and unspaced text. Adding spaces into the text does not change the PVL which remains near the OVP at or near word centre and similar to that of native readers of spaced alphabetic languages; however spacing does facilitate the landing site of younger children to be closer to the OVP. The developmental trend of eye movements in reading Thai seems to be that Thai readers rely less on visible visual cue as their reading skills increase at which point spaces between words facilitate reading time (first fixation and gaze duration) of skilled readers. Such results show that spaces between words are not essential for optimal eye movements in Thai. However, as spaces between words result in a decrease in reading time, spaces may aid in word recognition. These results have important implications for models of eye movements in reading which, at present, do not explicitly account for reading in scriptio continua languages.

Additionally, two more reading time studies were conducted to investigate other distinctive features of Thai orthography (and details are given in Appendix C). In Experiment A, the focus was the feedback consistency versus inconsistency of initial and final characters. Similar results to Experiment 1 were found in terms of spacing conditions of the texts. Feedback inconsistent grapheme-phoneme relationships
between letters and sounds slowed reading time for both initial and final consonants
but there were more errors for final consonants, possibly due to significantly fewer
final phonemes in the Thai phonological system. In Experiment B, the transparency of
tone realisations was investigated. It was found that participants read words with
transparent tone realisations faster and made fewer errors than for words with opaque
tone realisations. Together these results show that reading time is sensitive to the
influence of features of particular characters (frequency, grapheme-phoneme
consistency, transparency of tone realisation) and also to spacing between words. Even
though these features had some significant effects on reading time and reading
accuracy of Thai readers, we still did not know if they would have any effects on eye
movements of Thai readers or not.

The results of these two additional reading time experiments provide the bases
and hypotheses for further eye movement experiments; the effects found there need to
be followed up in further studies. On the other hand the results of the reading time
(Experiment 1) and eye movement (Experiment 2 and 3) studies of word-start and
word-end character frequencies provided definitive evidence that the PVL in Thai, a
scriptio continua language, is near the OVP at word centre as for spaces alphabetic
languages. This finding should be further investigated in other scriptio continua
languages such as Khmer (Cambodian) and Lao.
CHAPTER ONE

READING ACQUISITION, WRITING SYSTEM AND THAI LANGUAGE
1. READING ACQUISITION, WRITING SYSTEM AND THAI LANGUAGE

Reading is the process of gaining access to meaning from printed symbols (Ziegler and Goswami, 2005). Reading, unlike speaking, requires specialized tuition; children must be taught to acquire written language. Learning to read is a process of mapping written symbols (graphemes) to their sound representations (phonemes). In most languages, the relationship between written symbols and sound is systematic, but the relationship between written symbols and meanings is arbitrary (Ziegler & Goswami, 2006). The consistency of the orthography of a language, as well as the initial age at which children start learning to read are important factors in determining whether and how well children acquire reading skills. However, most children begin their formal schooling between the ages of four and seven years, depending on the policy of the country so there is not a large degree of variation in this factor. Of more importance is the nature of the orthography, and indeed the rate and success of reading acquisition has been found to vary with nature of the orthography being learned (Seymour, Aro & Erskine, 2003).

Eye movement behaviour during reading is an important moment-to-moment processing measurement by which the language comprehension process as well as general language processing and critical psycholinguistic issues in the normal reading process (Rayner, 1998; Starr & Rayner, 2001; Rayner & Pollatsek, 2006) can be determined. Most of these studies have investigated eye movements in reading languages that are written with spaces between the words, and these spaces between words have been generally accepted as the paramount visual cue in guiding eye movements in reading (Rayner, 1998). However, not all languages are written with
spaces between words and in such scriptio continua languages, especially ones with an alphabetic writing system, there is perhaps more limited information to guide eye movements. The aim of this thesis is to investigate eye movements in reading Thai, an alphabetic scriptio continua language, in order to investigate the factor(s) that serve as visual cues to guide eye movements when boundaries are and are not signaled by spaces between words.

This thesis is divided into seven chapters. In Chapter 1, background information about reading acquisition, writing systems of world languages and the Thai language are described in detail. The details of the measurements of eye movement study, approaches of eye movements in reading, models of eye movements in reading and experimental investigations of spaced versus scriptio continua languages and consideration of issues in eye movements in reading Thai, for example the effect of lack of spaces between words, are examined in Chapter 2. Chapter 3 is an empirical chapter in which the results of three reading time studies of Thai are presented with a view to determining the orthographic factors that should be used in the subsequent studies of eye movement in reading Thai text study. Chapter 4 provides a bridge between the reading times studies in Chapter 3 and the eye movement studies in Chapter 5 and 6. Chapter 5 is another empirical chapter concerned with adult good and poor readers’ eye movements in reading Thai text silently and aloud and Chapter 6 is also an eye movement study, the eye movements of child and adult good and poor readers when reading silently will be described and compared with the adult results in Chapter 5. Finally, Chapter 7 provides a general discussion of the results and conclusions of the studies of the thesis.
1.1. Reading Acquisition

One of the most important processes in reading is what is known as ‘phonological recoding’ (Share, 1995), the ability to map distinctive visual symbols to the units of sound they represent. Via phonological reading readers are able to read words that are present in their spoken lexicons and also, in alphabetic languages, and recode unknown words as well (Ziegler & Goswami, 2006). Although phonological recoding is an efficient strategy in learning to read, there can be some problems with graphemic (letter) and phonemic (sound) representations. In some languages, one letter can have more than one pronunciation, as in English, whereas in others it always pronounced the same, as in Italian or Spanish (Goswami, 2003). Similarly, in some orthographies, one phoneme can have multiple spellings, whereas in others it always spelled the same way. These variations across languages make it reasonable to expect that there will be differences in reading development across languages since it should be relatively easy to learn to read if one letter consistently represents one and the same phoneme, or if one phoneme always represents one letter whereas it should be more difficult to learn to read if either of these conditions does not hold (Ziegler et al., 2006).

Seymour, Aro, and Erskine (2003) studied literacy acquisition in European languages varying in orthographic complexity. They classified 13 European language orthographies on the basis of two dimensions; syllabic complexity and orthographic depth. Syllable complexity refers to the distinction between Romance languages, which have predominantly open (CV) syllables and few initial or final consonant clusters, and Germanic languages, which have numerous closed (CVC) syllables and complex initial and final consonant clusters. Orthographic depth contrasts alphabetic
writing systems on grapheme-phoneme correspondences (whether there are one-to-one (shallow) or one-to-many (deep) grapheme-phoneme correspondences) including multi-letter graphemes, context dependent rules, irregularities and morphological effects (Seymour et al., 2003). Their hypothetical classification of 13 European languages according to syllabic complexity and orthographic depth is shown in Table 1.

Table 1. **Hypothetical classification of participating languages relative to the dimensions of syllabic complexity (simple, complex) and orthographic depth (shallow to deep)**

<table>
<thead>
<tr>
<th>Orthographic Depth</th>
<th>Shallow</th>
<th>Deep</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syllable structure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple</td>
<td>Finnish</td>
<td>Greek</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Italian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spanish</td>
</tr>
<tr>
<td>Complex</td>
<td>German</td>
<td>Norwegian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(adapted from Seymour et al., 2003)

Children have been found to learn to read regular alphabetic orthographies such as Italian, Finnish and German at a fast rate of acquisition, attaining 90% accuracy in recoding grapheme-phoneme relationships within the first months of reading instruction; whereas children learning to read an irregular language, such as English, require approximately 3 to 4 years to reach the similar level of success (Seymour et al., 2003; Goswami, 2003). Comparative results are shown in Table 2 overleaf.
Table 2. Data for reading of familiar words and non-words by children learning different languages during their first year of instruction

<table>
<thead>
<tr>
<th>Language</th>
<th>Familiar Real Words (% correct)</th>
<th>Monosyllabic Non-words (% correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greek</td>
<td>98</td>
<td>97</td>
</tr>
<tr>
<td>Finnish</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>German</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Italian</td>
<td>95</td>
<td>92</td>
</tr>
<tr>
<td>Spanish</td>
<td>95</td>
<td>93</td>
</tr>
<tr>
<td>Swedish</td>
<td>95</td>
<td>91</td>
</tr>
<tr>
<td>Dutch</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Icelandic</td>
<td>94</td>
<td>91</td>
</tr>
<tr>
<td>Norwegian</td>
<td>92</td>
<td>93</td>
</tr>
<tr>
<td>French</td>
<td>79</td>
<td>88</td>
</tr>
<tr>
<td>Portuguese</td>
<td>73</td>
<td>76</td>
</tr>
<tr>
<td>Danish</td>
<td>71</td>
<td>63</td>
</tr>
<tr>
<td>English</td>
<td>34</td>
<td>41</td>
</tr>
</tbody>
</table>

(compiled from Seymour et al. 2003)

Studies of reading acquisition across European languages have found that at least three factors are crucial for explaining cross-language differences; consistency of grapheme-to-phoneme relationship, granularity (grain size) of orthographic and phonological representations, and teaching methods (Ziegler et al., 2005). As shown in Table 2, the reading accuracy of children learning to read languages with high consistency of grapheme-to-phoneme relationships, such as Greek and Italian is significantly higher than those with a lower consistency relationship such as Danish and English.

Psycholinguistic grain size theory suggests that the differences in reading accuracy and reading speed across orthographies reflect fundamental differences in the phonological recoding process and reading strategies that develop in response to the orthography. Children who learn to read more transparent grapheme-to-phoneme...
correspondence languages, e.g. Greek, have the advantage of using smaller grapheme units, like consonant or vowel sounds than children learning to read less orthographically consistent languages, such as English. The results of the studies have shown that children who learn to read a less transparent orthography language use larger grapheme units such as rimes (Goswami, 2005). However, recent developmental studies have found that children who learn to read less transparent orthography languages tends to develop variety of recoding strategies, combining grapheme-phoneme conversion strategies with recognition pattern of rimes that allow them to attempt whole-word recognition allowing them to have better reading accuracy and faster reading time (Goswami, 2005). However, it is important to be aware that inconsistency does not affect all psycholinguistic units to a similar extent. Smaller grain sizes tend to be more inconsistent than larger grain sizes (Treiman, Mullennix, Bijeljac-Basic & Richmond-Welty, 1995) for example, in English “i” can be pronounced [ɪ] as in ‘hit’ or [ai] as in ‘kind’, but “-iss” is always [ɪs]. The reduced reliability of small grain size in inconsistent orthographies may lead to development of recoding strategies at more than one grain size, which will take longer to develop than a single recoding strategy (Ziegler et al., 2005).

Brown and Deavers (1999) found that children learning to read English develop sublexical strategies at larger grain sizes according the rhyme of the word (rhyme analogy effects). The results of this study showed that both children and adults respond to irregular consistent non-word naming using both small and large grain sizes and there was a cross-over interaction between grain size effects and reading skills. They found a predominance of small grain size responses for less skilled readers and a predominance of large grain size for skilled readers, suggesting that reading
acquisition in English begins with mastery of small units with these becoming larger as reading skill improves. These findings support the theoretical framework of psycholinguistics gain size theory.

Moving to the last factor, teaching methods, there are two dominant approaches in teaching literacy, the whole-word method and the phonic method. For transparent orthography languages teaching literacy is quite naturally based on the phonics method. That is children learn to read starting from single letters. Children of these languages are taught letter-to-sound correspondences and acquire the phoneme and phonological systems through this and, as a result, children learning to read in these languages quickly become highly skill readers (Rayner, Foorman, Perfetti, Pesetsky and Siedenberg, 2001). On the other hand, children learning to read less transparent orthographies that use larger units, such as syllables or rimes, in teaching literacy may be more effective (Goswami, 2005). While there is some controversy over this issue and some indications that a phonics method of teaching literacy is better irrespective of the relative transparency of the language, it is believed that phonic methods are better for the relatively transparent orthography languages such as Italian or Greek while the whole-word methods are more suitable for the less transparent ones (Ziegler et al., 2005).

In conclusion, reading acquisition varies from one language to another depending on the consistency of grapheme-to-phoneme correspondences and other phonological rules of the language. For languages with transparent orthography, letters seem to be an efficient unit used by children learning to read. On the contrary, for languages with less transparent orthography small units (letters and phonemes) are not
enough and children need to use bigger units such as syllables, rimes or even the whole word units, e.g., ‘yacht’ or ‘choir’ (Ziegler et al., 2005, Goswami, 2005).

As shown above, the orthography of languages is different and this difference affects reading acquisition in that particular language. The details of different writing systems of the world languages are given in the following section.

1.2. Writing Systems

Every writing system has its own underlying linear organization: that is, every symbol in a language must be placed together in some sort of predictable order. For example, English is written in a horizontal line from left to right and the lines are ordered from top to bottom; Hebrew and Arabic are also written in horizontal lines but from right to left; Chinese was traditionally written vertically starting at the upper right whereas Mongolian was written vertically starting from the upper left (Rogers, 2005). However, not all languages follow such linear organization principles in a strict uncomplicated manner. Some languages, while written in a horizontal line from left to right, have some vowel characters and diacritics presented vertically above or below the main line the initial consonant. Examples of such languages are Thai, Lao, Khmer, and Black Tai.

Writing systems can be classified into two main types, alphabetic and ideographic according to the symbols that are used in the languages, and each of these is described in turn below.

1.2.1. Ideographic Writing Systems

Ideographic writing systems can be logographic or logophonetnic systems.
(1) Logographic systems use a large number of signs to represent morphemes (the minimal unit in a language that conveys meaning). A logogram (a sign in a logographic language) may represent a word or part of a word. Examples of logographic languages are Chinese, Nushu and Khitan.

(2) Logophonetic systems are similar to logographic systems, but consist of two types of signs; ones for morphemes and ones for sounds. Most of the logophonetic systems are also logosyllabic, meaning that generally each sign represents a syllable, rather than a single phone or a whole morpheme. These languages include Sumerian, Maya and Japanese.

1.2.2. Alphabetic Writing Systems

Alphabetic writing systems use characters or alphabets instead of signs to represent the sounds of the language. Alphabetic systems can be Consonant Alphabetic, Syllabic Alphabetic, or Full Alphabetic systems, as set out below.

(1) Consonant Alphabetic systems: In these systems vowels do not have any form, i.e. they are not represented at all. Such systems involve purely consonantal character where each character represents a consonant (and vowel) and speakers must remember which vowel co-occurs with which consonant. Examples of such languages are Proto-Sinaitic, Aramaic, Arabic and Hebrew.

(2) Syllabic Alphabetic systems: The characters in these systems fit into both a syllabary and alphabet. They are syllabic because each character represents a consonant and an inherent vowel, but the vowel is constant in any one character. Moreover, consonant characters can combine with other vowel characters to make different syllables. Most languages of South and South East Asia have such systems.
(3) Full Alphabetic systems: In these systems, the characters represent only a consonant or a vowel sound. These languages include most European languages like English, German, and Italian.

Another important aspect of the writing systems of different languages is the use of spaces between words. There are two writing styles: spaced and scriptio continua, as set out below.

(1) Spaced languages: In spaced languages words are combined into sentences, and these are written with spaces between words, as in most European languages.

(2) Scriptio continua languages: In scriptio continua languages, unlike spaced languages, writing is done without spaces between words. Examples of scriptio continua languages are South East Asian languages, such as Thai, Lao, and Khmer. In fact, in Thai a whole paragraph can be written without any spaces at all. Details of Thai language orthography are discussed in the next section.

1.3. Thai Language

Thai is the national and official language of Thailand. It is also a tonal language, that is, words are constructed and meaning is conveyed not only by consonants and vowels but also by tones which are essentially pitch-based variations (refer Section 1.3.1). Below the general characteristics of the Thai orthography are described ahead of a description of the specific characteristics of Thai consonants, vowels, tones and other aspects (refer Section 1.3.2). Thai is an analytic language, that is, it has no infections such as tense or plural morphemes.
1.3.1. **Thai Script**

Thai has its own alphabetic orthography. The Thai alphabet is a syllabic alphabetic script and it is written from left to right in a horizontal line starting from the top of the page. All consonants and some vowel characters are written in the main line; some specific vowel characters are presented vertically above or below initial consonant character(s); all tone markers are placed above initial consonant character(s); plus some diacritics and special symbols can be placed above final consonant character(s) (refer Figure 1 below). As a result, Thai is considered to have a complex orthography even though it has a high level of grapheme-phoneme correspondence.

![Diagram showing orthographic positions in Thai](image)

**Figure 1.** Diagram showing orthographic positions in Thai

(including the (usual) horizontal orthographic positions and vertical orthographic positions with ‘0’ as the main line which includes consonant and some vowel characters, ‘-1’ on which some vowels are written, ‘1’ on which other vowels are written, and ‘2’ on which tone characters are written).
Thai is written without spaces between words although spaces in Thai text are used as a form of punctuation, indicating the ends of phrases, clauses or sentences. Therefore, individual words are recognized by scanning the text for word boundaries. Word segmentation in Thai text depends very strongly on sentential context. Many Thai word strings are ambiguous, for example “ตากลม” can be read as “ตาก ลม” [tà:k lôm] (“exposed to wind”) or “ตา กลม” [tā: klôm] (“round eyes”) as in the following two sentences:

นิดมีกระต่ายตัวเล็กตากลมตัวหนึ่งชื่อปุกปุย
‘Nid has one small round-eyed rabbit named Pukpui.’

เก้าอี้ถูกทิ้งตากแดดตากลมอยู่สนามหลังบ้าน
‘The chair was left exposed to the sun and wind in the backyard.’

Thai words can be divided into three types: single words, compound words and complex words. Single words consist of only one morpheme but may have one or more syllables. Compound words are the combination of two or more single words. Complex words consist of two or more single words in which one word acts like a prefix or suffix.

For example, the word ‘การเดินทาง’ [kā:n dīːn tāːŋ] consists of two parts; the first part is the word ‘การ’ [kā:n] normally means ‘work/job’, but used as a prefix it changes the following compound word ‘เดินทาง’ [dīːn tāːŋ] meaning ‘travel’ from a verb
to a noun meaning ‘travel’. Thai orthographic syllables may consist of a consonant, a vowel, a tone marker (it is possible that vowel characters and tone markers may not be presented with a vowel and tone sounds being implicit in the characters) plus a possible final consonant or other diacritic. There are two types of syllable in Thai. The first one, live syllables which are either open syllables, consists of an initial consonant and a long vowel without a final consonant, or closed syllables with a short or long vowel that end with a nasal or approximant consonant. The second type is dead syllables which are open syllables containing short vowel sounds or closed syllables with a short or long vowel that ended with a stop consonant (refer details of final character in Table 8).

Thai words range from consisting only one consonant without any vowel, as in ‘น’ [nā] meaning ‘at / of’, to a word a paragraph long, such as the full name of Bangkok in Thai:

‘กรุงเทพมหานครอมรรัตนโกสินทร์มหินทรายุธยามหาดิลกภพนพรัตน์ราชธานีบุรีรมย์อุดมราชนิเวศมหาสถาน’

[krūŋ tʰēːp má hāː nā kʰāːn ṭā mǎːn rát tā nà kōː sīn má hīn tʰā rāː jūt tʰā jāː má hāː dî lōk pʰōp nōp pʰā rāt rāːt tʰā tʰāː nīː bū rīː rōm ṭū dōm rāːt tʰā nī wēːt má hāː sā tʰāːn ṭā mǎːn pʰī mǎːn ṭā wā tāːn sā tʰīt sāk kā tʰāːt tī jā wīt sā nū kām má prā sit]

Below, further details of consonants, vowels, tones, other symbols, and grrophytactic rules are discussed.
1.3.2. Thai Orthography

Consonants

Thai has a syllabic alphabetic system consisting of 44 basic consonant characters (plus 4 archaic characters) for 21 initial and nine final consonant sounds. These characters in Thai alphabetical order are ก, ข, ฃ, ค, ฅ, ฆ, ง, จ, ฉ, ช, ซ, ฌ, ญ, ฎ, ฏ, ฐ, ฑ, ฒ, ณ, ด, ต, ถ, ท, ธ, น, บ, ป, ผ, ฝ, พ, ฟ, ภ, ม, ย, ร, ล, ว, ศ, ษ, ส, ห, ฬ, อ, ฮ and the four archaic characters are ฤ, ฦๅ, ฦ, ฦๅ. All of these characters can occur in the initial position, but only 33 characters can occur in the final position. The grapheme-phoneme correspondences of Thai initial consonants are shown in Table 3 below.

Table 3. Initial consonant grapheme-phoneme correspondences

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manner of articulation</td>
<td>IPA Thai</td>
<td>IPA Thai</td>
<td>IPA Thai</td>
<td>IPA Thai</td>
<td>IPA Thai</td>
<td>IPA Thai</td>
</tr>
<tr>
<td>Stop</td>
<td>Voiceless Aspirated</td>
<td>pʰ</td>
<td>ผ</td>
<td>พ</td>
<td>ภ</td>
<td>ฅ</td>
</tr>
<tr>
<td></td>
<td>Voiceless Unaspirated</td>
<td>p</td>
<td>ฎ</td>
<td>ฎ</td>
<td>ฏ</td>
<td>ต</td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td>b</td>
<td>บ</td>
<td>บ</td>
<td>ฎ</td>
<td>ด</td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>ม</td>
<td>ม</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f</td>
<td>ร</td>
<td>ร</td>
<td>h</td>
<td>ฮ</td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>ล</td>
<td>ล</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>r</td>
<td>ร</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>w</td>
<td>ว</td>
<td>j</td>
<td>ย</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Initial consonant characters are separated into three consonant classes according to the voicing and aspiration which in turn determines the number of tone sounds that it can be realised with and the tone characters it can co-occur (refer Table 4). These classes are the High consonant class, which are the voiceless aspirated stop consonant sounds. Consonants in this class can occur with only three of the five Thai tones (low level, falling, and rising) and can be marked by only two tone characters (Mai Eak [ė] for low level, Mai Tho [ɔ] for falling, and no marker for rising tone); Mid consonant class, which can occur and can be marked with every tone marker ([ę] for low level, [ɔ̄] for falling, [ɔ̄] for high tone, and [ɔ̄] for rising tone, and no marker for mid level tone); and Low consonant class, which can occur with only three tones (mid level, falling, and high level), and can be marked by only two tone markers ([ɔ] for falling, and [ɔ̄] for high level, with the mid level tone being unmarked).

Since consonants in High and Low consonant classes cannot occur with all five tones and the tones that they occur with are not the same (except the falling tone), the consonant characters from these groups that have the same phoneme representation are called paired consonants as shown in Table 5. These paired consonants allow a particular initial consonant sound to co-occur with all five tone sounds. Further details of paired consonants and the combination of initial consonant characters and tone markers in different syllable types are shown in Tables A1 to A5 in Appendix A.
Thai has only initial consonant clusters, and these can be divided into three groups on the basis of orthography, as follows:

1) True consonant clusters: These are clusters which are both orthographically and phonologically realised, with the phonological realisation bearing a direct relationship to the orthography. There are fifteen of these. All start with stop consonants, and these are followed by laterals, trills, or bilabial approximants. These clusters are /kr/, /kl/, /kw/, /kʰr/, /kʰl/, /kʰw/, /kʰr/, /kʰl/, /kʰw/, /tr/, /pr/, /pl/, /pʰr/, /pʰl/.  

2) Class-change clusters: These are orthographic (but not phonological) clusters in which first consonant of the cluster, ʰ or ə, is silent and is used to change sonorant consonants from Low to High class, so that the rising tone can be
realized phonologically. These clusters are หม /h+m/, หน /h+n/, หง /h+ŋ/, หร /h+r/, หล /h+l/, หว /h+w/, หย /h+j/, หญ /h+j/, อย /ʔ+j/.

3) Pseudo-consonant clusters:

3.1) In the first pseudo-cluster type only the first consonant of the pair is pronounced, or is pronounced as another sound. These clusters are จร /c/, สร /s/, ศร /s/, ทร /ʔ/s/.

3.2) In the second pseudo-cluster type clusters are pronounced as two syllables, because the first consonant pronounced with /a/ sound. The combinations of these clusters are shown in Table 6 below.

Table 6.  Two syllable pseudo-consonant clusters

<table>
<thead>
<tr>
<th>First consonant class</th>
<th>Second consonant class</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>งน ม ย ร ล ว</td>
</tr>
<tr>
<td>Mid</td>
<td>งน ม ร ล ว</td>
</tr>
<tr>
<td>Low</td>
<td>งญ ม ย ร ล ว</td>
</tr>
</tbody>
</table>

Another characteristic that complicates Thai orthography is the change in grapheme-phoneme correspondences of consonants when they occur in the final position. Thai has only nine final consonant phonemes, /p, t, k, ʔ, m, n, ง, w, j/. Among these, only the voiceless glottal stop /ʔ/ does not have an orthographic form; it is sometimes considered as the phonetic realization of short vowel in open syllables [syllable without final consonant character]. There are eleven groups of consonant that undergo sound change when occur in final position, as shown in Table 7 overleaf.
Table 7. **Consonant grapheme-phoneme correspondence changes in final position**

<table>
<thead>
<tr>
<th>Orthographic Representation</th>
<th>Phonological Representation</th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>ข ค ฆ</td>
<td>kʰ</td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>จ</td>
<td>c [ʨ]</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>ช คʰ ฅ [ʨʰ]</td>
<td>t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ซ ศ ษ ไทย</td>
<td>s</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>ฏ ด</td>
<td>d</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>บ</td>
<td>b</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>ป ṃ</td>
<td>pʰ</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>ร</td>
<td>r</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>ล ṛ</td>
<td>l</td>
<td>n</td>
<td></td>
</tr>
</tbody>
</table>

Thai does not allow final consonant clusters phonologically, therefore final consonant cluster in orthography are pronounced as a single sound. These orthographic clusters are ชร /cʰ+r/ pronounced as /t/, ตร /t+r/ pronounced as /t/, ทร /tʰ+r/ pronounced as /t/, รถ /r+tʰ/ pronounced as /t/, รท /r+tʰ/ pronounced as /t/. The grapheme-phoneme correspondences of Thai final consonants are shown in Table 8.
Table 8. Final consonant grapheme-phoneme correspondences

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>Bilabial</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless Unaspirated</td>
<td>p บ ป ผ ฝ ฟ น</td>
<td>t จ ช ซ גד ด ต ท ฑ ฒ ต ถ ท ธ ศ ษ ส</td>
<td>k ข ค ฅ ฆ ฐ ถ ท ฑ ฒ ต ถ ท ธ ศ ษ ส</td>
<td>? ฌ ญ ณ น ร ฬ</td>
<td>-</td>
</tr>
<tr>
<td>Stop</td>
<td>m ฌ ญ ณ น ร ฬ</td>
<td>n ญ ณ น ร ฬ</td>
<td>j ย</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>w ว</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The relationship between graphemes (characters) and phonemes (sounds) can be divided into two types; feedforward consistency and feedback consistency. The relationship between graphemes and phonemes in feedforward situation is from grapheme-to-phoneme; if one grapheme represents one phoneme (one character represents one sound), then feedforward consistency is high, but if one grapheme represents more than one phoneme (one character represents many sounds), then feedforward consistency is low. On the other hand, in the feedback case the relationship is from phoneme-to-grapheme; if one phoneme represents one grapheme (one sound stands for one character) then feedback consistency is high, but if one
phoneme represents many graphemes (one character represents many sounds) then feedback consistency is low (Stone, Vanhoy & Van Orden, 1997).

Thai consonant character grapheme-to-phoneme relationships are quite complex. When considering the relationship of the grapheme-phoneme correspondences of characters separately in each position (either in initial or final), most are feedforward consistent, but many are feedback inconsistent. That is most characters have just a single sound associated with them, but many sounds have a number of orthographic representations.

On the other hand, when considering the entire system of Thai consonant characters, feedforward consistency is low because some Thai consonant characters undergo sound change when they occur in the final position. It is possible to say that there are 23 characters which are feedforward inconsistent, 15 characters which are feedback inconsistent and only six which are feedforward consistent.

**Vowels**

Vowel length is linguistically significant in Thai, i.e., Thai has both short and long vowel systems. Vowel characters can be written before, after, above, or below the initial consonant(s), or in some cases characters a combination of positions form together to represent one vowel phoneme. There are 18 monophthongs in Thai (nine short and nine long) and six diphthongs, plus three vowel characters that are actually pronounced as a vowel plus final consonants and are considered to be a short vowel sound with a final consonant sound phonologically. Short vowels, long vowels,
diphthongs, and vowel plus final consonant grapheme-phoneme correspondences are shown in Tables 9, 10, 11, and 12, respectively.

Table 9. Short vowel grapheme-phoneme correspondences

<table>
<thead>
<tr>
<th></th>
<th>Thai</th>
<th>Central</th>
<th>Thai</th>
<th>Back</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td>i</td>
<td>ᅌ</td>
<td>i</td>
<td>ᅌ</td>
<td>u</td>
</tr>
<tr>
<td><strong>Mid</strong></td>
<td>e</td>
<td>ɨ</td>
<td>ɤ</td>
<td>ɨ</td>
<td>o</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>ɛ</td>
<td>ɯ</td>
<td>a</td>
<td>ʊ</td>
<td>ɔ</td>
</tr>
</tbody>
</table>

Table 10. Long vowel grapheme-phoneme correspondences

<table>
<thead>
<tr>
<th></th>
<th>Thai</th>
<th>Central</th>
<th>Thai</th>
<th>Back</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td>i:</td>
<td>ᅌ</td>
<td>i:</td>
<td>ᅌ</td>
<td>u:</td>
</tr>
<tr>
<td><strong>Mid</strong></td>
<td>e:</td>
<td>ɨ</td>
<td>ɤ:</td>
<td>ɨ</td>
<td>o:</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>ɛ:</td>
<td>ɯ</td>
<td>a:</td>
<td>ʊ</td>
<td>ɔ:</td>
</tr>
</tbody>
</table>

Table 11. Diphthong grapheme-phoneme correspondences

<table>
<thead>
<tr>
<th></th>
<th>Thai</th>
<th>Thai</th>
<th>Thai</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short</strong></td>
<td>ia</td>
<td>ɨə</td>
<td>ia</td>
<td>ɨə</td>
</tr>
<tr>
<td><strong>Long</strong></td>
<td>i:a</td>
<td>ɨə</td>
<td>i:a</td>
<td>ɨə</td>
</tr>
</tbody>
</table>

Table 12. Vowel characters represented vowel plus final consonant

<table>
<thead>
<tr>
<th></th>
<th>Thai</th>
<th>Thai</th>
<th>Thai</th>
<th>Thai</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>am</td>
<td>ᅌ</td>
<td>aj</td>
<td>ɿ</td>
<td>aj</td>
<td>ɿ</td>
</tr>
</tbody>
</table>

1 Vowels in Tables 9 and 10 are presented following the IPA chart
With respect to the positioning of the characters and their components, Thai vowels can be put into ten groups, as follows:

1) Preceding vowels: utable, utable, utable (e.g. ่)
2) Following vowels: utable, utable (e.g. ำ)
3) Upper vowels: ่, ่, ่, ่ (e.g. ไต)
4) Lower vowels: ุ่, ุ (e.g. ห)
5) Preceding plus following vowels: utable, utable, utable, utable, utable (e.g. เ)
6) Preceding plus two following vowels: utable, utable (e.g. เ)
7) Preceding, upper plus following vowels: ไต, ไต (e.g. บัน)
8) Preceding, upper plus two following vowels: ไต, ไต (e.g. เ)
9) Upper plus following vowels: ไต, ไต (e.g. บัน)
10) Upper plus two following vowel: ไต (e.g. บัน)

Some vowel characters change their forms when they occur in closed syllables [syllable with final consonant character]. These vowels are shown in Table 13 below.

Table 13. Vowel characters change in closed syllable

<table>
<thead>
<tr>
<th>Open syllable form</th>
<th>Closed syllable form</th>
</tr>
</thead>
<tbody>
<tr>
<td>utable</td>
<td>ุCf</td>
</tr>
<tr>
<td>ำ</td>
<td>ำCf</td>
</tr>
<tr>
<td>utable</td>
<td>์Cf</td>
</tr>
<tr>
<td>า</td>
<td>าCf</td>
</tr>
<tr>
<td>ำ</td>
<td>ำCf, ำCf</td>
</tr>
<tr>
<td>ำ</td>
<td>-</td>
</tr>
<tr>
<td>ำ</td>
<td>ำCf</td>
</tr>
<tr>
<td>ำ</td>
<td>ำCf</td>
</tr>
</tbody>
</table>

**Tones**

There are five tones in Thai but only four tone markers, which are [่, ์, ้, ์].

These four tone markers interplay with initial consonant classes to get five tone sounds. Tone markers (if present) are always placed above initial consonant and if the
initial is a cluster they are placed above the second component of the pair. In cases where the vowel of the syllable is an Upper vowel or has an Upper vowel as a part of the vowel character, tone markers are placed above that vowel. Tone markers in Thai are shown in Table 14.

<table>
<thead>
<tr>
<th>Tone realization in Mid consonant class</th>
<th>Tone marker name</th>
<th>Tone marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid level</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low level</td>
<td>mai ek</td>
<td>之中</td>
</tr>
<tr>
<td>Falling</td>
<td>mai tho</td>
<td>之</td>
</tr>
<tr>
<td>High level</td>
<td>mai tri</td>
<td>之</td>
</tr>
<tr>
<td>Rising</td>
<td>mai cattawa</td>
<td>之</td>
</tr>
</tbody>
</table>

Examples of Thai words with different tone sounds:

isLoggedIn [pā:] means ‘to throw’

isLoggedIn [pà:] means ‘forest’

isLoggedIn [på:] means ‘aunt (older sister of father or mother)’

isLoggedIn [pá:] means ‘father’

isLoggedIn [pà:] means ‘father’

**Other Symbols**

There are a number of other symbols in Thai orthography, as follows:

i)  _วรรณ์/ (a+n) is another character standing for /a/ sound plus final /n/ but if it co-occurs with other final consonant character(s) the /n/ sound will change into that final,
ii) ڄ (/a/) is another form of the vowel _ƨ (/a/) when preceding a final consonant,

iii) Ӎ marks silent consonants at the end of a syllable,

iv) 럼 replaces the (ƨ) character of the short vowel (__, __) when occurring with a final consonant,

iv) _.nanoTime stands for reduplication of the whole word,

v) _Nano shows that the word is a short form of a longer word (not an abbreviation),

vi) น rallies means ‘et cetera’ or ‘and so on’.

vii) Finally, some distinct characters are used for Thai numbers. These are ๑ (one), ๒ (two), ๓ (three), ๔ (four), ๕ (five), ๖ (six), ๗ (seven), ๘ (eight), ๙ (nine), ๐ (zero),

1.4. Learning to Read in Thai

1.4.1. Literacy Instruction in Thailand

According to the Thai Basic Education Curriculum B.E. 2544 (2001), the official age for Thai children to enter the education system at Grade 1 is 7 years of age. However, Thai children normally attend kindergarten school (equal to Preschool level in Australia) for three years prior to this. Therefore, the usual age for Thai children to start formal education is 3 to 4 years of age. Thai children learn alphabet character
names in Kindergarten. Teachers may teach them to read some basic words such as colours, fruits, animals, days, and months, but linguistic knowledge and graphotactic rules are properly taught at the primary level (starting from Grade 1).

In the text books used to teach Thai children in grade 1 to 6, literacy lessons are based on the whole-word approach with additional phonic strategies to teach children how to map characters to sounds and gain a better understanding of the phonological realisation of different combinations of initial consonant classes and tone characters (Sirikanjanapong, 1996).

Of particular important for this thesis is the fact that Thai children start learning to read with spaced texts, they study with spaced text until the end of Grade 1 when lessons change to use unspaced text from the start of Grade 2. Examples of passages from Grade 1 and 2 text books\(^2\) are shown below:

**Text from Grade 1 text book (children learn to read using spaced text):**

เวลา เที่ยง แดด ร้อน กล้า กบ โจม วิ่งเล่น นา แกว แกว มะลิ นั่งเล่น ที่ ศาลา ทาน้ำ

[\[wːlāː tʰːiːŋ dɛːt tʰːr̚ːn klâː kàp tʰːŋ mïːn jîːm wîːŋlên nāː nàː kēːw kâːp mâːlî nâːŋlên tʰː sāːlāː tʰːnâːːm\]]

Literal translation: ‘time noon sunshine hot Kla and Jom run+play in field Kaew and Mali sit+play at pavilion riverbank’.

Translation: ‘At noon, even though it’s hot Kla and Jom are playing and running in the rice field while Kaew and Mali play in the pavilion on the riverbank.’

\(^2\) All text books used to teach Thai subject in Thai primary and secondary school must be proved by Office of Basic Education Commission, Ministry of Education.
Text from Grade 2 text book (children learn to read using unspaced text):

คุณครูพูดว่า “วันนี้แดดไม่ร้อน ลมพัดเย็นสบาย บนต้นไม้มีนกนอยมาอาศัย

[kʰūnkrūːː pʰúː t waː wān nǐː dēː t màː j rīː n lōm pʰâːt jēn sâbāː j bōn tōnmâː j kâː mīː nōk nōː j māː ʔīːː sāːj]

Literal translation: ‘teacher speak talk “today sunshine not hot wind blow cool comfortable on tree have bird small come’

Translation: ‘The teacher said “The weather today is nice not too hot with cool breeze. There is a little bird living in the tree.’

1.4.2. Issues in Literacy Development and Reading

There are a number of points of interest in the above review that are of interest for this thesis. As Thai children begin to learn to read with spaced text and then move onto unspaced text in Grade 2 (around 8 years of age), it is of great interest to study the effects of the absence of spaces between words on reading of children in Grade 1 compared with Grade 2. Thai orthography seems to be complicated even though its phonological system is simple and Thai is written without spaces between words. For these reasons it will be of great interest to investigate eye movements in reading Thai. It is possible that some of the complexities of Thai orthography might be used by Thai readers to control eye movements in reading. It will be of interest to investigate the factors that allow Thai readers to competently read such a complex text and how these develop over age. The results of these investigations should help provide information not only on reading in Thai, but also on how to improve models of eye movements in reading, so that they are capable of explaining eye movement behaviour with both spaced and unspaced texts. In the next chapter we move on to investigate studies of eye movements in reading.
CHAPTER TWO

EYE MOVEMENTS IN READING
2. EYE MOVEMENTS IN READING

2.1. Introduction

Eye movements in reading are complex because the eyes do not move smoothly across the text. On the contrary, during reading, the eyes rather move back and forth, i.e., there are both progressions and regressions (Rayner, 1979; Rayner, 1998; O'Regan, 1990). In reading, the eyes move in a sequence of very fast, relatively well-coordinated movements known as saccades. These saccadic movements are punctuated by fixations, periods of relative stability in the direction of the visual axis, during which visual information can be extracted (Radach & Kennedy, 2004). Eye movements in reading are believed to be a window onto the cognitive processes involved in reading and indeed eye movement data provide a relatively natural method for investigating psycholinguistic issues since they are part of the normal reading process (Rayner & Pollatsek, 2006). However, while eye movement data may be very informative regarding the lexical processing and understanding reading, they are not perfect reflections of the mental activities associated with comprehension (Rayner et al., 2006).

The study of eye movements in reading research can be divided into three eras. The first began in 1879, when Professor Emile Javal initially observed the role of eye movements in reading, and continued through until 1920. In this era, basics facts about eye movements, such as saccade latency and perceptual span, were discovered (Rayner, 1998). The second era, from the 1920s to the mid-1970s, coincided with the reign of the behaviourist movement in experimental psychology, and in this period eye movement studies tended to have a somewhat applied focus. There were only a few studies conducted during this era and relatively little information emerged about the
cognitive processes underlying eye movements. The third era began in the mid-1970s and continues until the present day. In concert with the “cognitive revolution”, which began around the time of the publication of Neisser’s book, ‘Cognitive Psychology’ (1969), most studies in this period have been concerned more with the implications of eye movements during reading for cognitive processes. In addition, the development of advanced technologies during this period has allowed researchers to collect and analyse a large and varied amount of data on eye movements during reading (Rayner, 1998).

This chapter provides general background information regarding the measurement of eye movements, approaches to and models of eye movement control in reading, as well as a review of studies of eye movements in reading, especially in the area of spaced and scriptio continua (unspaced) languages.

2.2. Measurements of Eye Movements in Reading

The fact that in reading the eyes do not move smoothly and uniformly forward along the line of text, but rather move back and forth, suggests that eye movements in reading involve a number of factors. Consequently, the analysis of eye movements in reading is normally based on a number of measures as follows: reading rate, average fixation duration, saccade size, number of fixations or refixations, regressions, landing site position, first fixation duration, and gaze duration (Rayner & Pollatsek, 1996; Rayner, Fischer & Pollatsek, 1998; Rayner 1998). These measures follows are elaborated below.
Reading rate is the number of words being read per minute, for example adult English readers have been found to read normal sentences (written with spaces between words) at the average rate of 261 words per minute (Rayner et al, 1998).

Average fixation duration is the average time in milliseconds (msec.) that a particular reader spends fixating a particular point in the text, averaged across a large amount of text. The average fixation duration for English readers is about 200 to 250 msec. (Rayner, 1998).

Fixation duration: First fixation duration is the time a reader spends on just the first fixation of a currently fixated word; while gaze duration refers to the time that reader spends fixating the word for the first time before subsequently leaving it regardless of the number of fixations; total reading time is the total time spent on a word and may include regressions back to the word after having left it.

Saccade size is a measure of the extent of the rapid movements of the eye across the text (saccades). Saccade size is usually measured by the numbers of letters between each fixation. Depending on viewing distance, 3 to 4 letter spaces are equivalent to 1 degree of visual angle. The normal saccade size of English readers reading spaced text is typically 7 to 9 letter spaces, i.e., around 2° of visual angle (Rayner, 1998).

Landing site position is the point at which readers focus their first fixation in a word and is measured in terms of letter spaces into a word. As a study by Rayner (1979) showed, readers tended to have a preferred viewing location (PVL) for their first fixation during reading. In European languages all or most readers have the same PVL – to the left of the centre of the word. This PVL is closer to the Optimal Viewing
Position (OVP), the word-centre, which allows the readers fastest word recognition. There can, however, be individual variation, and certain conditions may affect the PVL, e.g., European readers’ PVL changes if there are no spaces between words, as will be discussed in more detail on page 50.

Number of fixations or refixations is the number of fixations within single words. Studies have shown that first landing position, word length and word frequency all affect the number of fixations. For example, English readers tend to make more fixations if the first fixation is not at the usual PVL or if the word is relatively long, or is a low frequency word, and even more fixations if the word is low frequency long word.

Regressions refer to the percentage of eye movements back to a word prior to the currently fixated one. The word to which the regression occurs might be the word preceding the fixated word or any word in the text that comes before the currently fixated word. Studies of English readers’ eye movements have shown that for most readers around 10 to 15% of their eye movements during reading are regressions (Rayner, 1998).

2.3. Approaches in Eye Movement Studies

It is possible to divide theories of eye movement control in reading into four main types (Epelboim, Booth & Steiman, 1994). The first is text-independent eye movement control (Haber, 1976), in which, as the name indicates, parameters of eye movements such as progressive and regressive saccades, size of saccade, or fixation duration are adjusted by the readers in each reading session on the basis of global features of the text. That is, eye movements are not associated with either the physical
parameters, e.g., word length, or any higher order variables, such as semantics, of the
text being read. In this approach, it is assumed that the eye moves the same number of
characters every saccade regardless the word length (Epelboim et al., 1994). However,
Rayner (1979) found that in reading European texts, readers tend to show a “preferred
viewing location (PVL)”, that is the first saccade in a word tends to fall to the left of
the word centre. This shows that eye movements are not the same from one saccade to
another, as they depend on the length of words, so this approach, at least in its strong
form, cannot be accepted.

The second type of theory is Suppes’ probabilistic eye movement model or
text-dependent probabilistic control model (Suppes, 1990). In this it is assumed that
moment-to-moment saccadic programming depends on local characteristics of the text
in specific ways, and the importance of probabilistic factors is also stressed in this
approach. Suppes (1990) concluded from the results of a series of studies that at every
fixation, the reader must choose one of several types of eye movement operations, such
as staying within a word, going to the next word, going back to the previous word or
skipping the next word. The probability of particular selections at a given time varies
depending on local visual, syntactic and semantic properties of the text (word length,
word frequency, grammatical structure, and cognitive processing of the text currently
fixated).

The third type of theory is that of meaning-guided eye movement control
(Just & Carpenter, 1980). As the purpose of reading is to understand the meaning of
the text, it is reasonable to postulate that the eyes will execute the next saccade only
after the reader extracts meaning from the currently fixated word. In this approach eye
movements in reading are based on semantic and syntactic features of the text, and
fixation duration will depend upon the type, meaning, and familiarity of the word. Accordingly, readers should spend less time fixating words with high than low frequency. For example, O’Regan (1979) found that short informative context words such as “eat” are fixated more often than short grammatical function words like “the”.

The final type is referred to as **visually-guided eye movement control**. This is similar to Suppes’ probabilistic eye movement model, but focuses more on the effects of visual cues in eye movement behaviour. The results of studies (Rayner, 1979; Just & Carpenter, 1980) showing that readers tend to have a particular PVL and different saccade size depending on word length, suggest that the length of the word currently fixated and the word to its right have an effect on programming of saccades in reading. In keeping with this approach, O’Regan (1990) proposed a variant of the visually-guided eye movements approach, the “strategy-tactics” eye movement theory, which places importance on local physical features of the text especially the spaces between words, and introduced the notion of “optimal viewing position” (OVP).

The OVP is a position near the centre of a word which, research has shown, allows fastest recognition of the word (O’Regan, Levy-Shoen, Pynte & Brugaillere, 1984; O’Regan, 1990; O’Regan & Jacobs, 1992). Readers of European writing systems tend to land their eye close to this OVP, to the left of the centre of the word. This suggests that there is use of a prominent cue or cues to word onset, offset, and length, e.g., the spaces between words, that could be used to control saccades to land near the OVP. The notion of the OVP, unlike that of the PVL, has been derived from studies of reading in European languages in which word boundaries are visibly presented. Thus, the OVP may or may not be applicable to languages without visible word boundaries. On the other hand, the PVL refers to any particular position that the
majority of readers in a particular language and script tend to first fixate in a word. Contemporary studies in eye movements are usually based on this OVP approach because most of them have studied eye movements in reading European languages, languages written with spaces between words. However, not all languages are written with spaces between words, an issue of central interest in this thesis.

2.4. Models of Eye Movement Control in Reading

Since there are many factors that influence reading such as the transparency of orthography and complexity of the syntax, it is almost impossible, at present, to develop a model that can explain all the phenomena that have been observed in reading because a model that can explain everything would be extremely complex (Rayner et al., 2006). Therefore for a model of reading to be of some value, it needs to be able to explain (not just describe) at least one significant part of the reading process, e.g., where and when the eyes move, but be simple enough so that it is useful tool for understanding which phenomena it can explain and which it cannot (Rayner et al., 2006).

At present, there is some research that has resulted in the development of quantitative models of eye movements in reading. These models of can be generally categorised into two classes, a) models that explain eye movements as the result of lexical and other ongoing comprehension or cognitive processes (Processing Models) and b) models that explain eye movements as mainly controlled by the oculomotor factors and are only indirectly related to ongoing language processing (Oculomotor Models). The primary models are E-Z Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998), Strategy-tactics (O’ Regan, 1990, 1992), SWIFT (Engbert, Longtin, & Kliegl, 2002), Glenmore (Reilly & Radach, 2003a), SERIF (McDonald, Carpenter &
Shillcock, 2005), Mr. Chips (Legge, Klitz & Tjan, 1997) and the Competition/Interaction model (Yang & McConkie, 2001). Though these models differ on a number of dimensions, they all share certain low-level features, as none attempt any serious modelling of higher-level processes such as how text is parsed or how discourse structures are constructed (Rayner et al., 2006).

Since the eye movement studies in this thesis will focus on the physical features that orchestrate eye movements during reading of Thai readers, only the E-Z Reader, Strategy-tactics and Glenmore models, good examples of processing, oculomotor and oculomotor-cognitive models respectively will be illustrated in this section.

2.4.1. E-Z Reader Model

The E-Z reader model (Rayner et al., 1998, Reichle et al., 1998) is a Processing model. It is based on the assumption that eye movements during reading are influenced by on-going cognitive (linguistic) processing. It can be regarded as a simulation of reading when higher-level linguistic processing is running smoothly, but as a consequence the model does not explain inter-word regression (Reichle, Rayner & Pollatsek, 2003). The E-Z Reader model should be viewed as a ‘default’ reading process, i.e., the process of encoding words as a forward-moving ‘driving engine’ in reading (Reichle et al., 2003). It explains how lexical processing influences the progress of the eyes through the text. It provides a theoretical framework for understanding word identification, visual processing, attention, and oculomotor control determining when and where the eyes move during reading (Reichle, et al., 2003).
In this model, it is considered reasonable to work on the hypothesis that linguistic processing affects eye movements in two different ways. First, there is a relatively low-level linguistic process that keeps the eyes moving forward. Second, higher-level processing occurs in parallel with this low-level processing and is effective when this higher-level processing is having difficulty extracting further information. For example, if there is a failure to encode a word or if errors are detected, this system intervenes to tell the first system to either stay fixated or go back and make an attempt to gain the information (refixations).

In this model, fixation locations are determined by a combination of visual, oculomotor and linguistic factors, while the fixation durations are determined by only linguistic processing. There are five sequential processes to determine when and where the eyes move during reading:

a) ‘Familiarity check’ – a quick assessment of whether or not word identification is imminent
b) Completion of lexical access – a later stage of word identification, in which the actual identification occurs
c) Labile stage of saccadic programming – the motor commands that determine whether it is necessary to move the eyes or to cancel the saccade execution
d) Non-labile stages of saccadic programming – commands the actual saccade execution after the decision was made in the labile stage to move the eyes
e) Actual saccades – the actual movements of the eyes
Figure 2. A schematic diagram of the E-Z Reader model (Reichle et al., 1998)

Figure 2 shows the influences of lexical processing on oculomotor controls and vice versa. After the initial landing site on the fixated word, lexical process starts checking the familiarity of the words to determine whether the eyes need to move to the next word in order to complete the lexical process or not. If there is the need to gain more lexical information because of the unfamiliarity of the fixated word then the saccadic program will move the eyes to a new location before current lexical process is complete. Conversely, if there is no need of saccadic execution then the lexical process will be completed and the attention system will shift attention to the next target and the whole process will be repeated again.

The mean time required to complete the first two stages of lexical processing in this model are determined by the word’s normative frequency of occurrence and word’s predictability within the sentence context. The labile and non-labile stages are the commands that determine whether the eyes should move (refer Figure 2 above). This also determines the refixation within the word, the model’s assumption being that
if the fixation location is ‘bad’, i.e. not at the OVP, then the oculomotor system begins programming a second saccade to the center of the word as soon as it fixated. It is believed that low-frequency words will tend to be refixated more than the high-frequency words.

This model assumes that lexical processing is serial in the sense that only one word is processed at a time and so the E-Z Reader model predicts that longer, less frequent and less predictable words will be (1) fixated longer, and (2) skipped less often. However, there are some problems that need to be improved in the model. First, E-Z Reader does not contain a model of word recognition. Second, it is quite vague about lexical access and does not specify what codes, for example orthographic, phonological or semantic, are accessed (Rayner et al., 2006).

The E-Z Reader model is based on the results of the studies of eye movements in reading of European languages which are spaced texts, thus the main saccadic target is based at the word level. As stated earlier, the E-Z Reader model is a cognitive processing model and it should be a good model to explain eye movement control in reading Thai, a language in which word segmentation strongly relies on the contextual meaning in which the word co-occurs. However, the scriptio continua nature of the text should be problematic for the predictions of the E-Z Reader model since there are no spaces between words to facilitate word segmentation during reading.

As the E-Z Reader model was not originally constructed to process unspaced text, it cannot predict the processing of eye movements in reading normal Thai text. However, predictions can be made for conditions in which spaces between words are inserted in Thai text. Under these conditions, the E-Z Reader model should predict:
– landing site location should be at the word centre assuming that it is the OVP,
– first fixation duration and gaze duration should be strongly related to the relative frequency of occurrence of the words and the predictability level of the context, as well as the length of the word being fixated.

2.4.2. **Strategy-tactics/S-T**

Strategy-Tactics/S-T (O’Regan, 1990, 1992) is an Oculomotor Model. The core principle of this model is that the visual properties of the text, such as word length and spaces between words, and the operating characteristics of the visual and oculomotor systems, such as visual acuity and saccade accuracy, mainly determine fixation locations. Then fixation durations are determined by where the eyes are fixated. This Strategy-tactics model originated from two observations, that a) words are identified most rapidly if the fixation locations are slightly to the left of word centre (the Optimal Viewing Position) and b) refixations are also less likely to occur if the fixation falls in this position.

The main idea of this model is that pre-determined oculomotor strategies and tactics and visuomotor constraints are the main driving force of eye movements during reading. The refixation decision in the S-T model is determined solely by the oculomotor tactics based on the initial landing position of the eyes in that particular word. The tactics are as follows: if the eyes land near the OVP, only a single fixation will occur on this word. On the other hand, if the eyes land at the beginning or end of the word, refixation will be programmed immediately without any linguistic process.
Figure 3. A schematic diagram of strategy-tactics model (Reilly, 2010)

The S-T model, which is an oculomotor model, should be the most inappropriate model to predict the eye movement control in reading the scriptio continua languages. This is because the core assumptions of the model rely mainly on the physical conditions of the text being read. Thus the weakest point of this model is that they may not be easily applied to texts in which there are no visual cues that specify word boundaries, e.g., in scriptio continua languages such as Thai.

Possible predictions of the S-T model for eye movements in reading Thai are that:

- the eyes should move consistently with equal saccadic size across the text being read regardless of the target word since there are no visible visual cue available; and
- landing site location as well as the fixation duration, gaze duration, refixations and regressions should vary from word to word depending on the initial landing site on fixated word.
2.4.3. Glenmore Model

The Glenmore model (Reilly et al., 2003a) is an oculomotor-cognitive model of eye movement control in reading which can account within one mechanism for preview and spillover effects as well as for regressions, progressions and refixations (refer Figure 4). The Glenmore model allows the interplay of factors from multiple levels of representation and assumes that potential target words are represented on a salience map and their salience values depend on the particular visual configuration. It also assumes that at the beginning of each fixation, low level visual information is available which allows saccadic triggering without any cognitive influence and this saliency value representing potential targets will change in response to information of the ongoing linguistic processing (Reilly et al., 2003a). For example, if a reader is fixated on a letter in the right of a medium length word, and the next word (N+1) is short and the N+2 word is of medium size, then it is likely that the word N+2 will have highest initial salience value and will be the target for the next inter-word saccade (Reilly et al., 2003a). In addition, the Glenmore model includes a visual input module, a word processing module, fixate centre and a saccade generator modules, producing the actual saccadic movement (Reilly et al., 2003). The visual information is transferred to the saliency map representation and to a linguistic processing module that implements processing on the letter and word level within an interactive activation framework.

The Glenmore model allows the processing of words in parallel with a limited amount of competition between words in a given fixation. Unlike the E-Z Reader model that focuses on processing one word at a time, the Glenmore model assumes that lexical processing simultaneously occurs on the fixated word and the one on the
right (Rayner et al., 2006). The Glenmore model also allows a substantial influence of linguistic processing from the word recognition process on decisions about when to move the eyes. The other main feature of the model is the use of a saliency map that acts as a ground for the interplay of bottom-up visual features of the text and top-down lexical features. These factors combine to create a pattern of activation that selects one word as the saccade target (Reilly et al., 2003a)

Figure 4. An overview of Glenmore Model (Reilly et al., 2003a)
The Glenmore model consists of three main components: a) a saliency map [selects saccade targets], b) an interactive-activation network system [identifies words], and c) a saccade-generator [initiate and executes the actual eye movements].

There are two units in an interactive-activation system: the letter units and word units. In addition, there is a fixate centre, which controls the decision when to execute the saccadic eye movements. The details of each unit in the Glenmore model are as follows:

a) Input units – the model assumes a visual field of 30 characters.

b) Letter units – these units receive bottom-up activation from the input units plus top-down activation from the word units. The letter unit is connected to its relevant word unit to determine the saccade target selection.

c) Saliency units – saliency units receive activation from both input and letter units. The input from the letter units provides a top-down cognitive contribution to the saliency values for a specific area of visual field. The saliency map role is to support the saccade target selection process.

d) Word units – word units receive inputs from the letter units and send the activation back to the letter units in return. There are seven word units, maximum numbers found in the visual field of 30 characters, in this model.

e) Fixate centre unit – this unit is a single unit with connections and recurrent connection from the letter units. The fixate centre unit determines when to execute a saccadic targets.

f) Saccade generator – the unit that generates and executes the actual saccadic movements.
This Glenmore model, especially the extended version of the model, seems to be the best model to explain eye movement control during reading scriptio continua languages since the model covers the eye movements on both letter and word levels at the same time. Even though it was not constructed to explain eye movements in reading unspaced text, its design principles, such as interactive activation and saliency, should permit the dynamic emergence of word boundary information from word and letter statistics. This is because the model works on both the letter (bottom-up process) and word (top-down process) levels and includes the saliency map construct, which creates the advantage of scanning the text to construct word boundaries during reading.

Predictions of an extended Glenmore model for eye movement control in reading scriptio continua languages should be as follows:

- landing site location should rely on the saliency of the letter and word being fixated, such that landing site would be further into the word if the start and end letters of the word or the frequency of occurrence of the word are high, and less far into the word if these frequencies are low; and
- fixation duration, gaze duration and refixations should be associated with word length and word frequency.

2.5. **Eye Movements in Reading Spaced and Unspaced Texts**

As shown in Section 1.2, the world’s written languages can be divided into two groups on the basis of writing styles; spaced and scriptio continua languages. Since most research on eye movements in reading has investigated European languages (which are spaced), it has been widely accepted that spaces serve as visual cues to specify word boundaries. In this vein, modern theories of reading eye movements treat spaces between words as vital visual markers that guide fixations from one position in
the text to the next (Epelboim et al, 1994). This section provides a brief history of studies of eye movements in reading using the visually-guided eye movements approach (refer Section 2.2) to examine the importance of spaces as a visual cue in guiding eye movements in reading.

2.5.1. Reading With and Without Spaces in Spaced Languages

It has been debated whether spaces between words play an important role in guiding eye movements or whether they simply facilitate reading. Accordingly, there have been many studies which have manipulated text in European spaced languages in order to investigate the effects of spaces on eye movements in reading. In such studies text is either presented without spaces or the spaces are replaced with other characters such as ‘x’ or a number.

Many of the studies on reading with and without spaces have found that average reading rate decreases approximately 30% when spaces are deleted from texts (Spragins, Lefton & Fisher, 1976; Kolers, 1968; Pollatsek & Rayner, 1982; Inhoff, Pollatsek, Posner & Rayner, 1989; Rayner et al., 1996 and Rayner, 1998). For example, Spragins et al. (1976) found that the average reading rate of English unspaced texts was significantly decreased by approximately 48%, and readers found that reading unspaced texts was difficult. European readers’ PVL when reading normal spaced text is at the left of word centre. This position allows them the fastest word recognition and thus they spend less time fixating each word. However, European readers’ PVL has been found to shift to the beginning of the word when reading unspaced texts (Rayner et al., 1996; Rayner, 1998). This shows that European readers used different oculomotor strategies when reading spaced and unspaced texts.
In contrast to these studies, Epelboim et al. (1994) studied the eye movements of native English speakers, bilingual English speakers, or second language (L2) speakers of English. Subjects were asked to read English and their native language texts presented with and without spaces both silently and aloud. Unspaced texts in this study were constructed simply by deleting the spaces from normal text. (This meant that the unspaced texts were narrower on the page – by approximately 15% - than the spaced texts.) It was found that subjects did not have any difficulty reading unspaced text; i.e., they read unspaced text with the same comprehension level. This was so, even though the average reading rate was slower, but not significantly so, for unspaced than spaced texts. The results overall suggest that the PVL when reading unspaced texts is the same as when reading spaced texts, which implies that oculomotor strategy does not change as a function of reading spaced vs. unspaced texts (refer Figure 5 below).

![Probability of landing position of subject ME reading spaced vs. unspaced texts](image)

**Figure 5.** Probability of landing position of subject ME reading spaced vs. unspaced texts (from Epelboim et al., 1994)
Epelboim et al. (1994) claimed that word recognition, not spaces, is the factor that plays the most important role in guiding eye movements during reading. Spaces simply facilitate word recognition. So, for Epelboim et al. the same model of eye movements used to explain reading spaced texts could also explain reading unspaced texts.

The differences in conclusions between the Epelboim et al. and the other studies beg resolution. Rayner et al. (1998) conducted an experiment using the same procedure as Epelboim and colleagues - subjects were asked to read spaced and unspaced sentences. In addition, they added word frequency as a factor because it is widely considered to be a determiner of the ease or difficulty of word identification and lexical access (Rayner & Fischer, 1996 and Rayner et al., 1998) - subjects were given texts with target high and low frequency words and passages. In accord with the earlier findings (e.g., Spragins et al., 1976), they found that readers had difficulties reading unspaced texts. The average reading rate decreased by 54% for reading sentences and by 44% for passages. Moreover, fixation duration was longer in the unspaced condition and on low frequency words; forward saccade size was larger in the spaced condition and subjects tended to make more regressions when reading unspaced texts.

These Rayner et al. (1998) results also suggest that readers use different oculomotor patterns with and without spaces. When spaces were present subjects tended to land their first fixation close to word centre, the usual PVL of European readers, and close to the OVP. However, their PVL changed to the beginning of words when spaces were absent (refer Figure 6 below).
Figure 6. Landing position of subjects when reading spaced and unspaced texts (Rayner et al., 1998)

Both the first fixation duration and gaze duration on the target word were longer if the word was of low frequency and even longer if that low frequency word occurred in the unspaced condition. These results show that the absence of spaces between words interfered with both word identification and eye movement control.

In their earlier studies Pollatsek and Rayner (1982, 1996) showed that spaces aid readers’ word identification in two ways - by specifying word boundaries, and by guiding the readers’ eye movements. Rayner et al. (1998) also conducted a further experiment which manipulated text in five conditions - normal spaced, unspaced, filled space with ‘x’, wide spaced, and a flanker condition (examples of these are shown in Table 15). This experiment aimed to test whether word recognition had more effect on eye movements in reading than spaces between words as claimed by Epelboim et al. (1994).
Table 15. **Illustration of the spacing conditions in the Rayner et al. (1998) experiment**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sample Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal spaced</td>
<td>This is a sample sentence.</td>
</tr>
<tr>
<td>Unspaced</td>
<td>Thisisasamplesentence.</td>
</tr>
<tr>
<td>Filled space</td>
<td>Thisxisxaxsamplexsentence.</td>
</tr>
<tr>
<td>Wide spaced</td>
<td>This is a sample sentence.</td>
</tr>
<tr>
<td>Flanker</td>
<td>Thisxisxaxxamplesxsentence.</td>
</tr>
</tbody>
</table>

It was found that average reading rates decreased according to condition in the following order - normal spaced, wide spaced, flanker, filled space, then unspaced. It was clear that subjects found the normal spaced texts easiest to read while unspaced texts were the most difficult to read, the difference in reading rate between these conditions being 58%. Fixation durations, forward saccade size and regressions, along with reading rate, indicated that the absence of spaces between words, either when they were deleted or replaced by another character (‘x’ in this experiment), affected reading - subjects spent more time on the words and needed to move their eyes back and forth more in order to gain lexical access for words. The situation was worse if the target words were low frequency. The PVL for all spaced conditions was near the word centre, but the PVL for unspaced conditions was at the beginning of the word, similar to the finding of their first experiment (refer Figure 7 below). Therefore spaces between words had stronger effects on eye movements in reading than did word recognition.
These studies show that the absence of spaces for readers of usually spaced texts confuses the reader as to where the word starts and ends, resulting in greater time being required to recognize the word. This suggests that spaces between words are a vital visual cue for readers to move from the currently fixated word to the next one; and that spaces between words do not simply facilitate word recognition as claimed by Epelboim et al. (1994). The results also support the notion that during reading readers of spaced texts tend to have a preferred viewing location (PVL) as stated in the visually-guided eye movements approach, i.e., just to the left of the word centre. However, this PVL for European readers is dependent on the presence of spaces between words, for it shifts to the start of words when spaces are absent. Thus European readers use different oculomotor patterns when reading spaced and unspaced texts.

With regard to landing site distribution, the results of studies of eye movements in reading European languages reveal that readers’ showed PVL effects and their PVL was normally near the OVP point as stated earlier (Rayner et al., 1996, 1998).
Furthermore, the results of studies on reading isolated words (O’Regan et al., 1984; O’Regan et al., 1992) found that if the first fixation location is at the OVP then this results in the shortest fixation duration for that word (the OVP effect). However, Vitu, McConkie, Kerr and O’Regan (2001) conducted a study to determine whether the OVP effect observed during reading, as revealed by a relationship between fixation location within a word and fixation duration, and found the opposite results. Their surprising results showed that the fixation duration at the OVP was the longest one not the shortest as expected, but the probability of refixation was low. They called this phenomenon the Inverted-Optimal Viewing Position effect (Inverted-OVP effect).

In spite of this, their study found similar results concerning the probability of refixations, as those found in previous studies; i.e., the probability was lowest if the first fixation was near or at the OVP. However, their study focussed only on the first fixation duration compared to the second fixation (refixation) duration (if it occurred) within a word, and they did not take gaze duration into account in their study. It might be possible that even though the fixation duration at the OVP is the longest compared to fixations on the other points in a word, the number of fixations is lowest when this first fixation location is at the OVP. Therefore, overall gaze duration for that particular word in which the first fixation is on the OVP should be the shortest compared to gaze duration for a word for which the first fixation is on the outer non-central part of word.

In conclusion, studies of eye movements in reading spaced languages have found that spaces between words, frequency of words and word length have effects on oculomotor control during reading. The OVP is the most effective target for the eye during the execution of saccades and allows the fastest word recognition.
2.5.2. Reading With and Without Spaces in Scriptio Continua Languages

The studies in the previous section showed that readers of languages which normally use spaces to indicate word boundary performed poorer when the spaces were removed, and their preferred viewing location (PVL) changed from the word centre or left of the word centre in normal spaced text to the first character of the word when reading unspaced text; that is they used different oculomotor patterns when reading spaced vs. unspaced texts. On this basis it might be expected that readers of scriptio continua languages, which normally do not use spaces at the word boundaries, would also show that the PVL is on the first character of the fixated word. Moreover, it is possible that scriptio continua readers should be able to perform better if spaces are provided, thereby giving them visual cues. That is, the average reading rate should be faster, and fixation duration shorter, and PVL might also shift to the OVP for spaced versus unspaced texts for scriptio continua readers.

Japanese script is normally written without spaces between words, but it does offer other ‘visuo-lexical’ cues to word boundaries. Modern Japanese is written in a mixture of three basic scripts: Kanji, a morphographic script; and Hiragana and Katakana, two syllabaries. In general, Kanji characters are visually more complex than Katakana or Hiragana characters, therefore, the location of Kanji characters can help to provide visual cues for where to fixate next in the text (Kambe, 1986 & Saida, 1993 in Kajii, Nazir and Osaka, 2001). The three types of Japanese characters are shown in Figure 8 below.
Figure 8. Examples of the Japanese word meaning ‘Japanese’ in three different writing systems: Kanji, Hiragana and Katakana (Kajii et al., 2001)

Studying reading in Japanese, Kajii, et al. (2001) asked eight subjects to read 20 Japanese short paragraphs consisting of 1096 words, including 711 Kanji characters, 140 Katakana characters, 1197 Hiragana characters, 7 Arabic numbers and 180 punctuation marks. The paragraphs were written using all three types of characters (as shown, for example, in Figure 9 below).

The results showed that the PVL for these Japanese subjects was the first character of the word (refer Figure 10), similar to English readers when reading unspaced English text (Rayner, et al., 1998). However, the pattern of probabilities of Japanese readers refixating a word during reading was opposite to that of European language readers (Rayner et al., 1998). In European languages the probability to refixate is lowest if the first fixation is at or near the PVL [left of the centre of the
word] (refer Section 2.5), but in Japanese the probability to refixate was highest if the first fixation was at the PVL (start of the word). This may be because the PVL for European languages is near the centre of the word, making the need to refixate less probable, whereas for Japanese the PVL was at the first character, making the probability of the need for refixation quite high, depending on word length. The average fixation duration for the Japanese readers reading normal unspaced Japanese text (275 msec.) was shorter than for English readers reading unspaced English text (378 msec.), but a little bit longer than English readers reading spaced text (253 msec.) in Rayner et al. (1998).

![Figure 10. Mean proportion of initial fixations (Kajii et al., 2001)](chart)

This study also examined the effect of character type on fixation probability to ascertain whether the Kanji characters provide visual cues that guide eye movements. Subjects were asked to read two- and three-character words consisting of pure Kanji, pure Hiragana, or mixed Kanji-Hiragana. It was found that the Kanji characters attracted significantly more fixations than the other two character types. Moreover, the preference for fixation at the beginning of words disappeared in pure Hiragana and the PVL effect for Japanese language was evident only in parts of words that contained
Kanji characters. This confirms the hypothesis that Kanji characters do indeed provide ‘visuo-lexical’ cues to word boundaries in Japanese. The findings of the Kajii et al. study support the hypothesis that the PVL for unspaced text is the first character of the word. However, as they did not investigate eye movement control of Japanese readers when reading spaced Japanese text, we still do not know whether spaces would facilitate reading Japanese or not.

Sainio, Hyönä, Bingushi and Bertram (2007) did manipulate spaces between words in Japanese. They recorded eye movements of native Japanese readers reading either Hiragana (syllabic) or mixed Kanji-Hiragana (ideographic-syllabic) texts in spaced and in unspaced conditions (refer Figure 11 for example of texts). They found that spaces between words facilitate both word identification and eye guidance only when the readers read Hiragana (syllabic) script as indicated by the shift of initial landing site, but not when the text contains Kanji script (refer Figure 12), and concluded that spaces between words serve as an effective segmentation cue only when Japanese readers read Hiragana. Presumably the Kanji characters serve as an effective segmentation cue in Kanji-Hiragana text, such that the additional space information is redundant (Kajii, et al., 2001).
Table 1.

<table>
<thead>
<tr>
<th>Text type</th>
<th>Spaced</th>
<th>Unspaced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hiragana</td>
<td>なりた くうこうの けいえいは</td>
<td>なりたくうこうのけいえいは</td>
</tr>
<tr>
<td></td>
<td>げんじょうでは おせじにも ゆうりようとは いえない。</td>
<td>げんじょうではおせじにもゆうりようとはいえない。</td>
</tr>
<tr>
<td>Kanji</td>
<td>成田 空港の 経営は 現状では</td>
<td>成田空港の経営は現状ではおさ</td>
</tr>
<tr>
<td>Hiragana</td>
<td>お北側にも 優良とは いえない。</td>
<td>お北側にも優良とはいえない。</td>
</tr>
<tr>
<td>English</td>
<td>The management of Narita airport can by no means said to be excellent</td>
<td></td>
</tr>
</tbody>
</table>

Figure 11. Sample of stimulus configuration (Sainio et al., 2007)

Figure 12. Mean proportion of initial fixations (Sainio et al., 2007)

The effect of spaces between words has also been investigated in reading Chinese. Bai, Yan, Zang, Liversedge and Rayner (2008) manipulated the presence/absence of spaces in four conditions: normal unspaced text, and then by spaces in three different ways either a) between each character; b) appropriately at word boundaries, or c) inappropriately between characters, thus creating non-words (refer Figure 13 for examples). The results showed that the average fixation duration for reading Chinese text with between-character spaced text (214 msec.) was shorter than for inter-word spaced text (227 msec.) which in turn was also shorter than for
reading normal unspaced texts (246 msec.). The probability of refixations was lowest when reading normal unspaced text. Surprisingly, the average fixation duration a non-word spaced condition (228 msec.) was even faster than the normal unspaced text condition (246 msec.). Moreover, there was no significant effect of spacing between words on the oculomotor control of Chinese readers. Unfortunately, this study did not compare landing site distribution across space conditions, so it is not yet known whether spaces have the same effects on landing site distribution in reading Chinese as in reading Japanese.

<table>
<thead>
<tr>
<th>1) Normal unspaced condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>科学技术的飞速发展给社会带来了巨大的变化。</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2) Single character spacing condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>科学技术的飞速发展给社会带来了巨大的变化。</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3) Word spacing condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>科学技术的飞速发展给社会带来了巨大的变化。</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4) Nonword spacing condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>科学技术的飞速发展给社会带来了巨大的变化。</td>
</tr>
</tbody>
</table>

**Figure 13. Example of Chinese stimuli from four spacing conditions (Bai et al., 2008)**

Turning to Thai, Kohsom and Gobet (1997) investigated the effects of adding spaces to Thai text by asking eight native Thai speakers, who speak English as their second language, to read Thai and English texts in eight conditions obtained by the factorial crossing of three factors: spaced versus unspaced, English or Thai script, and coherent or incoherent segmentations of the texts. The length of paragraphs and the number of words and characters within each paragraph were similar between the two languages. Participants were given a cue before each paragraph presented on the screen
to show whether the next paragraph would be spaced or unspaced. Participants reading aloud was recorded, and their reading speed and errors were measured. Examples of some paragraphs are shown in Figure 14 below.

**Figure 14. Examples of Thai texts used in the experiment (Kohsom et al., 1997)**

The results showed that Thai readers read meaningless text significantly more slowly than coherent text in both spacing conditions, but that their reading rate was faster when spaces were present. The effects of spacing was stronger in English, that is the better reading rate for spaced than unspaced text was greater in English than in Thai. The results also showed that the participants made less errors when reading spaced texts. In conclusion, Thai readers, although they are familiar with reading text without spaces, can use the information provided by spaces to improve their reading rate and accuracy. Therefore, spaces seem to be a universal cue to word boundaries across languages (Kohsom et al., 1997). However, this study examined only the effects of spaces on reading rate and accuracy, and did not provide any information on eye movement behaviour of Thai readers when reading Thai texts. So it is unknown whether Thai readers use different oculomotor patterns from European readers reading spaced text, as Japanese readers were found to do (Kajii et al., 2001).
Reilly, Radach and Luksaneeyanawin (2003) conducted an experiment to investigate eye movements in Thai readers when reading normal Thai unspaced texts. The participants, 23 native Thai Linguistics students at Chulalongkorn University, Bangkok, were asked to read the short story “Deaf Sim” (2,117 words long) (refer Figure 15). The average word length was 3.3 letters (3 letters are equal to 1 degree of visual angle). Their eye movements were recorded using the EyeLink II apparatus.

Reilly et al. found that average fixation duration of Thai readers reading unspaced text was 204 msec. This is even faster than English readers in Rayner et al, 1998, reading spaced text (229 msec.). The average gaze duration of Thai readers was 229 msec., their average progressive saccade size was 5.8 letters and their average regressive saccade size was 4.9 letters. Surprisingly, Reilly et al. found that the PVL of Thai readers when reading unspaced text was near the centre of the word, similar to the PVL of European readers (refer Figure 16). Therefore they concluded that word centre is the effective target for the eye during the execution of saccades in Thai (unspaced text) as well as in European languages (spaced text).
Since Thai does not use spaces between words, Reilly et al. (2003) conducted further studies to investigate what features Thai readers might use for specifying word boundaries and locating the centre of the word. They found that start and/or end characters seemed to be potential candidates that Thai readers use as visual segmentation cues. Frequencies of start and end characters for the text used in their experiment are shown in Table 16 below.

<table>
<thead>
<tr>
<th>Start Character</th>
<th>% (Frequency)</th>
<th>End Character</th>
<th>% (Frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ይ /e:/</td>
<td>17.3 (366)</td>
<td>ง /ŋ/</td>
<td>19.2 (408)</td>
</tr>
<tr>
<td>ฌ /n/</td>
<td>6.1 (120)</td>
<td>ง /ŋ/</td>
<td>15.5 (328)</td>
</tr>
<tr>
<td>ฎ /aj/</td>
<td>5.9 (115)</td>
<td>ง /w/</td>
<td>9.3 (197)</td>
</tr>
<tr>
<td>ฑ /thi:/</td>
<td>4.3 (91)</td>
<td>ง /k/</td>
<td>7.6 (162)</td>
</tr>
<tr>
<td>น /k/</td>
<td>3.8 (80)</td>
<td>ง /a/</td>
<td>5.1 (108)</td>
</tr>
<tr>
<td>ฎ /kh/</td>
<td>3.7 (78)</td>
<td>ง /thi:/</td>
<td>4.9 (108)</td>
</tr>
<tr>
<td>ณ /h/</td>
<td>3.5 (74)</td>
<td>ง /j/</td>
<td>4.5 (96)</td>
</tr>
<tr>
<td>ฎ /aj/</td>
<td>3.4 (73)</td>
<td>ง /m/</td>
<td>3.9 (82)</td>
</tr>
<tr>
<td>ฏ /ʔ/</td>
<td>3.4 (72)</td>
<td>ง /d/</td>
<td>3.4 (73)</td>
</tr>
<tr>
<td>ผ /ʨ/</td>
<td>2.8 (60)</td>
<td>ง /w/</td>
<td>3.3 (70)</td>
</tr>
</tbody>
</table>

Note: numbers in brackets represent the numbers of occurrences
Re-examination of the results in terms of character frequencies showed that the PVL was closer to the word centre when the fixated word ended with a high frequency final character than if the word ended with a low frequency final character (refer Figure 17).

![Landing site distributions on 6-8 letter words in Thai a function of word ending frequency]

**Figure 17.** Effects of final script frequency (Reilly et al., 2003)

The frequency data in Table 16 were calculated by Reilly et al. from the story ‘Deaf Sim’ used in their experiment. Similar to the Kajii et al. (2001) study with Japanese, the Reilly et al. study investigated only eye movements of Thai readers when reading normal unspaced text, so we still do not know whether spaces between words will have any effect on the eye movement pattern and eye movement control of Thai readers. We know from the results of Kohsom et al. (1997) that Thai readers can use information provided by spaces between words to improve their reading accuracy (faster reading speed, less errors). If the findings of Reilly et al. (2003) regarding the PVL of Thai readers is a robust effect; i.e., Thai readers’ PVL is always near the word centre, then spaces between words should not have any significant effect on reading eye movements of Thai readers. That is, spaces should simply facilitate reading rate
and accuracy by Thai readers, but should not play any important role in guiding eye movements as in European languages readers.

Most of the studies of eye movements in reading with and without spaces between words discussed in this section showed that there are differences in eye movement patterns and control when reading spaced and unspaced texts. Generally speaking, the readers’ PVL when reading spaced text is normally at or near the word centre, whereas the PVL of readers reading unspaced text is at the beginning of the word. Moreover, readers can read spaced texts faster than unspaced ones. However, the results of the one study of eye movements so far in Thai (Reilly et al., 2003) contradict these results; Thai readers read unspaced text very quickly and their PVL is near the centre of the word (Reilly et al., 2003). The questions of whether Thai readers really target the centre of the word during reading, and what factors or features of Thai language control eye movements in reading Thai and allow targeting of the word centre (when no visual boundaries are available) still require thorough investigation.

2.6. Eye Movements in Reading of Children

Reading, unlike speaking, requires specialized tuition so it is useful to investigate the development of reading skill, especially with regard to the issues raised above regarding PVL, OVP and reading spaced and unspaced texts, etc. Children must be taught to acquire written language, and the consistency of orthography of the language, as well as the initial age at which children start learning to read are important factors in determining whether and how well children develop reading skills. Developmental studies have shown that the rate of acquisition and success of reading development varies with nature of orthography being learned (Seymour, Aro & Erskine, 2003). Most children begin their formal schooling between the ages of 4 and 7
years, depending on the policy of the country, and, while the age of which children begin to learn to read has some effect on reading acquisition (Goswami et al., 2005), of more importance is the nature of the orthography (refer 1.2 for more details). However, most studies of children eye movements are with European languages. Here it will be interest to study both adult and child Thai readers. Some findings with European children’s reading are set out below.

Similar to the studies of eye movements in reading by adults, most eye movement studies of reading of children have investigated European children. There are certain known developmental trends in eye movements in reading European languages: as reading skill increases fixation duration, frequency of fixations, and frequency of regressions decrease, while saccade length increases (Buswell, 1922 in Rayner, 1998). Such trends have been found consistently despite differences in procedures, such as the manner in which eye movements were recorded. Eye movements of normal children when reading normal spaced texts have shown that their average reading speed is slower, number of fixations is higher and saccade size is smaller compared to adults. Children’s reading proficiency is also related to the level of transparency of the grapheme-phoneme correspondence rules in a particular language (Rayner, 1998).

McConkie, Zola, Grimes, Kerr, Bryant and Wolff (1991) investigated English children’s eye movements during reading. They found that even in their first year of reading children showed the same preferred landing location patterns as adults – their PVL was near the centre of the word. This study also showed that the primary difference between English language children and adults’ eye movements was the frequency of refixation before moving to another word: adults refixated 5-letter words
15% of the time, whereas first grade children refixated such words on 57% of occasions (Rayner, 1998).

Søvik, Arntzen and Samuelstuen (2000) conducted research to examine the relationship between four eye movement parameters (recognition span, average fixation duration, number of progressive saccades, and number of regressive saccades) and reading speed in child readers. Twenty 12-year-old Norwegian children were asked to read three passages both silently and aloud. These participants were grouped into Good and Poor readers according to the results of reading tests. Reading accuracy and reading speed were measured as well as the eye movement measures. The results showed that poor readers have shorter recognition span, longer average fixation duration, and a higher number of both progressive and regressive saccades. Regression analyses indicated significant relations between recognition span, average fixation duration, number of regressive saccades and speed in silent reading. In reading aloud, only average fixation duration and the number of regressive saccades were significantly related to reading speed. Therefore, they concluded that the average fixation duration and number of regressive saccades were the most efficient predictors for reading speed in children (both good and poor readers).

2.7. Issues in Eye Movements in Reading Thai Text

The eye movement studies in this thesis will focus on eye movement control in reading Thai text. This is because Thai language has several distinctive characteristics of interest for eye movements in reading.
Most importantly, Thai does not use spaces to specify word boundaries as in English or most other languages in the world, and so how Thai readers locate word boundaries and control saccadic eye movements is of great interest.

Second, the definition of ‘word’ in Thai is tricky. A compound word might change to become a whole sentence in a different context; for example, ‘คนขับรถ’ [khōn khàp rót] can refer to either a single word, the noun ‘driver’, or to a three-word sentence ‘Man driving car [lit]’ (‘A man is driving a car’). As a result of the spaces issue and the word definition issue, many Thai word strings are ambiguous and can be segmented into two different meaningful lexical readings, depending on the context, e.g. ‘ตาลม’ can be read as ‘ตาลม’ [tà:k lōm] (“exposed to wind”) or ‘ตาลม’ [tā: klōm] (“round eyes”). Moreover, since word segmentation is solely derived by scanning the text this might affect eye movement control of Thai readers if pseudo-spaces between words are inserted into the text.

Finally, studies have compared eye movements in silent reading and oral reading in European languages and have found that average fixation duration is longer and saccade size shorter with oral reading [275/225 msec; 6/8 letter] (Rayner, 1984 & 1998). It is of interest to study whether Thai readers have the same pattern as European readers in this regard.
CHAPTER THREE

EXPERIMENT 1: READING TIME STUDIES
3. EXPERIMENT 1: READING TIME STUDIES

As described in Section 1.3, Thai is written without spaces between words; there are no obvious visible cues to specify word boundaries during reading and the features of Thai orthography that allow Thai readers to perceive word boundaries are still questionable. However, an eye movements study on reading Thai conducted by Reilly and colleagues (2003) found the relative frequencies of occurrences of characters at word-start and word-end position seemed to be features that Thai adult readers use as a visual cues to guide their eye movements (refer to section 2.5.2). The higher frequency characters (especially at word-end) allowed readers to land their eyes closer to the PVL, the word-centre area. The reading time study in this chapter aimed to investigated how the relative frequency of start and end character frequencies might affect reading accuracy (in error rates and error types) of native Thai speakers, not only adults as in Reilly et al (2003), but also children. The results of this study are of interest in themselves and will also be used as a basis for further investigations of eye movements in reading Thai text by adults (Chapter 4) and children (Chapter 5). Here in this chapter, the effects of characters with high, medium or low relative frequency of occurrence in word-start and word-end positions on reading time and accuracy in Thai children and adult controls are reported.

There are further quite complex features of Thai orthography which may be related to higher linguistic processes such as (i) the letter-to-sound correspondences of Thai characters (Feedback Consistency Effects) at the start and end positions and (ii) the combination of initial consonant classes and tone realisation in Thai (Consonant Class and Tone Realisation Effects). These features were also investigated with children and adults using reading time as the dependent measure, and reading
accuracy. The results of the two reading time studies on these factors are reported in Appendix C, because their follow-up in eye movement studies is beyond the scope of this thesis. All the experiments contained in the main body of the thesis (reading time with children and adults (Chapter 3), eye movements with adults (Chapter 4) and eye movements with children (Chapter 5) include as a central concern, the relative frequency character in the start and end of words.

The main objective of this experiment is to investigate the effects character frequencies on reading ability of Thai children across age, with adult data as a reference point. It examined the effects of high, medium and low relative frequency of characters in the start and end position of target word on reading in both spaced and unspaced conditions. The dependent variable was reading ability of participants as measured by their reading speed, error rate, and error types.

3.1. Participants

There were a total of 100 participants altogether: 20 first graders \([\bar{x} = 6.45 \text{ year-old (Min = 5; Max = 7)}]\), 20 second graders \([\bar{x} = 7.8 \text{ year-old (Min = 7; Max = 9)}]\), 20 fifth graders \([\bar{x} = 10.8 \text{ year-old (Min = 10; Max = 12)}]\), 20 sixth graders \([\bar{x} = 12.1 \text{ year-old (Min = 11; Max = 13)}]\) (all recruited from the Thanombutra School, Bangkok, Thailand) and 20 adults \([\bar{x} = 29.2 \text{ year-old (Min = 19; Max = 45)}]\) (10 males and 10 females in each child and adult group).

There were two screening processes involved before the final sample was determined before the testing. The first was to group the participants into good or poor reading level (Reading Ability Screening Test), and the second to check the speech production ability of the selected participants (Shadowing Task).
3.1.1. **Reading Ability Screening Test**

In the reading test, there were different procedures for child and adult participants. For child participants, a preliminary reading examination was used to screen all students in first, second, fifth, and sixth grades at the school before selecting the participants for each group according to their reading score. The Reading Ability Test used in the screening process here was modified from some questions of former examination for the National Test in Thai language subject and lessons from Grade 1, 2, 5 and 6 text books. There were two sections in the Test; Section 1 comprised of 40 comprehensive questions with Graphotactic understanding (10 questions), Phonological understanding (10 questions), Semantics (10 questions), Vocabulary (10 questions) subtests and Section 2 was 20 Spelling questions for which children were asked to write the correct spelling of the words from the list of written pronunciation of each word that was given (refer Appendix B.1A to 1D for details of the test). There were two difficulty levels of the test, the easier version was used to test first and second graders; the more difficult version was used for fifth and sixth graders. On the basis of these tests, 10 Good and 10 Poor readers (5 girls and 5 boys in each) were selected for each grade. Good readers were considered to be those who had reading test scores above the 70th percentile, and Poor readers those who had reading test scores below the 30th percentile. Of a total of 223 children who were screened, there were 73 students who could be classified as Good readers, 76 students who could be classified as Poor readers, and 74 students who fell in between. From these five girl and five boy good readers and five girl and five boy poor readers in each grade were selected.

As it was difficult to conduct a similar examination for adult participants; the highest education level of participants was used instead as the criterion by which to
group them into a Hi-Ed (Tertiary Education) group, approximately age-equivalent to the child Good readers or P/Mid-Ed (Secondary Education) which is approximately age-equivalent to Poor readers. Hi-Ed readers are those adults whose final education level is higher than Grade 9 and P/Mid-Ed readers are those with Grade 9 or lower education level.

3.1.2. Speech Production Screening Test

All the selected participants were tested with three lists of items, one of words, one of non-words and a third of sentences, which cover all the sounds in the Thai phonological and tonological system, in a shadowing (word and sentence repetition) task to check their pronunciation ability (refer Appendix B.2 for the lists of word, non-word and sentences used in this task). This was to obviate attributing any errors to inefficient speech production performance by the participants. The results of this shadowing task showed that none of the 100 participants had any problems pronouncing any of the sounds in the Thai language.

3.2. Experiment 1: Relative Frequency of Initial and Final Characters Effects

Reilly et al. (2003) found that the relative frequency of characters at the start and end positions of words served as a visual cue in guiding eye movements in reading of Thai readers. This experiment follows on from that study and investigated the effects of the relative frequency of characters in the initial and/or final position on reading accuracy of Thai readers. The three independent variables in this experiment were (i) spaced versus unspaced sentences, (ii) the relative frequency of the start
characters in target words (high, medium or low) (iii) and the relative frequency of the end characters in target words (high, medium or low). Participants were presented with sentences and asked to read these aloud. The dependent variables were derived from recordings of these productions (refer to 3.2.2).

3.2.1. Materials

The stimuli were 54 test sentences and 20 practice sentences. The 54 test sentences consisted of the factorial combination of 2 writing styles (spaced/unspaced) x 3 start character frequencies (high/medium/low) x 3 end character frequencies (high/medium/low) x 3 exemplars (presented in 3 separate blocks). The relative frequency of start and end characters was calculated from the two million word database of Thai language created by the Centre for Research in Speech and Language Processing (CRSLP), Chulalongkorn University, Bangkok, Thailand. There were three levels of relative frequency of characters in both the start and end position. In this study, the characters with the relative frequency equal to or above the sixty-fifth percentile are classed as High frequency group; the relative frequency of characters that fall between the sixty-fifth and thirty-fifth percentile are classed as Medium frequency while the characters that have relative frequency below the thirty-fifth percentile belong to the Low frequency group (refer details in Table 17 and Table 18).
Table 17.  **Relative frequency of characters in the initial position**

Those characters above the black line fall in the High Frequency class; those between the black and grey line are in the Medium Frequency class; and those below the grey line are in the Low Frequency class.

<table>
<thead>
<tr>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
</tr>
</thead>
<tbody>
<tr>
<td>ถ</td>
<td>213038</td>
<td>12.1534</td>
<td>ฦ</td>
<td>22001</td>
<td>1.2551</td>
</tr>
<tr>
<td>ก</td>
<td>125298</td>
<td>7.148</td>
<td>ง</td>
<td>19811</td>
<td>1.1302</td>
</tr>
<tr>
<td>ฑ</td>
<td>105756</td>
<td>6.0332</td>
<td>ฃ</td>
<td>11491</td>
<td>0.6555</td>
</tr>
<tr>
<td>น</td>
<td>92368</td>
<td>5.2694</td>
<td>ฉ</td>
<td>8168</td>
<td>0.466</td>
</tr>
<tr>
<td>อ</td>
<td>89901</td>
<td>5.1287</td>
<td>ช</td>
<td>7975</td>
<td>0.455</td>
</tr>
<tr>
<td>ก</td>
<td>78336</td>
<td>4.4689</td>
<td>ง</td>
<td>6024</td>
<td>0.3437</td>
</tr>
<tr>
<td>จ</td>
<td>75273</td>
<td>4.2942</td>
<td>ฑ</td>
<td>4807</td>
<td>0.2742</td>
</tr>
<tr>
<td>สง</td>
<td>75098</td>
<td>4.2842</td>
<td>ง</td>
<td>4261</td>
<td>0.2431</td>
</tr>
<tr>
<td>ม</td>
<td>74538</td>
<td>4.2522</td>
<td>ฅ</td>
<td>4210</td>
<td>0.2402</td>
</tr>
<tr>
<td>ฉ</td>
<td>71094</td>
<td>4.0558</td>
<td>พ</td>
<td>2876</td>
<td>0.1641</td>
</tr>
<tr>
<td>ง</td>
<td>67004</td>
<td>3.8224</td>
<td>ฏ</td>
<td>1061</td>
<td>0.0605</td>
</tr>
<tr>
<td>น</td>
<td>64520</td>
<td>3.6807</td>
<td>ฏ</td>
<td>883</td>
<td>0.0504</td>
</tr>
<tr>
<td>ค</td>
<td>62273</td>
<td>3.5525</td>
<td>ฌ</td>
<td>801</td>
<td>0.0457</td>
</tr>
<tr>
<td>ฅ</td>
<td>60440</td>
<td>3.448</td>
<td>ฏ</td>
<td>643</td>
<td>0.0367</td>
</tr>
<tr>
<td>ฅ</td>
<td>59746</td>
<td>3.4084</td>
<td>ฎ</td>
<td>525</td>
<td>0.03</td>
</tr>
<tr>
<td>ฅ</td>
<td>48781</td>
<td>2.7829</td>
<td>ฎ</td>
<td>450</td>
<td>0.0257</td>
</tr>
<tr>
<td>ฅ</td>
<td>45536</td>
<td>2.5977</td>
<td>ฏ</td>
<td>33</td>
<td>0.0019</td>
</tr>
<tr>
<td>ฅ</td>
<td>43055</td>
<td>2.4562</td>
<td>ฎ</td>
<td>15</td>
<td>0.0009</td>
</tr>
<tr>
<td>ฅ</td>
<td>42441</td>
<td>2.4212</td>
<td>ฎ</td>
<td>8</td>
<td>0.0005</td>
</tr>
<tr>
<td>ฅ</td>
<td>31721</td>
<td>1.8096</td>
<td>ฎ</td>
<td>6</td>
<td>0.0003</td>
</tr>
<tr>
<td>ฅ</td>
<td>29579</td>
<td>1.6874</td>
<td>ฅ</td>
<td>5</td>
<td>0.0003</td>
</tr>
<tr>
<td>ฅ</td>
<td>28911</td>
<td>1.6493</td>
<td>ฅ</td>
<td>3</td>
<td>0.0002</td>
</tr>
<tr>
<td>ฅ</td>
<td>24559</td>
<td>1.401</td>
<td>ฅ</td>
<td>3</td>
<td>0.0002</td>
</tr>
<tr>
<td>ฅ</td>
<td>23967</td>
<td>1.3673</td>
<td>ฅ</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>ฅ</td>
<td>23619</td>
<td>1.3474</td>
<td>ฅ</td>
<td>1</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Table 18. **Relative frequency of characters in the final position**

Those characters above the black line fall in the High Frequency class; those between the black and grey line are in the Medium Frequency class; and those below the grey line are in the Low Frequency class.

<table>
<thead>
<tr>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
</tr>
</thead>
<tbody>
<tr>
<td>ง</td>
<td>189460</td>
<td>15.0436</td>
<td>ค</td>
<td>4215</td>
<td>0.3347</td>
</tr>
<tr>
<td>บ</td>
<td>165776</td>
<td>13.1631</td>
<td>ง</td>
<td>4166</td>
<td>0.3308</td>
</tr>
<tr>
<td>ำ</td>
<td>107099</td>
<td>8.5039</td>
<td>ญ</td>
<td>4157</td>
<td>0.3301</td>
</tr>
<tr>
<td>ง</td>
<td>77743</td>
<td>6.173</td>
<td>เร</td>
<td>3569</td>
<td>0.2834</td>
</tr>
<tr>
<td>ง</td>
<td>77288</td>
<td>6.1369</td>
<td>ช</td>
<td>2602</td>
<td>0.2066</td>
</tr>
<tr>
<td>ง</td>
<td>62490</td>
<td>4.9619</td>
<td>ฉ</td>
<td>1810</td>
<td>0.1437</td>
</tr>
<tr>
<td>ง</td>
<td>56122</td>
<td>4.4562</td>
<td>;charset</td>
<td>1796</td>
<td>0.1426</td>
</tr>
<tr>
<td>ง</td>
<td>55284</td>
<td>4.3897</td>
<td>คะ</td>
<td>1466</td>
<td>0.1164</td>
</tr>
<tr>
<td>ง</td>
<td>51049</td>
<td>4.0534</td>
<td>ชำ</td>
<td>1263</td>
<td>0.1003</td>
</tr>
<tr>
<td>ง</td>
<td>44547</td>
<td>3.5372</td>
<td>ช</td>
<td>1182</td>
<td>0.0939</td>
</tr>
<tr>
<td>ง</td>
<td>42348</td>
<td>3.3625</td>
<td>ฅ</td>
<td>953</td>
<td>0.0757</td>
</tr>
<tr>
<td>ง</td>
<td>40184</td>
<td>3.1907</td>
<td>ฑ</td>
<td>676</td>
<td>0.0537</td>
</tr>
<tr>
<td>ง</td>
<td>35689</td>
<td>2.8338</td>
<td>ฑ</td>
<td>468</td>
<td>0.0372</td>
</tr>
<tr>
<td>ง</td>
<td>28682</td>
<td>2.2774</td>
<td>ฑ</td>
<td>456</td>
<td>0.0362</td>
</tr>
<tr>
<td>ง</td>
<td>27347</td>
<td>2.1714</td>
<td>ฅ</td>
<td>232</td>
<td>0.0184</td>
</tr>
<tr>
<td>ง</td>
<td>25080</td>
<td>1.9914</td>
<td>ฆ</td>
<td>227</td>
<td>0.018</td>
</tr>
<tr>
<td>ง</td>
<td>21953</td>
<td>1.7431</td>
<td>ฅ</td>
<td>146</td>
<td>0.0116</td>
</tr>
<tr>
<td>ง</td>
<td>19382</td>
<td>1.539</td>
<td>ฅ</td>
<td>84</td>
<td>0.0067</td>
</tr>
<tr>
<td>ง</td>
<td>14072</td>
<td>1.1174</td>
<td>ฅ</td>
<td>67</td>
<td>0.0053</td>
</tr>
<tr>
<td>ง</td>
<td>13546</td>
<td>1.0756</td>
<td>ฅ</td>
<td>43</td>
<td>0.0034</td>
</tr>
<tr>
<td>ง</td>
<td>10901</td>
<td>0.8656</td>
<td>ฅ</td>
<td>40</td>
<td>0.0032</td>
</tr>
<tr>
<td>ง</td>
<td>9826</td>
<td>0.7802</td>
<td>ฅ</td>
<td>40</td>
<td>0.0032</td>
</tr>
<tr>
<td>ง</td>
<td>8412</td>
<td>0.6679</td>
<td>ฅ</td>
<td>30</td>
<td>0.0024</td>
</tr>
<tr>
<td>ง</td>
<td>7802</td>
<td>0.6195</td>
<td>ฅ</td>
<td>13</td>
<td>0.001</td>
</tr>
<tr>
<td>ง</td>
<td>7467</td>
<td>0.5929</td>
<td>ฅ</td>
<td>13</td>
<td>0.001</td>
</tr>
<tr>
<td>ง</td>
<td>7300</td>
<td>0.5796</td>
<td>ฅ</td>
<td>6</td>
<td>0.0005</td>
</tr>
<tr>
<td>ง</td>
<td>7271</td>
<td>0.5773</td>
<td>ฅ</td>
<td>5</td>
<td>0.0004</td>
</tr>
<tr>
<td>ง</td>
<td>6657</td>
<td>0.5286</td>
<td>ฅ</td>
<td>4</td>
<td>0.0003</td>
</tr>
<tr>
<td>ง</td>
<td>4385</td>
<td>0.3482</td>
<td>ฅ</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>ง</td>
<td>4226</td>
<td>0.3356</td>
<td>ฅ</td>
<td>1</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

(Upper and Lower vowel characters, tone characters and diacritics are presented with null character - ◌)
From Table 18, it is notable that, unlike characters in the start position, some upper and lower vowel characters, tone characters, diacritic scripts as well as punctuation symbols are in the list. However, only characters that occur in the main horizontal line will be used as end position characters in this experiment.

The target word was placed in or near the middle of the sentences, all of which were grammatical and meaningful sentences. The difficulty levels of target words and sentence frames were designed to be age appropriate to avoid ceiling effects for the older and floor effects for the younger participants. Thus there were two sets of stimuli: one for first and second grade students and another for fifth and sixth graders, and adults. There were nine types of target word conditions arrived at by factorial combination of the relative frequency (high, medium, low) of the start and end characters as shown in Table 19 and Table 20 below.

Table 19. Examples of sentences with high, medium or low frequency initial and final character using in Experiment 1 (Unspaced condition)

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Start Character</th>
<th>End Character</th>
<th>Mean Word Length</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>High</td>
<td>High</td>
<td>6.5</td>
<td>สุดาพาแก่นมาท่ศาลา</td>
</tr>
<tr>
<td>HM</td>
<td>High</td>
<td>Medium</td>
<td>6.7</td>
<td>มาณีรำมาภันน์</td>
</tr>
<tr>
<td>HL</td>
<td>High</td>
<td>Low</td>
<td>6.2</td>
<td>เอกทะสมุบกียน์</td>
</tr>
<tr>
<td>MH</td>
<td>Medium</td>
<td>High</td>
<td>4.5</td>
<td>วินัยถูกเพื่อนขับรถด่อยเวลา</td>
</tr>
<tr>
<td>MM</td>
<td>Medium</td>
<td>Medium</td>
<td>4.5</td>
<td>กนกถูกเพื่อนโทรศัพท์เวลา</td>
</tr>
<tr>
<td>ML</td>
<td>Medium</td>
<td>Low</td>
<td>4.2</td>
<td>ศุมลถูกเพื่อนขับรถด่อยเวลา</td>
</tr>
<tr>
<td>LH</td>
<td>Low</td>
<td>High</td>
<td>4.2</td>
<td>คุณครสอนคำว่าคำว่าคำว่าในชั้นเรียน</td>
</tr>
<tr>
<td>LM</td>
<td>Low</td>
<td>Medium</td>
<td>3.7</td>
<td>นักเรียนคำว่าคำว่าคำว่าในชั้นเรียน</td>
</tr>
<tr>
<td>LL</td>
<td>Low</td>
<td>Low</td>
<td>4.2</td>
<td>ภูมิตอานคำว่าคำว่าในพจนานุกรม</td>
</tr>
</tbody>
</table>
Table 20. Examples of sentences with high, medium or low frequency initial and final character using in Experiment 1 (Spaced condition)

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Start Character</th>
<th>End Character</th>
<th>Mean Word Length</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>High</td>
<td>High</td>
<td>6.5</td>
<td>สุดา พา แม่ หนี ที่ ศาลา</td>
</tr>
<tr>
<td>HM</td>
<td>High</td>
<td>Medium</td>
<td>6.7</td>
<td>แม่น พา ราช หนี ที่ บ้าน</td>
</tr>
<tr>
<td>HL</td>
<td>High</td>
<td>Low</td>
<td>6.2</td>
<td>เอก พา เมฆ หนี ที่ บึงน้ำ</td>
</tr>
<tr>
<td>MH</td>
<td>Medium</td>
<td>High</td>
<td>4.5</td>
<td>วินัย ถูก เพื่อน ข้อความ ตลอด เวลา</td>
</tr>
<tr>
<td>MM</td>
<td>Medium</td>
<td>Medium</td>
<td>4.5</td>
<td>นายก ถูก เพื่อน โทร ตลอด เวลา</td>
</tr>
<tr>
<td>ML</td>
<td>Medium</td>
<td>Low</td>
<td>4.2</td>
<td>ศูนย์ ถูก เพื่อน ข้อขี้ ตลอด เวลา</td>
</tr>
<tr>
<td>LH</td>
<td>Low</td>
<td>High</td>
<td>4.2</td>
<td>ศูนย์ สอน คี่ว่า ยาน ปัน ท้องเรียน</td>
</tr>
<tr>
<td>LM</td>
<td>Low</td>
<td>Medium</td>
<td>3.7</td>
<td>นักเรียน อ่าน คี่ว่า ฐาน ปัน ข้อเรียน</td>
</tr>
<tr>
<td>LL</td>
<td>Low</td>
<td>Low</td>
<td>4.2</td>
<td>ภาษณ์ี หา คี่ว่า anggan ปัน พจนานุกรม</td>
</tr>
</tbody>
</table>

3.2.2. Apparatus

Stimulus sentences were presented stationary on the centre of a 17-in. NEC monitor with 1,024 x 768 pixel resolution and 60 Hz refresh rate approximately 60 centimetres from the participants in Cordia New font size 22 one sentence at a time. Participants’ reading aloud was recorded via the program Adobe Audition 1.5 and from this, reading time, error frequency and error type were derived. Using PRAAT\(^3\) program version 5.1.44, acoustic *.wav files were segmented in order to derive the actual reading time of the target word from the starting point, excluding the silent after the end of the proceeding word if any, to the end point of the target word as shown in Figure 18.

\(^3\) Paul Boersma and David Weenink (www.praat.org)
3.2.3. Procedure

Each participant was tested individually in a room with minimal noise interference in the school for child participants and in a sound-attenuated room at CRSLP, Chulalongkorn University for adult participants. There were 20 practice trials (10 spaced and 10 unspaced sentences presented separately; starting with spaced) before the testing session involving 54 test trials. Of these 20 practice sentences, there were two blocks of nine (3 start character frequencies x 3 end character frequencies), one block of spaced first then a block of unspaced text and two other practice sentences acting as decoy or warm-up trials at the start of the Test session. Participants read all sentences aloud and there was no response time limit.

As can be seen in Table 19 and Table 20, the sentence frame was fixed within each of the three blocks of nine trials such that within the three sentences with High start character frequency and the three sentences with Medium and within these with Low start character frequency, the single word that preceded and followed the target
word – either the same words or at least words having the same end character (preceding word) or start character (following word) to avoid the effects of the frequency of that particular character on the start and end character of the target words. Thus each of the three sets of three sentences trials starting with the same start character frequency had similar meanings. The stimuli were arranged in a rolling design so that each member (high, medium or low end character) of the same start-character defined sentence frame was not presented in the same block.

For the 54 test sentences a similar control was instigated, thus there were 18 different sentence frames⁴, based on start frequency of groups of target words. Of these six were frames for High start character target words, with three exemplars for spaced and three exemplars for unspaced sentences. Similarly, there were six frames for Medium start character target words and other six for Low start character target words. Each frame was used for three sentences that varied in terms of the frequency of the end characters of the target. So for example, for High start character target words, the same frame was used for High, Medium, and Low end characters. (Refer Table 19 and Table 20 for example sentences and Appendix B.3A and B.3B for the design of these sentences along with a full list of the Thai sentences that were used).

The test stimuli were arranged in a rolling design so that each member (high, medium or low end character) of the same sentence frame was not presented in the same block to avoid giving the participants specific information about the experimental

⁴ Each frame had the same meaning across frequency of end character variations, although the actual words may have changed slightly.
manipulations. After reading each stimulus sentence, participants responded by hitting a key (SPACEBAR) in order for the next stimulus to be presented on the screen.

### 3.2.4. Hypotheses

Since the Thai language is normally written without spaces between words, it is unclear how Thai readers specify word boundaries. On the basis of the results of Reilly and colleagues (2003) it is possible that start and end characters are keys that aid word recognition process in reading Thai. Therefore, it is possible to hypothesise that:

A) High frequency start and/or end characters may assist younger children to **specify word boundaries** when reading unspaced sentences which would in turn lead to:

1. **faster word recognition**, as measured here by reading time, for words with higher frequency start and/or end characters, and
2. **less errors** reading target words with higher frequency start and/or end characters

B) Regarding the role of reading instruction in Thai and participants’ age, it is possible to hypothesise that:

3. Given that young children up until the end of 1\textsuperscript{st} grade are initially learn to read spaced text, 1\textsuperscript{st} and 2\textsuperscript{nd} grade participants should have much slower word recognition and more errors reading **unspaced** than **spaced** texts and this difference in performance should be less pronounced for 5\textsuperscript{th} and 6\textsuperscript{th} grader and adult participants

4. **Good** reader groups should have less of a difference in performance reading unspaced than spaced texts than would **poor** readers (refer Figure 19).
C) More generally in terms of overall development, it can be hypothesised that:

5. As the age of the participants increases, their reading skills should increase (they should make less errors and their reading time should be faster)

6. Good readers’ performance should be better than that of poor readers (both in accuracy and reading time)

### 3.3. Results

The results of this experiment are presented in two sections; the reading time and error results.

#### Reading Time

The overall results for reading time (RT) are shown in Table 21, broken down by age group and number of syllables and Table 22, broken down by age group and reading level. Figure 20 also shows the reading time results of each age group by the number of syllables. The overall means of reading time show that the participants read
short words (with less syllables) faster than longer words. This is a reassuring effect showing that reading time is sensitive to word length and thus a valid response measure. Table 21 and Table 22 also show that reading time decreases over age and is inversely related to reading ability effects that further attest to the validity of the reading time measure and Table 23 shows that participants read words faster when spaces between words were available.

**Table 21. Overall mean reading time of participants in Experiment 1**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Number of Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td></td>
<td>561.42</td>
<td>659.99</td>
<td>729.88</td>
</tr>
<tr>
<td>Second</td>
<td></td>
<td>540.12</td>
<td>626.80</td>
<td>686.00</td>
</tr>
<tr>
<td>Fifth</td>
<td></td>
<td>475.20</td>
<td>527.03</td>
<td>562.32</td>
</tr>
<tr>
<td>Sixth</td>
<td></td>
<td>431.09</td>
<td>473.01</td>
<td>502.23</td>
</tr>
<tr>
<td>Adult</td>
<td></td>
<td>385.85</td>
<td>436.21</td>
<td>459.72</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>478.74</td>
<td>544.61</td>
<td>588.03</td>
</tr>
</tbody>
</table>

**Table 22. Mean reading time of participants with different reading ability**

<table>
<thead>
<tr>
<th>Grade</th>
<th>First</th>
<th>Second</th>
<th>Fifth</th>
<th>Sixth</th>
<th>Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good readers</td>
<td>572.90</td>
<td>519.92</td>
<td>503.31</td>
<td>467.06</td>
<td>415.08</td>
</tr>
<tr>
<td>Poor readers</td>
<td>727.96</td>
<td>715.36</td>
<td>539.73</td>
<td>491.50</td>
<td>621.44</td>
</tr>
</tbody>
</table>

**Table 23. Mean reading time of participants with different reading ability**

<table>
<thead>
<tr>
<th>Grade</th>
<th>First</th>
<th>Second</th>
<th>Fifth</th>
<th>Sixth</th>
<th>Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spaced</td>
<td>605.44</td>
<td>570.59</td>
<td>529.87</td>
<td>472.77</td>
<td>437.43</td>
</tr>
<tr>
<td>Unspaced</td>
<td>674.83</td>
<td>647.86</td>
<td>530.36</td>
<td>478.62</td>
<td>434.99</td>
</tr>
</tbody>
</table>
The data were analysed using the lmer program for linear mixed-effects modelling (lme4 package; Bates & Starkar, 2006) in the R system for statistical computing (R Development Core Team, 2006). Mixed effects analysis has three main advantages over more traditional analyses of variance; it is more generalisable results, it is more flexible and it makes use of a modelling approach, as set out below.

The greater generalisability of data analysis using mixed-effect models derives from the fact multiple random effects can be used to test for any given fixed effects, at the same time in the same analysis. For example, in an experiment in which multiple subjects and items are used mixed-effect models can be used to test for main effects.
and interactions of the particular variables of interest using both subjects and items as random factors in the same analysis\(^5\). Related to this point the analysis is more flexible, for the distinction between fixed and random factors is less clear-cut in mixed model analyses. This is because, theoretically, what was a fixed factor, e.g. position of a word in a sentence, could be used as a random factor if one wanted to test whether the results of a particular fixed factor (e.g. character frequency) was robust across sentence position. Finally, mixed effects analyses are conducted within a modelling framework, more akin to regression analyses than fixed analysis of variance. Thus a particular model, e.g. using particular variables as fixed effect factors and selected interactions between these, and using particular variables as random effect factors can be tested. Further models may then also be tested, e.g. using more or fewer interactions of fixed effects, changing a variable from being a fixed to a random effect. Differences between models can be tested statistically to arrive at the model that best describes the data. It should be noted, however, that while differences between models can be tested statistically, there is still debate about appropriate p levels for the significance of particular effects within particular models (Baayen, 2008). Here we will follow the usual practice (Baayen, 2008) of reporting t-values and taking to be significant any effect with an absolute t-value $\geq 2.00$.

For these reasons mixed-effect models were considered to be the most suitable statistical analysis method for the current study with the fixed effects of age

\(^5\) This contrasts with the way in which the effect of particular fixed effects over subjects and items has been treated over the past 30 years or so. Foster and Dickinson (1976) proposed separate analyses across subjects (resulting in an F1 F-value) and across items (F2 F-value) and then combining these in a minF value. Apart from the fact that the random effects are not considered simultaneously, a major practical drawback of this method is that experimenters have tended to report F1 and F2 separately even when only one of these is significant and especially when min F itself is not significant, leaving open the increased possibility of Type I errors.
group/Grade (1st, 2nd, 5th, 6th, and adults), reading ability/Ability (good/poor reader), Spacing (spaced/unspaced), start character frequency/Initial (high/medium/low), end character frequency/Final (high/medium/low), word frequency/Word (high/low) and number of syllables/Syllable (1/2/3), which was used as a covariate in the model to control for the fact that the target words have different length. As syllable was treated as a covariate it was tested for the main effects only, not any interaction effects because there was no theoretical reason to expect interactions. The random effects in the analysis were words and subjects.

In this and all subsequent linear mixed-effects modelling analyses any t-value equal to or greater than 2.0 was taken to indicate a significant difference (Baayen, 2008). Analyses were conducted by first entering all main effects and 2-way interaction for all factors except word frequency and syllable and then successively removing any main and interaction effects that were not significant. To determine which solution best fit the data, the different resulting models were testing against one another with that model yielding significantly better results being the one that is reported in the results section.

The analysis showed that the main effects of Grade, Ability, Spacing, and Syllable were significant. As shown in Table 24 below, there was a significant effect of number of syllables on reading time of the participants (t = 13.42), a reassuring result. The differences between age groups are significant in all comparisons (between 1st vs. 2nd grade (t = 10.47); 2nd vs. 5th grade (t = 19.41); 5th vs. 6th grade (t = 18.43); and 6th grade vs. Adults (t = 13.37)). Thus there was, as expected, a general improvement in reading speed over ages (see Figure 21). Good readers’ reading time was also significantly shorter than that of poor readers’ (t = -48.43) and participants
read target words in spaced text significantly faster than in unspaced text \( t = -3.97 \).

There were no significant main effects of frequencies, neither start character, end character or word on reading time of the participants all \( ts < 2 \).

Table 24. **Mixed – effect models results of participants’ reading time**

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>437.5925</td>
<td>8.5913</td>
<td>50.93</td>
</tr>
<tr>
<td>Syllable</td>
<td>49.2654</td>
<td>3.6703</td>
<td>13.42</td>
</tr>
<tr>
<td>Grade.1st&lt;&gt;2nd</td>
<td>30.9185</td>
<td>2.9523</td>
<td>10.47</td>
</tr>
<tr>
<td>Grade.2nd&lt;&gt;5th</td>
<td>94.2017</td>
<td>4.8522</td>
<td>19.41</td>
</tr>
<tr>
<td>Grade.5th&lt;&gt;6th</td>
<td>54.4204</td>
<td>2.9523</td>
<td>18.43</td>
</tr>
<tr>
<td>Grade.6th&lt;&gt;Adult</td>
<td>39.4824</td>
<td>2.9523</td>
<td>13.37</td>
</tr>
<tr>
<td>Ability.G&lt;&gt;P</td>
<td>-96.0409</td>
<td>1.983</td>
<td>-48.43</td>
</tr>
<tr>
<td>Spacing.S&lt;&gt;U</td>
<td>-17.5175</td>
<td>4.412</td>
<td>-3.97</td>
</tr>
<tr>
<td>Grade.1st&lt;&gt;2nd:Ability.G&lt;&gt;P</td>
<td>35.4074</td>
<td>5.9047</td>
<td>6</td>
</tr>
<tr>
<td>Grade.2nd&lt;&gt;5th:Ability.G&lt;&gt;P</td>
<td>-159.378</td>
<td>5.9077</td>
<td>-26.98</td>
</tr>
<tr>
<td>Grade.6th&lt;&gt;Adult:Ability.G&lt;&gt;P</td>
<td>-18.5352</td>
<td>5.9047</td>
<td>-3.14</td>
</tr>
<tr>
<td>Grade.1st&lt;&gt;2nd:Spacing.S&lt;&gt;U</td>
<td>7.8815</td>
<td>2.751</td>
<td>2.86</td>
</tr>
<tr>
<td>Grade.2nd&lt;&gt;5th:Spacing.S&lt;&gt;U</td>
<td>-54.7393</td>
<td>7.9683</td>
<td>-6.87</td>
</tr>
<tr>
<td>Grade.6th&lt;&gt;Adult:Spacing.S&lt;&gt;U</td>
<td>-8.2907</td>
<td>2.751</td>
<td>-3.01</td>
</tr>
<tr>
<td>Ability.G&lt;&gt;P:Spacing.S&lt;&gt;U</td>
<td>74.5364</td>
<td>1.7433</td>
<td>42.76</td>
</tr>
<tr>
<td>Initial:Ability.G&lt;&gt;P</td>
<td>0.519</td>
<td>0.2319</td>
<td>2.24</td>
</tr>
<tr>
<td>Final:Spacing.S&lt;&gt;U</td>
<td>-2.671</td>
<td>1.0809</td>
<td>-2.47</td>
</tr>
</tbody>
</table>
Figure 21. Reading time as a function of age group

Turning to the interaction effects, the mixed-effects models results showed significant interaction effects of age groups by ability and age groups by spacing. The interaction effects were significant in contrasts of all age groups except between grade 5 and grade 6 participants in both factors with the stronger interaction effects being between 2nd and 5th graders (between 1st and 2nd grade * reading ability (t = 6); 2nd and 5th grade * reading ability (t = -26.98); 6th grade and adults * reading ability (t = -3.14); 1st and 2nd grade * spacing (t = 2.86); 2nd and 5th grade * spacing (t = -6.87); and 6th grade and adults * spacing (t = -3.01)). As shown in Figure 22, the differences between reading times of good and poor readers was greater for younger child participants than the older participant groups. Spaces played important role in first and second graders by allowing the participants to read the words faster while there were
less effects on the other age groups (see Figure 23). There was also a strong interaction
effects of reading ability and spacing ($t = 42.76$); as can be seen in Figure 24 good
readers’ reading time when reading target words in spaced and unspaced texts was
equivalent while spaces between words assist poor readers by allowing them to read
target words in spaced text faster.

![Graph showing reading time as a function of age group and reading ability](image)

**Figure 22.** Reading time as a function of age group and reading ability
Figure 23. Reading time as a function of age group and spacing

Figure 24. Reading time as a function of spacing and reading ability
Regarding the frequency of characters at the word-start and word-end positions, even though there was no significant main effect of these frequencies on the participants’ reading time, the interaction effects between character frequencies and reading ability and spacing conditions were significant. The results of mixed-effects models analysis showed a significant interaction effect of start character frequency by reading ability (t = 2.24). Figure 25 shows that the difference between reading time of good and poor readers was greater when the target word started with high relative frequency start character than either medium or low, in turn, start characters.

![Figure 25. Reading time as a function of reading ability and start character frequency](image)
The relative frequency of end characters only showed interaction effects with spacing conditions ($t = -2.47$). As can be seen in Figure 26, participants’ reading time for target words ending with high and medium frequency characters was lower than those ending with low frequency when reading spaced text. Conversely, participants’ reading time for the target words ending with low frequency characters was lowest in the unspaced condition.

![Reading Time](image)

**Figure 26.** Reading time as a function of spacing and end character frequencies

**Error Type**

Errors were judged by a phonetically-trained native Thai speaker. Errors were categorized into six types, as set out below.
1. Phonological Errors:

There were three types of phonological errors:

- **Tone errors**: participants pronounced the incorrect tone
- **Consonant / Vowel errors**: participants pronounced incorrect consonants or vowels that appeared not to be due to confusability via orthographic similarity (refer to Orthographical Errors below)
- **Sound omission errors**: participants did not pronounce particular sound(s) or even syllable(s)

2. Orthographical Errors:

- Orthographical errors: participants pronounced the incorrect consonant sound substituting a consonant of similar orthography, such as in the example pairs of consonant shown below:

  ![Example pairs of consonant](image)

3. Semantic Errors:

There were two types of semantic errors as follows.
Lexical substitution errors: participants uttered (substituted) a word consistent in meaning with the context of the preceding or following words or the sentence rather than the target word.

Incorrect segmentation errors: participants uttered a word that was consistent with a misplaced word boundary (especially in unspaced sentences) and so pronounced an incorrect word.

The results are tabulated in Table 25. It can be seen that the number of errors decreased over age and that good readers made less errors than poor readers in every grade. Since there were two levels of difficulty of target words (for Grade 1&2 and Grade 5&6), it was expected that Grade 1 and Grade 5 participants should have similar mean numbers of errors while Grade 2 should similar to Grade 6. However, the results show that, even though the target words for Grade 5 participants were more difficult, they made less errors than Grade 2 participants. In conclusion, there were generally more phonological errors than orthographic or semantic errors, although it is notable that adult good readers made no errors and adult poor readers made no orthographic errors. Within the phonological errors there were always more tone errors than consonant and vowel errors.
Table 25.  **Error types and mean numbers of errors of each participant group**

<table>
<thead>
<tr>
<th>Group</th>
<th>Error type</th>
<th>Good Readers</th>
<th>Poor Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade 1</td>
<td>Grade 2</td>
<td>Grade 5</td>
</tr>
<tr>
<td>Tone</td>
<td>17</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Consonant &amp; Vowel</td>
<td>11</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Sound Omission</td>
<td>8</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Orthography</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Lexical Substitution</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Incorrect Segmentation</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td><strong>51</strong></td>
<td><strong>38</strong></td>
<td><strong>34</strong></td>
</tr>
</tbody>
</table>

**3.4. Discussion**

The results of the reading time study in this chapter showed that reading time performance improved (participants made less errors and read faster) monotonically with age. Additionally and as expected, good reader groups always performed better than poor reader groups in every grade. First and second grade good readers can read unspaced texts, although the poor readers in these grades did have some problems reading unspaced text as shown by their increased reading time when no spaces between words were available. Child participants always read spaced texts faster than unspaced, but the difference was greater in 1\textsuperscript{st} and 2\textsuperscript{nd} grade. This means that all children can indeed read unspaced text, but that spaces facilitate reading especially for poor young readers.

Given that young children can read unspaced text, the question then is how this is done? The answer to this question is still unclear as the results of this study showed that there were no significant main effects of start and end character frequency on
participants reading time. However, the significant interaction effects results of start character frequency by reading ability and of end character frequency by spacing condition suggest some contribution of the character at word-start and word-end positions on reading of Thai participants. The fact that younger child participants can read unspaced text shows that Thai children appear to be able to locate word boundaries effectively and efficiently in the early years of reading instruction using character frequency, and that this may be in part due to the frequency of characters in particular positions.

With regard to the frequency of characters in the word-start position higher start character frequency significantly facilitates reading times and dose so better in good than poor readers. Thus good readers tend to be those who can use word-start character frequency significantly better than their poor reader counterparts to segment sentences into words.

Turning to the end character frequencies, the results showed that when there were visible word boundaries (in the spaced condition) participants can use the information of word-end character frequencies in the word recognition process resulting in shorter reading times for words ending with higher frequencies. However, this effect of being able to use high end-character frequency better in spaced than unspaced conditions is ameliorated in medium frequency end-word characters, and disappear completely in low frequency end-word characters.

The results of this relative frequency of character effects study suggest that both start and end character may play important roles in word recognition in reading Thai. The results of this study confirm the finding of Reilly et al. (2003) that the
relative frequency of start and end characters served as a visual word boundaries of Thai readers. The results of this reading time study here seem to give some preliminary indications of how eye movements in reading of Thai readers might develop over age. The facts that (i) there were more pronounced start frequency effects in good than poor readers, (ii) there were stronger effects of end frequency effects in spaced than unspaced conditions, and (iii) there were strong interaction effects between reading ability and spacing, and (iv) the effects of spaces between on reading time decreased over age, suggests a shift in reading strategy over age. It appears that there may be decreasing reliance on visible visual cues like spaces between words and increasing reliance on character frequency in the start and end position especially the start character frequency over age, which might signal a change in eye movements landing site distribution of the Thai readers. These age by character frequency effects were not significant here and this may be due to the relatively small interaction effects of start- and end-word frequencies in this reading time study; reading time may not be a very sensitive measure of character frequency effects and eye movement studies are required to investigate this further (see Chapter 4 and especially Chapter 5). Nevertheless, some speculations about the development of the use of start and end character frequency may be posited here.

For example, when reading unspaced text; Thai children might land their eyes in the initial area as a result of there being no visible word boundaries available and so their attention is captured by high frequency start characters. Over age, such a visual capture effect may be ameliorated and children may come to use the initial character to guide their eyes to a more optimal position – towards word centre, thus resulting in faster reading times.
Conversely, as for the adults spaces between words should simply facilitate their eye movements control and not cause any shift on landing site distribution. The spacing by end-word character frequency effect shows that end-word frequency can be used when there are spaces between words. It is then possible that as a function of age readers, and perhaps more so for good readers, may learn to use end-word character frequency to guide eye movements in reading. These are purely speculations based on the results of this reading time study and developmental eye movement studies are required to address the question of whether there are any changes in eye movement strategies over age. The results of such studies will be presented in Chapters 4 and 5.
CHAPTER FOUR

EXPERIMENT 2: ADULT READERS’ EYE MOVEMENTS IN READING THAI
4. EXPERIMENT 2: ADULT READERS’ EYE MOVEMENTS IN READING THAI

4.1. Introduction

Experiment 2 is a study of native Thai readers’ eye movements in reading Thai text. In this experiment, eye movements of Thai adults with higher or lower education levels was investigated as a function of character frequency (high, medium and low) in word-start and word-end position, as well as spacing between words. In addition, another factor, reading sentences silently or aloud was included as a further repeated measures factor. Thus the study employed a 2 [Hi-Ed\textsuperscript{6} Readers vs. P/Mid-Ed\textsuperscript{7} Readers] x (2 [spaced/unspaced] x 2 [aloud/silent] x 3 [high, medium, low character frequency in start position] x 3 [high, medium, low character frequency in end position]) design with repeated measures on all factors. The statistical analysis of eye movement data in this experiment used the mixed-effects model analysis method (R program version 2.10.1 with ‘lme4’\textsuperscript{8}, ‘reshape’\textsuperscript{9}, ‘sciplot’\textsuperscript{10} and ‘ggplot2’\textsuperscript{11} packages) to analyse the data.

The main objectives of the experiment are to investigate the low-level visual information of Thai orthography that might allow Thai readers to read such an alphabetic scriptio continua language fluently. More specifically the aims are to investigate what features might provide visual cues to guide eye movements of Thai readers, the location of the preferred viewing location (PVL) of Thai readers, as well

\textsuperscript{6} Hi-Ed refers to undergrad and post-graduate education
\textsuperscript{7} P/Mid-Ed refers to primary or middle School to the first year of high school education (Grade 6 to 9)
\textsuperscript{8} Developed by Bates, D. and Maechler, M. (2009)
\textsuperscript{9} Developed by Wickham, H. (2009)
\textsuperscript{10} Developed by Morales, M. (2010) with code developed by R Development Core Team and general advice from the R-help listserv community, especially Murdoch, D.
\textsuperscript{11} Developed by Wickham, H. (2010)
as whether or not the oculomotor control of Thai readers is different when reading silently versus reading aloud. The three dependent variables in this experiment are landing site location (first fixation location), first fixation duration, and gaze duration on the target word embedded in the centre of the sentences.

4.2. Method

4.2.1. Participants

There were 30 participants: 15 Hi-Ed Readers and 15 P/Mid-Ed Readers, all native Thai-speaking adults, and all current students or general staff at Chulalongkorn University, Bangkok. Educational level of participants was used as the criterion to group them. The lowest education level of participants in the P/Mid-Ed group (all general staffs) was Grade 6 and highest level Grade 9 [mean age was 30.3 years (range = 21 – 42 years)] while lowest education level of participants in Hi-Ed Readers (all students) was first year undergrad student and highest level was third year PhD candidature [mean age was 22.8 years (age = 18 – 36 years)].

4.2.2. Materials

The stimuli were 54 test sentences and 20 practice sentences. The 54 test sentences comprised the factorial combination of spaced/unspaced text x 3 start character frequencies (high, medium, low) x 3 end character frequencies (high, medium, low) x 3 repetitions with each repetition presented in a separate block of 18 trials. The relative frequency of start and end characters was calculated from the two million word database (CRSLP Corpus version 1) of Thai language created by the Centre for Research in Speech and Language Processing (CRSLP), Chulalongkorn University, Bangkok, Thailand. This database was the initial step in setting up a
national corpus of Thai words which was later expanded into a ten million word database (CRSLP Corpus version 2). This CRSLP Corpus 2 will be used to analyse eye movements of Thai readers reading spaced and unspaced texts in Analysis II and III (refer 4.2.2 and 4.2.3 for details). These two corpora were made available for this project as part of the CRSLP, Chulalongkorn University – MARCS Auditory Laboratories, University of Western Sydney – Department of Computer Science, National University of Ireland (Maynooth) collaboration to expand the knowledge of eye movements in reading Thai.

To determine the three levels of relative frequency of characters in start and end position here characters with relative frequency equal to or above the 65th percentile of frequency in Corpus 1 were classified as High frequency; characters that had relative frequencies falling between the 64th and 35th percentile were classified as Medium frequency, while characters with relative frequency below the 35th percentile belong to the Low frequency class. A full listing of characters, their number of occurrences (n) per the two million word corpus, and percentage of these numbers of occurrences are shown in Table 26 (start characters) and Table 27 (end characters). These are the same as Table 17 and 18 in Chapter 3, but are reproduced here for ease of reference.
Table 26. **Relative frequency of characters in the initial position**

Those characters above the black line fall in the High Frequency class; those between the black and grey line are in the Medium Frequency class; and those below the grey line are in the Low Frequency class.

<table>
<thead>
<tr>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
</tr>
</thead>
<tbody>
<tr>
<td>ต</td>
<td>213038</td>
<td>12.1534</td>
<td>ย</td>
<td>22001</td>
<td>1.2551</td>
</tr>
<tr>
<td>ไ</td>
<td>125298</td>
<td>7.148</td>
<td>ง</td>
<td>19811</td>
<td>1.1302</td>
</tr>
<tr>
<td>ท</td>
<td>105756</td>
<td>6.0332</td>
<td>ช</td>
<td>11491</td>
<td>0.6555</td>
</tr>
<tr>
<td>ณ</td>
<td>92368</td>
<td>5.2694</td>
<td>ร</td>
<td>8168</td>
<td>0.466</td>
</tr>
<tr>
<td>ร</td>
<td>89901</td>
<td>5.1287</td>
<td>ศ</td>
<td>7975</td>
<td>0.455</td>
</tr>
<tr>
<td>ต</td>
<td>78336</td>
<td>4.6489</td>
<td>อ</td>
<td>6024</td>
<td>0.3437</td>
</tr>
<tr>
<td>ิน</td>
<td>75273</td>
<td>4.2942</td>
<td>ม</td>
<td>4807</td>
<td>0.2742</td>
</tr>
<tr>
<td>ม</td>
<td>75098</td>
<td>4.2842</td>
<td>น</td>
<td>4261</td>
<td>0.2431</td>
</tr>
<tr>
<td>ย</td>
<td>74538</td>
<td>4.2522</td>
<td>บ</td>
<td>4210</td>
<td>0.2402</td>
</tr>
<tr>
<td>ช</td>
<td>71094</td>
<td>4.0558</td>
<td>ป</td>
<td>2876</td>
<td>0.1641</td>
</tr>
<tr>
<td>จ</td>
<td>67004</td>
<td>3.8224</td>
<td>ฉ</td>
<td>1061</td>
<td>0.0605</td>
</tr>
<tr>
<td>ฉ</td>
<td>64520</td>
<td>3.6807</td>
<td>ท</td>
<td>883</td>
<td>0.0504</td>
</tr>
<tr>
<td>ฑ</td>
<td>62273</td>
<td>3.5525</td>
<td>ฑ</td>
<td>801</td>
<td>0.0457</td>
</tr>
<tr>
<td>ฑ</td>
<td>60440</td>
<td>3.448</td>
<td>ธ</td>
<td>643</td>
<td>0.0367</td>
</tr>
<tr>
<td>ฑ</td>
<td>59746</td>
<td>3.4084</td>
<td>ฎ</td>
<td>525</td>
<td>0.03</td>
</tr>
<tr>
<td>ธ</td>
<td>48781</td>
<td>2.7829</td>
<td>ฏ</td>
<td>450</td>
<td>0.0257</td>
</tr>
<tr>
<td>ฏ</td>
<td>45536</td>
<td>2.5977</td>
<td>ฐ</td>
<td>33</td>
<td>0.0019</td>
</tr>
<tr>
<td>ฐ</td>
<td>43055</td>
<td>2.4562</td>
<td>ฑ</td>
<td>15</td>
<td>0.0009</td>
</tr>
<tr>
<td>ฑ</td>
<td>42441</td>
<td>2.4212</td>
<td>ฑ</td>
<td>8</td>
<td>0.0005</td>
</tr>
<tr>
<td>ฑ</td>
<td>31721</td>
<td>1.8096</td>
<td>ฑ</td>
<td>6</td>
<td>0.0003</td>
</tr>
<tr>
<td>ฑ</td>
<td>29579</td>
<td>1.6874</td>
<td>ฑ</td>
<td>5</td>
<td>0.0003</td>
</tr>
<tr>
<td>ฑ</td>
<td>28911</td>
<td>1.6493</td>
<td>ฑ</td>
<td>3</td>
<td>0.0002</td>
</tr>
<tr>
<td>ฑ</td>
<td>24559</td>
<td>1.401</td>
<td>ฑ</td>
<td>3</td>
<td>0.0002</td>
</tr>
<tr>
<td>ฑ</td>
<td>23967</td>
<td>1.3673</td>
<td>ฑ</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>ฑ</td>
<td>23619</td>
<td>1.3474</td>
<td>ฑ</td>
<td>1</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Table 27. Relative frequency of characters in the final position

Those characters above the black line fall in the High Frequency class; those between the black and grey line are in the Medium class; and those below the grey line are in the Low class.

<table>
<thead>
<tr>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
<th>Character</th>
<th>n of 2 million</th>
<th>% of 2 million</th>
</tr>
</thead>
<tbody>
<tr>
<td>ง</td>
<td>189460</td>
<td>15.0436</td>
<td>ค</td>
<td>4215</td>
<td>0.3347</td>
</tr>
<tr>
<td>น</td>
<td>165776</td>
<td>13.1631</td>
<td>ฅ</td>
<td>4166</td>
<td>0.3308</td>
</tr>
<tr>
<td>ว</td>
<td>107099</td>
<td>8.5039</td>
<td>ฤ</td>
<td>4157</td>
<td>0.3301</td>
</tr>
<tr>
<td>ศ</td>
<td>77743</td>
<td>6.173</td>
<td>ฦ</td>
<td>3569</td>
<td>0.2834</td>
</tr>
<tr>
<td>ฉ</td>
<td>77288</td>
<td>6.1369</td>
<td>ฦๅ</td>
<td>2602</td>
<td>0.2066</td>
</tr>
<tr>
<td>ช</td>
<td>62490</td>
<td>4.9619</td>
<td>ษ</td>
<td>1810</td>
<td>0.1437</td>
</tr>
<tr>
<td>ซ</td>
<td>56122</td>
<td>4.4562</td>
<td>ฏ</td>
<td>1796</td>
<td>0.1426</td>
</tr>
<tr>
<td>ฌ</td>
<td>55284</td>
<td>4.3897</td>
<td>ฑ</td>
<td>1466</td>
<td>0.1164</td>
</tr>
<tr>
<td>ญ</td>
<td>51049</td>
<td>4.0534</td>
<td>ฎ</td>
<td>1263</td>
<td>0.1003</td>
</tr>
<tr>
<td>ฎ</td>
<td>44547</td>
<td>3.5372</td>
<td>ฏ</td>
<td>1182</td>
<td>0.0939</td>
</tr>
<tr>
<td>ฏ</td>
<td>42348</td>
<td>3.3625</td>
<td>ฏ</td>
<td>953</td>
<td>0.0757</td>
</tr>
<tr>
<td>ฐ</td>
<td>40184</td>
<td>3.1907</td>
<td>ฑ</td>
<td>676</td>
<td>0.0537</td>
</tr>
<tr>
<td>ฑ</td>
<td>35689</td>
<td>2.8338</td>
<td>ฏ</td>
<td>468</td>
<td>0.0372</td>
</tr>
<tr>
<td>ฒ</td>
<td>28682</td>
<td>2.2774</td>
<td>ฏ</td>
<td>456</td>
<td>0.0362</td>
</tr>
<tr>
<td>ฑ</td>
<td>27347</td>
<td>2.1714</td>
<td>ฏ</td>
<td>232</td>
<td>0.0184</td>
</tr>
<tr>
<td>ฎ</td>
<td>25080</td>
<td>1.9914</td>
<td>ฎ</td>
<td>227</td>
<td>0.0180</td>
</tr>
<tr>
<td>ฏ</td>
<td>21953</td>
<td>1.7431</td>
<td>ฏ</td>
<td>146</td>
<td>0.0116</td>
</tr>
<tr>
<td>ฏ</td>
<td>19382</td>
<td>1.539</td>
<td>ฏ</td>
<td>84</td>
<td>0.0067</td>
</tr>
<tr>
<td>ฏ</td>
<td>14072</td>
<td>1.1174</td>
<td>ฏ</td>
<td>67</td>
<td>0.0053</td>
</tr>
<tr>
<td>ฏ</td>
<td>13546</td>
<td>1.0756</td>
<td>ฏ</td>
<td>43</td>
<td>0.0034</td>
</tr>
<tr>
<td>ฏ</td>
<td>10901</td>
<td>0.8656</td>
<td>ฏ</td>
<td>40</td>
<td>0.0032</td>
</tr>
<tr>
<td>ฏ</td>
<td>9826</td>
<td>0.7802</td>
<td>ฏ</td>
<td>40</td>
<td>0.0032</td>
</tr>
<tr>
<td>ฏ</td>
<td>8412</td>
<td>0.6679</td>
<td>ฏ</td>
<td>30</td>
<td>0.0024</td>
</tr>
<tr>
<td>ฎ</td>
<td>7802</td>
<td>0.6195</td>
<td>ฏ</td>
<td>13</td>
<td>0.0010</td>
</tr>
<tr>
<td>ฏ</td>
<td>7467</td>
<td>0.5929</td>
<td>ฏ</td>
<td>13</td>
<td>0.0010</td>
</tr>
<tr>
<td>ฏ</td>
<td>7300</td>
<td>0.5796</td>
<td>ฏ</td>
<td>6</td>
<td>0.0005</td>
</tr>
<tr>
<td>ฏ</td>
<td>7271</td>
<td>0.5773</td>
<td>ฏ</td>
<td>5</td>
<td>0.0004</td>
</tr>
<tr>
<td>ฏ</td>
<td>6657</td>
<td>0.5286</td>
<td>ฏ</td>
<td>4</td>
<td>0.0003</td>
</tr>
<tr>
<td>ฏ</td>
<td>4385</td>
<td>0.3482</td>
<td>ฏ</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>ฏ</td>
<td>4226</td>
<td>0.3356</td>
<td>ฏ</td>
<td>1</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
From Table 27, it is noticeable that, unlike the characters in the start position in Table 26, some upper and lower vowel characters, tone characters, diacritic scripts as well as punctuation symbols are in the list. However, only characters that occur in the main horizontal line are used as characters in the end position to construct stimuli in this experiment.

The target word for each sentence was placed at or near the centre of the sentence, and all sentences were grammatical and meaningful. The average length of target words was similar for each condition and was equated within each start character frequency condition, such that word length in the three conditions High Start – High/Medium/Low End was approximately equal and in the three conditions Medium Start – High/Medium/Low End and in the three conditions Low Start – High/Medium/Low End were also approximately equal. The nine target word conditions, arrived at by a factorial combination of the frequency of the start and end characters are shown in Table 28 and Table 29, along with examples (As can be seen these are the same as in Tables 19 and 20 in Chapter 3 and are reproduced here for ease of reference.)
Table 28. Examples of sentences with high, medium or low frequency start and end characters used in Experiment 2 (Unspaced condition)

Target words are in red and underlined here for illustration

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Start Character</th>
<th>End Character</th>
<th>Mean Word Length</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High</td>
<td>6.5</td>
<td>ชายต้องเขียนตัวภาษาเพื่อสอนเด็ก</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Medium</td>
<td>6.7</td>
<td>หญิงต้องเขียนตัวเลขเพื่อใช้เป็นโจทย์</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>Low</td>
<td>6.2</td>
<td>ข้าวม่าเขียนภาพเพื่อให้ทำรายงาน</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>High</td>
<td>4.5</td>
<td>นักเรียนใหม่ชื่อซัลทร์มาจากต่างจังหวัด</td>
</tr>
<tr>
<td>5</td>
<td>Medium</td>
<td>Medium</td>
<td>4.5</td>
<td>เด็กใหม่ที่ชื่อโชติร์มาจากกรุงเทพมหานคร</td>
</tr>
<tr>
<td>6</td>
<td>Medium</td>
<td>Low</td>
<td>4.2</td>
<td>คุณครูคนใหม่ชื่อชัยวุฒิมาจากโรงเรียนชื่อดัง</td>
</tr>
<tr>
<td>7</td>
<td>Low</td>
<td>High</td>
<td>4.2</td>
<td>ลุงพงเป็นคนเล่นฆ้องที่เก่งมากที่สุด</td>
</tr>
<tr>
<td>8</td>
<td>Low</td>
<td>Medium</td>
<td>3.7</td>
<td>กล้าเป็นคนยิ้มหงายคงจะมีเรื่อง</td>
</tr>
<tr>
<td>9</td>
<td>Low</td>
<td>Low</td>
<td>4.2</td>
<td>ยายม่าเป็นคนอารมณ์จึงไม่มีใครคบ</td>
</tr>
</tbody>
</table>

The sentence frame for each of the nine conditions was fixed for the words that preceded and followed the target word – either exactly the same words or at least words having the same end character (preceding word) or start character (following word) were used to avoid the effects of the frequency of that particular character on the start and end characters of the target words. Thus the three sentences in each of the nine conditions consisted of target words starting with the same start character frequency tended to have similar meanings.
Table 29. **Examples of sentences with high, medium or low frequency start and end characters used in Experiment 2 (Spaced condition)**

Target words are in red and underlined here for illustration.

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Start Character</th>
<th>End Character</th>
<th>Mean Word Length</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 High</td>
<td>High</td>
<td>High</td>
<td>6.5</td>
<td>ชาย ต้อง เขียน ต่างๆ เพื่อ สอน เด็ก</td>
</tr>
<tr>
<td>2 High</td>
<td>Medium</td>
<td></td>
<td>6.7</td>
<td>หญิง ต้อง เขียน ตัวเลข เพื่อ ใช้ เป็นโจทย์</td>
</tr>
<tr>
<td>3 High</td>
<td>Low</td>
<td></td>
<td>6.2</td>
<td>เข้าวาม่า เขียน ตาราง เพื่อ ทำ รายงาน</td>
</tr>
<tr>
<td>4 Medium</td>
<td>High</td>
<td></td>
<td>4.5</td>
<td>นักเรียน ใหม่ ชื่อ ชลาร มาจาก ต่างจังหวัด</td>
</tr>
<tr>
<td>5 Medium</td>
<td>Medium</td>
<td></td>
<td>4.5</td>
<td>เด็ก ใหม่ ที่ ชื่อ ไตรภุช มาจาก กรุงเทพมหานคร</td>
</tr>
<tr>
<td>6 Medium</td>
<td>Low</td>
<td></td>
<td>4.2</td>
<td>คุณครู คน ใหม่ ชื่อ ชลาร มาจาก โรงเรียน ชื่อ คอง</td>
</tr>
<tr>
<td>7 Low</td>
<td>High</td>
<td></td>
<td>4.2</td>
<td>ลุงพง เป็น คน เล่น ฆ้อง ที่เก่ง มากที่สุด</td>
</tr>
<tr>
<td>8 Low</td>
<td>Medium</td>
<td></td>
<td>3.7</td>
<td>กล้า เป็น คน สีฟ้า จึง มักจะ มี เรื่อง</td>
</tr>
<tr>
<td>9 Low</td>
<td>Low</td>
<td></td>
<td>4.2</td>
<td>ยายม่า เป็น คน  pornos จึง ไม่มี ใคร คบ</td>
</tr>
</tbody>
</table>

**4.2.3. Apparatus**

Participants’ eye movements were recorded with an SR Research Ltd. (Toronto, Ontario, Canada) EyeLink II eye tracker (sampling rate = 500 Hz) that monitored the position of the participants’ left or right eye (depending on which eye was able to be tracked better in the initial calibration phase) every 2 msec. Participants viewed the stimuli on a 21-in. NEC monitor with 1,024 x 768 pixel resolution and 60 Hz refresh rate approximately 50 centimetres from their eyes. The stimuli were presented stationary in the centre of the computer screen in Cordia New font size 22 one sentence at a time with a 30 seconds time limit, after which the next stimulus sentence was automatically presented. To maintain attention, there was a comprehension question for participants to answer after every test sentence.
4.2.4. **Procedure**

Each participant was tested individually in a room with minimal noise and light interference (at the CRSLP, Chulalongkorn University, Bangkok, Thailand). There were 20 practice trials comprising two blocks of nine (3 start character frequencies x 3 end character frequencies), one block of nine with spaced text first then a block of nine with unspaced text and then two other practice sentences (1 spaced, 1 unspaced) acting as decoy or warm-up trials at the start of the Test session. The Test session comprised the 54 test sentences. Of these 27 were spaced and 27 were unspaced and test trials were presented randomly within three blocks of 18 trials (9 spaced and 9 unspaced) with one spaced and one unspaced sentence for each of the 3 (start) x 3 (end) character frequency conditions in each block. The stimuli were arranged in a rolling design so that members (high, medium or low end character) of the similar start-character-defined sentence frame were not presented in the same block. All participants were asked to do the test twice, first reading all the stimuli silently then two weeks later reading all the stimuli aloud. This was always done in the silent then aloud condition order for two reasons; first because silent reading is the main focus of this and other eye movement in reading studies, and secondly because reading aloud may have resulted in greater memory for the stimulus material than silent reading.

After reading a stimulus sentence, participants responded by hitting a key (SPACEBAR) and the comprehension question for that test sentence was then presented on the screen. After participants responded ‘true’ or ‘false’ to the question via the Right Shift (TRUE) and Left Shift (FALSE) keys, the next stimulus sentence was presented on the screen. The participants were required to answer the question before they could move on to the next stimulus trial.
The eye movement data were then processed using the DataViewer program (version 1.9.197) (SR Research Ltd., Toronto, Ontario, Canada) to define areas of interest (IA) for each target word for all words in the sentences and to extract the fixation data of each participant. There were two levels of processing; word and letter levels and these two methods of segmentation into IAs were used to derive eye fixations for analysis of both the spaced and unspaced text. Figure 27 shows an example for segmentation at the word level for a spaced text sentence. As can be seen for spaced text the space between two words was included as the leading space for the second word. For segmentation of unspaced text at the word level (refer Figure 28) the same method was applied but there were no spaces to contend with in the segmentation process. Figure 29 shows an example for segmentation at the letter level for a spaced text sentence. As can be seen at the letter level, the space is counted as a separate “letter”. For segmentation of unspaced text at the letter level (refer Figure 30) the same method was applied but there were no spaces to contend with in the segmentation process.

The three dependent variables were derived from these segmentations: from the word fixation segmentations, the duration of each fixation on a particular word was recorded and from this the total gaze duration per word was derived; from the letter segmentation the position (landing site) and the duration of the first fixation were derived.

Figure 27. Snapshot of word segmentation for spaced sentences
The eye movement fixation data of participants reading stimuli silently and aloud were then analysed using the lmer program for linear mixed-effects modelling (lme4 package; Bates & Sarkar, 2006) in the R system for statistical computing (R Development Core Team, 2006) (refer details in 3.3). For the reasons given in Chapter 3 mixed-effect models were considered to be the most suitable statistical analysis method for the current studies of eye movements in reading Thai. There were four nominally fixed factors - education (2 levels), start character frequency (3 levels), end character frequency (3 levels), writing style (2 levels) and reading mode (2 levels) and two nominally random factors – subjects and items (repetitions across the word start and word end frequency factors).

4.2.5. Hypotheses

From the results of the Reilly et al. (2003) studies, it is possible to hypothesise that the eye movement patterns and control by Thai readers should be similar to those of the readers of European languages. That is the word centre should be the most effective target (OVP) for the eyes during saccadic execution in Thai (unspaced text)
and their PVL should be at or near this position as well. The relative frequency of occurrence of characters at the start and end positions of words should affect the eye movements of Thai readers; characters occurring with higher frequency at word boundaries should better direct the eyes to word centre than lower frequencies. According to the findings of Reilly et al. (2003), end character frequency should have stronger effects than start character on landing site (refer Section 2 for details). There are no clear hypotheses for the effect of character frequency on word boundary positions for first fixation and gaze duration, except that there should either be reduced or no effects for these dependent variables.

4.3. Results

The overall comprehension rate for test sentences was 97% for Hi-Ed readers and 91% for the P/Mid-Ed readers indicating that participants read and understood the sentences to a high level of proficiency. Statistical analysis of the results in this study was divided into three parts. In the first two parts the target words (refer Table 28 and Table 29) were the focus of analysis and in the third part the focus was expanded to all words in the sentences. In all three sets of analyses the same three dependent variables (landing site location, first fixation duration and gaze duration) were analysed each in a separate mixed-effect model analysis as set out below.

Analysis Set I (Target Words, Predetermined Planned Fixed Effects) involved primarily the use of the predetermined factorial manipulations as fixed effects in the analysis of each of the three dependent variables. These fixed effects were education level, spacing, reading mode and frequency of start and end characters of the target
words. In addition, word frequency and word length of the target words were added in as control variable fixed effects. The random variables were participants and items (repetitions across frequency conditions. The frequencies of words and characters used in this analysis were the log frequency of occurrence of words, and of start and end characters based on the two-million word database (CRSLP Corpus version 1) from which the stimuli were constructed.

Analysis Set II (Target Words, Additional Fixed Effects) was similar to Analysis Set I, but used (i) the frequency of occurrence of start and end characters of target words calculated from the larger 10 million word database (CRSLP Corpus version 2), i.e. the database that was not used to create the target word character frequencies and (ii) another fixed factor that was not pre-determined in the experimental design – the relative frequency of the middle character of each target word. Middle characters were defined as the central character of the word along with any upper or lower vowel characters and tone characters presented over or below that particular middle character if there was an odd numbers of horizontal characters in the word or, if there was an even number of characters in the word, the mean frequency of the middle two characters. As for Analysis Set I word frequency and word length were also included as control fixed effects. All frequencies in this Analysis Set II were calculated from the 10-million word database (CRSLP Corpus version 2).

Analysis Set III (Sentences, Additional Fixed Effects) was similar to Analysis Set II except that all words in every sentence were analysed, not just the target words. In addition a third random factor was included in this analysis set, given that all words in the sentence were analysed. This was an overall measure of the distinctiveness of each horizontal character position summed over each word. The distinctiveness of each
horizontal position was given by a combination of whether the central line character extended above the central line, i.e., whether the character was tall (1) or not (0); whether the character included an upper vowel or not (1/0); a lower vowel or not (1/0); a tone character or not (1/0); the combined frequency of characters and upper and lower components in the Thai 10-million word corpus; and the combined frequency of characters and upper and lower components at that horizontal position in that particular horizontal position in a Thai word.

The eye movement data in these analyses were those if Hi-Ed and P/Mid-Ed readers reading, silently and aloud, the stimuli in spaced and unspaced conditions. The dependent variables – landing site location, first fixation duration and gaze duration – were derived using DataViewer program (refer section 5.2.4 for details). The details of all the analyses including other models that were tested are presented in detail in Appendix E2 and E3.

4.3.1. Analysis Set I (Target Words, Predetermine Planned Fixed Effects)

In this analysis, the eye movement data of all 30 participants reading the stimuli silently and aloud were analysed in terms of the landing site location, first fixation duration and gaze duration on the target word. The independent variables were the education level of the participants (edu), mode of reading (mode), a relative frequency of occurrences (High, Medium and Low) of start (startfreq) and end characters (endfreq) of the target word, the word frequency of the target word (wordfreq), target word length (wlength) and the spacing condition (spacing) of the sentence frames.
The average word length of target words was 4.98 letters. An overall summary of the data for mode and spacing condition for each of the three dependent variables is given in Table 30.

Table 30. **Mean landing site location, first fixation duration and gaze duration of Analysis Set I**

<table>
<thead>
<tr>
<th></th>
<th>Silent</th>
<th></th>
<th>Aloud</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spaced</td>
<td>Unspaced</td>
<td>Spaced</td>
<td>Unspaced</td>
</tr>
<tr>
<td>Landing Site Location (characters)</td>
<td>2.43</td>
<td>2.37</td>
<td>2.32</td>
<td>2.24</td>
</tr>
<tr>
<td>First Fixation Duration (msec.)</td>
<td>227.85</td>
<td>232.38</td>
<td>245.27</td>
<td>249.68</td>
</tr>
<tr>
<td>Gaze Duration (msec.)</td>
<td>349.51</td>
<td>366.99</td>
<td>445.28</td>
<td>440.84</td>
</tr>
</tbody>
</table>

Note: mean word length equals 4.98 letters

The mixed-effect models analyses of (i) landing site location, (ii) first fixation duration and (iii) gaze duration in Analysis Set I are described in turn in the following sections.

**Landing Site Location**

The results of mixed-effect analyses for landing site are shown in Table 31. As can be seen in Table 30, there was an overall effect of mode; landing site was further into the word when participants read silently than when they read aloud (t= 3.139). This is consistent with readers being more conservative in landing site when reading aloud and more efficient (closer to the OVP) when reading silently. This mode effect did not interact with any other effects and so is not discussed further. There was also an expected effect of word length; quite understandably, for longer words the landing site is further into the word (t= 8.336). The remainder of the effects for landing site
concern the effect of start frequency on landing site and various interactions with end frequency, word length, spacing, and education level and these are discussed below.

Table 31.  
**Mixed-effect models results of Analysis Set I on landing site location**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.780949</td>
<td>0.190430</td>
<td>4.101</td>
</tr>
<tr>
<td>edu.L</td>
<td>-0.036763</td>
<td>0.055484</td>
<td>-0.663</td>
</tr>
<tr>
<td>modesilent</td>
<td>0.141084</td>
<td>0.044947</td>
<td>3.139</td>
</tr>
<tr>
<td>wlength</td>
<td>0.277960</td>
<td>0.033345</td>
<td>8.336</td>
</tr>
<tr>
<td>startfreq.L</td>
<td>-0.981512</td>
<td>0.354465</td>
<td>-2.769</td>
</tr>
<tr>
<td>startfreq.Q</td>
<td>-0.778989</td>
<td>0.269631</td>
<td>-2.889</td>
</tr>
<tr>
<td>edu.L:startfreq.Q</td>
<td>0.114456</td>
<td>0.052705</td>
<td>2.172</td>
</tr>
<tr>
<td>edu.L:wordfreq.L</td>
<td>-0.124241</td>
<td>0.049565</td>
<td>-2.507</td>
</tr>
<tr>
<td>wlength:startfreq.L</td>
<td>0.195279</td>
<td>0.063172</td>
<td>3.091</td>
</tr>
<tr>
<td>wlength:startfreq.Q</td>
<td>0.110311</td>
<td>0.049550</td>
<td>2.226</td>
</tr>
<tr>
<td>startfreq.L:endfreq.L</td>
<td>-0.332113</td>
<td>0.153526</td>
<td>-2.163</td>
</tr>
<tr>
<td>startfreq.L:spacing.L</td>
<td>0.240737</td>
<td>0.104202</td>
<td>2.310</td>
</tr>
<tr>
<td>startfreq.Q:spacing.L</td>
<td>0.322582</td>
<td>0.086340</td>
<td>3.736</td>
</tr>
</tbody>
</table>

The most striking effect is that start character frequency strongly influences Thai readers’ eye movement patterns. As can be seen in Figure 31, both Hi-Ed and P/Mid-Ed participants landed their eyes further into the target word if the start character frequency was high, but the landing site moved backward toward the start of the word when the frequency was lower. This effect can be seen to be quite linear for the Hi-Ed readers but for the P/Mid-Ed readers the effect is more quadratic – the increment into the word is greater between low and medium than between medium and high frequency (t= 2.172), but irrespective of this small difference the overall effect
was the same—characters with higher frequency in the start position led to more optimal landing site.\(^\text{12}\)

Figure 31. Landing site location of participants by start character frequency

\(^{12}\) There was also an interaction of education level word frequency (\(t= -2.507\)); irrespective of start frequency or any other factors, for Hi-Ed readers word frequency has little if any effect on landing site, whereas for P/Mid-Ed readers higher frequency words assist them to land further into the word.
End character frequency also has a linear qualifying effecting on the linear effect of start character frequency ($t = -2.163$). As can be seen in Figure 32, the effect of start character frequency is more linear for medium and high end character frequencies than for low end frequencies. Note that the major effect appears to be start character frequency (there is a start frequency main effect) with end character frequency qualifying the magnitude and pattern of the start character frequency effect.

Figure 32. Landing site location as a function of start and end character frequencies

That the landing site was indeed optimal is shown by the interaction of start frequency (linear and quadratic) and word length ($t = 3.091$ and $2.226$ respectively). As can be seen in Figure 33, in general, as word length increased fixations landed a greater number of letters into the word. Moreover, despite some minor exceptions with
longer words (possibly due to the smaller number of longer words – see Appendix d for word list), the overall linear effect of higher start frequency resulting in fixations closer to the centre of the word was maintained.

Figure 33. Landing site location as a function of start character frequency and word length

Of particular importance in this study is whether landing site is affected by spacing between the words. Figure 34 shows that the start character frequency effect is essentially maintained across spaced and unspaced Thai text, although for normal unspaced text the effect of start character frequency is linear ($t= 2.310$) whereas for spaced text (when word boundaries were presumably more clear) there was a quadratic effect of start character frequency ($t= 3.736$); from a landing site less than 2 characters for low character frequency landing site jumped to well above 2 for both medium and high frequency ($t= 3.736$). Nevertheless these effects are striking; for both
spaced text and unspaced text first fixations land near the centre of the word and land
easier word centre the greater the start character frequency.

Figure 34.  Landing site location as a function of spacing and start character
frequency

Together these results show that Thai readers use the relative frequency of the
characters at the start position of words as a cue to word boundaries. The higher
frequency of these characters facilitates Thai readers by providing the visual cues that
guide their eyes to the position (near the centre of the word) that allows faster word
recognition. While there is some qualification of this start frequency effect by end
character frequency, spacing and education level, these are quantitative effects of
degree rather than qualitative differences; in general the higher the frequency of the
start character the further into the word and the closer to the word centre is the landing
site.
First Fixation Duration

The mixed-effect models analysis results are shown in Table 32. There was a main effect of mode of reading (t = -5.84) and an interaction effect between mode of reading and word frequency (t = -2.39). As can be seen in Figure 35, first fixation durations on the target word were longer when participants read the text aloud. Higher word frequencies helped reduce the first fixation duration in silent reading but in reading aloud first fixation duration remained relatively high irrespective of word frequency. The only other effect was a quadratic effect of start character frequency (t = -2.33), due to slightly longer fixation durations for Medium start character frequency (Mean = 245.66) than either Low (Mean = 231.85) or High (Mean = 234.30) start character frequency. This unexpected effect is possibly due to the use of log frequency to calculate start character frequency; as a result of this there were more items in the medium character frequency category (24 characters) than the low (18 characters) or high (12 characters) frequency categories and this may have biased the frequency calculations.

Table 32. Mixed-effect models results of Analysis I on first fixation duration

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>246.196</td>
<td>4.788</td>
<td>51.41</td>
</tr>
<tr>
<td>modesilent</td>
<td>-21.024</td>
<td>3.602</td>
<td>-5.84</td>
</tr>
<tr>
<td>startfreq.Q</td>
<td>-9.306</td>
<td>3.994</td>
<td>-2.33</td>
</tr>
<tr>
<td>modesilent:wordfreq. L</td>
<td>0.12.149</td>
<td>5.092</td>
<td>-2.39</td>
</tr>
</tbody>
</table>
Gaze Duration

As shown in Table 33, the gaze duration analysis showed only that mode of reading was significant both for its main effect (t= 11.482) and in its interaction effect with end character frequency (t= -2.113 [Linear]). As can be seen in Figure 36, gaze duration was longer when reading aloud and particularly long when the target word ended with a low frequency character than when it ended with greater (medium or high) character frequencies. This general influence of end character frequency on gaze duration may indicate that when reading target word with a low frequency end character, participants are possibly less certain about whether the word has in fact ended and therefore tend to add additional fixations (presumably refixations). However, for medium or high frequency end characters, participants are more certain
that the word has ended and so move onto the next word without additional fixations (refer Figure 36).

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>355.301</td>
<td>21.212</td>
<td>16.750</td>
</tr>
<tr>
<td>modealoud</td>
<td>85.640</td>
<td>7.458</td>
<td>11.482</td>
</tr>
<tr>
<td>modealoud:endfreq.L</td>
<td>-27.687</td>
<td>13.103</td>
<td>-2.113</td>
</tr>
</tbody>
</table>

Figure 36. Gaze duration as a function of end character frequency and mode

4.3.2. Analysis Set II (Target Words, Additional Fixed Effects)

In Analysis Set II landing site location, first fixation duration and gaze duration were analysed. The fixed effects were education level (edu), mode of reading (mode),
relative frequency of occurrence (High or Low) of start (first), middle (mid) and end characters (last) of the target word, the word frequency (wordfreq), target word length (wlength) and the spaced condition (spacing) of the sentence frames. The log character frequency used in Analysis Set II was calculated from the 10-million word database (CRSLP Corpus version 2). Means for length of target words, landing site location, first fixation duration and gaze duration across mode and education level are the same as in Analysis Set I and are shown in Table 30. Analyses for landing site location, first fixation duration and gaze duration in are set out below.

**Landing Site Location**

Three main effects were significant, mode of reading ($t = 3.081$), word length ($t = 9.410$), and middle character frequency ($t = -2.661$). Similar to the results in Analysis Set I, landing site location of participant was closer to the start area of word when reading aloud (refer Table 34). Surprisingly, there was no effect of start or end character frequency at all in this analysis when changing the database to the larger one and adding middle character frequency. In fact the frequency effect found in this analysis is for the middle frequency ($t = -2.661$); in general the higher the frequency of the middle characters, the further to the left of word centre was the landing site location. This can be seen in Figure 37 as a function of word length. Except for 7-character words high frequency middle characters had the effect of directing the first fixation landing site further to the left of word centre than did low frequency middle characters. It is unclear why the higher frequency middle characters pulled the eyes backward towards word start rather than sending them to the PVL (word-centre). However, it should be noted that the frequency of the characters in this middle position of the words was not part of the orthographic features used in constructing the stimuli
used in this experiment. Further studies in which middle frequency is systematically manipulated would be of use to investigate this further.

Table 34. Mixed-effect Models results of Analysis Set II on Landing Site Location

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.33345</td>
<td>0.18900</td>
<td>7.055</td>
</tr>
<tr>
<td>wlength</td>
<td>0.24438</td>
<td>0.02597</td>
<td>9.410</td>
</tr>
<tr>
<td>mid</td>
<td>-0.61692</td>
<td>0.23184</td>
<td>-2.661</td>
</tr>
<tr>
<td>modesilent</td>
<td>0.13863</td>
<td>0.04500</td>
<td>3.081</td>
</tr>
</tbody>
</table>

Figure 37. Landing site location as a function of word length and middle character frequency

First Fixation Duration

Only mode of reading (t = -5.303) had significant effect on first fixation duration in this analysis (see Table 35). In concert with the other effects for mode so far, first fixation duration was longer when participants read aloud and this was true for
both Hi-Ed and P/Mid-Ed readers as shown in Figure 38. No other factors had any effects on first fixation duration.

Table 35. Mixed-effect Models results of Analysis Set II on First Fixation Duration

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>225.4590</td>
<td>12.1530</td>
</tr>
<tr>
<td>modesilent</td>
<td>-17.4874</td>
<td>3.2973</td>
</tr>
</tbody>
</table>

Figure 38. First fixation duration as a function of mode of reading and education

Gaze Duration

As shown in Table 36, the analysis for gaze duration showed only main effects for mode of reading (t= 11.521) and word frequency (t= -2.273). These two main effects are plotted together in Figure 39: gaze duration for target words when reading
aloud was always longer than for silent reading; and gaze duration on low frequency target words was longer than that on high frequency words.

Table 36.  **Mixed-effect Models results of Analysis Set II on Gaze Duration**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>336.431</td>
<td>22.182</td>
<td>15.167</td>
</tr>
<tr>
<td>modealoud</td>
<td>85.902</td>
<td>7.456</td>
<td>11.521</td>
</tr>
<tr>
<td>wordfreq.L</td>
<td>-62.522</td>
<td>27.504</td>
<td>-2.273</td>
</tr>
</tbody>
</table>

Figure 39.  **Gaze duration as a function of target word frequency and mode of reading**

4.3.3.  **Analysis Set III (Sentences, Additional Fixed Effects)**

In Analysis Set III, eye movement data of all participants reading all the words in each sentence analysed using the same independent variables as in Analysis Set II,
i.e., the education level of the participant (edu), mode of reading (mode), a relative frequency of occurrence (High and Low) of start (first), middle (mid) and end characters (last) of each word in the sentence, the word frequency (wordfreq), word length (wlength) of each word and the spaced condition (spacing) of the sentence frames. The character frequency use in this analysis was also calculated from the CRSLP Corpus version 2 as in Analysis Set II. The average word length was 3.92 characters. Mean landing site location on the words in the sentences, first fixation duration and gaze duration across spaced and unspaced and for reading aloud or silently are given in Table 37.

Table 37. Mean landing site location, first fixation duration and gaze duration of Analysis Set III

<table>
<thead>
<tr>
<th></th>
<th>Silent</th>
<th></th>
<th>Aloud</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spaced</td>
<td>Unspaced</td>
<td>Spaced</td>
<td>Unspaced</td>
</tr>
<tr>
<td>Landing Site Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(characters)</td>
<td>1.92</td>
<td>2.27</td>
<td>2.05</td>
<td>2.36</td>
</tr>
<tr>
<td>First Fixation Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(msec.)</td>
<td>214.70</td>
<td>218.10</td>
<td>237.22</td>
<td>244.60</td>
</tr>
<tr>
<td>Gaze Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(msec.)</td>
<td>268.49</td>
<td>275.16</td>
<td>324.75</td>
<td>330.07</td>
</tr>
</tbody>
</table>

Note: mean word length was 3.92 characters

Mixed-effect models analyses for landing site location, first fixation duration and gaze duration are set out in turn below.

**Landing Site Location**

The results of mixed-effect models analysis of landing site location on the sentences are shown in Table 38. As can be seen only the main effects of mode of reading (t= 4.938), word length (t= 14.008) and first character frequency (t=2.755) were significant.
As can be seen in Table 38, the effect of mode was similar to that for target words: when reading aloud participants’ landing site location was closer to the beginning of the word. The main results of start character frequency and word length are shown together in Figure 40. As words get longer the landing site understandably moves a greater number of characters into the word. In addition, in a result similar to the overall effects in Analysis Set I, the high frequency start characters assisted the eyes in moving further into the word towards the OVP, than did low frequency start characters, irrespective of word length.
First Fixation Duration

Analysis details for first fixation are set out in Table 39. Similar to the results for target words, first fixation duration was longer when reading aloud ($t = -15.75$; refer Table 37). In addition, there was a main effect of word frequency ($t = -2.23$), and an interaction between spacing and word frequency ($t = 2.67$).
Table 39. Mixed-effect models results of Analysis Set III on first fixation duration

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>244.562</td>
<td>3.868</td>
<td>63.23</td>
</tr>
<tr>
<td>modesilent</td>
<td>-24.696</td>
<td>1.568</td>
<td>-15.75</td>
</tr>
<tr>
<td>wordfreq.L</td>
<td>-4.648</td>
<td>2.081</td>
<td>-2.23</td>
</tr>
</tbody>
</table>

The effect of word frequency and its interaction with spacing are shown in Figure 41. As can be seen spaces between words facilitated shorter first fixation duration when word frequency was high, but spacing had little or no effect for low frequency words. Thus it is possible that spacing between words allows readers to identify words (especially high frequency words) more quickly in the first fixation and move on to the next word.

Figure 41. First fixation duration as a function of word frequency and spacing
Gaze Duration

Mixed-effect models analysis results on gaze duration for the sentences are shown in Table 40. There was the, by now, expected effect of mode (t= 20.53) due, as can be seen in Table 37, to longer gaze duration for reading aloud. As can be seen in Table 40, there was both a main effect of word frequency (t = -2.71) and word frequency x education level (t = -2.21); there was a slight advantage for Hi-Ed over P/Mid-Ed for low word frequency resulting in shorter gaze duration, but there was no such effect for high word frequency.

Most interestingly there was a main effect for spacing (t= 4.31), graphically represented for each reading mode in Figure 43. In a similar effect to that found for first fixation spaces between words appears to have assisted in word recognition, for total gaze duration on a word is less when there are spaces between the words.

Table 40. Mixed-effect Models results of Analysis Set III on Gaze Duration

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>220.526</td>
<td>3.795</td>
<td>58.12</td>
</tr>
<tr>
<td>modealoud</td>
<td>27.074</td>
<td>1.319</td>
<td>20.53</td>
</tr>
<tr>
<td>spacing.L</td>
<td>7.236</td>
<td>1.681</td>
<td>4.31</td>
</tr>
<tr>
<td>wordfreq.L</td>
<td>-5.552</td>
<td>2.051</td>
<td>-2.71</td>
</tr>
<tr>
<td>edu.L:wordfreq.L</td>
<td>2.915</td>
<td>1.317</td>
<td>-2.21</td>
</tr>
</tbody>
</table>
**Figure 42.** Gaze duration as a function of education and word frequency

**Figure 43.** Gaze duration as a function of spacing and mode of reading
4.4. Discussion

There was a main effect of mode in all three sets of analyses showing that when reading aloud (compared to reading silently), Thai readers’ (i) eyes land closer to the beginning of the word, (ii) have longer first and gaze durations. Despite slightly greater effects of mode on first fixations for higher word frequencies and on gaze durations for low and medium end word character frequencies in Analysis Set I, this effect of mode was consistent across all three analysis sets. Moreover, none of the other effects were qualitatively affected by mode: all effects discussed further below were the same across modes.

The results for the three independent variables are now summarised in turn across the three sets of analyses ahead of a more general interpretation of the data.

Landing Site: Analysis Set I showed a consistent effect of the start word frequency on landing site; despite some minor differences of degree due to end frequency, spacing, and education, the higher the start word character frequency, the further into the word towards word centre were located the Thai readers’ first fixation location. In Analysis Set II in which the middle frequency measure was added, this first character effect appears to have been replaced by an effect of middle word character frequency, as if the middle character frequency overshadowed the first character effect. However, it should be noted that middle frequency was not part of the original design of the experiment and so it is possible that this effect may in part be due to uncontrolled variations of middle character frequencies. That such an interpretation may be correct is borne out by the results of Analysis Set III in which, once again there was a clear effect of the start word character frequency. As the sample of words analysed in Analysis Set III was greatly increased by including all words in
the sentences, it is possible that any anomalous effects of Analysis Set II due to (a) testing for variables not used to design the study and (ii) basing the analysis on a more limited sample of words may have been overcome by the greater power afforded by the analysis of more words.

It thus appears that the dominant variable affecting landing site is start word character frequency; Thai readers use this to direct their eyes to an optimal viewing position. Moreover, this is an effective cue both in spaced text and also in unspaced text; Thai readers use start word character frequency to direct eye movements even when these characters are “disguised” or “hidden” by the lack of spaces between words.

First Fixation: The most dominant variables affecting first fixation are word frequency and spacing. It appears that the slowing effect (longer first fixations) of reading aloud mode is potentiated with high frequency words (Analysis Set I) and that the benefit of higher word frequency on reducing the duration of the first fixation is only apparent in spaced text.

Gaze Duration: For Gaze duration higher word frequency also assists in reducing gaze duration (Analysis Sets II and III); higher frequency words are presumably recognised more quickly necessitating less overall gaze duration. In Analysis Set I there is also some indication that the higher the word end character frequency the more certain are readers in determining that the end of the word has in fact been reached in both reading aloud and silently, but more so for reading aloud. And in Analysis Set III there is an overall affect of spacing – gaze duration is reduced for spaced text.
The OVP in Thai: The most striking finding here is that the landing site in for Thai readers reading Thai script is close to the centre of the word, the OVP, just as it is European readers reading European script. Moreover this was the case irrespective of spacing; unlike European readers whose PVL moves from the OVP near word centre for spaced script to initial area of the word when the script is unnaturally changed to unspaced (Rayner, 1998), Thai readers’ landing site does not shift between unspaced and spaced script. The PVL of Thai readers is always close to the centre of the words regardless of the space conditions of the text or the modes of reading, even though the landing site was a little further to the left of the word centre when reading aloud. Thus it can be suggested that the oculomotor control of Thai adults reading Thai script does not differ from that found in the English and other European languages.

Word recognition in Thai: Despite the lack of spacing effects on landing site, effects of spacing do emerge in first fixation and gaze duration; it appears that spacing between words, while not affecting oculomotor control, does reduce looking times. This suggests that spacing allows quicker word recognition, thus allowing readers to move onto the next word more quickly.

What Drives Thai Eye Movements? The critical factor for directing eye movements in Thai appears to be the first character frequency in words; higher frequencies result in landing the eyes closer to the word centre. This is interesting given that they confirm that letter frequencies affect eye movements in Thai but that it is the start character frequencies that do the work here rather than the end characters, as found in by Reilly et al. (2003). Further studies on different Thai orthographic features, such as position of vowels and other diacritics, are required to conclude if this
Start character frequency is the main visual cue that guides eye movements of Thai reader.
CHAPTER FIVE

EXPERIMENT 3: CHILD READERS’ EYE MOVEMENTS IN READING THAI
5. EXPERIMENT 3: CHILD READERS’ EYE MOVEMENTS IN READING THAI

5.1. Introduction

In Experiment 3 Thai children’s eye movements in reading sentences were investigated using material similar to that used in the eye movement study with adults in Experiment 2. There were four groups of child participants in this experiment, two ages (younger children and older children) with two groups of children at each age – children with good or poor reading ability. As for the adults, children’s reading was investigated as a function of character frequency (high, medium and low) in word-start and word-end position, as well as spaces between words. In contrast to Experiment 2, only silent reading was investigated, and the data from the Thai adults reading silently in Experiment 2 were included in the second part of the analysis for comparison purposes. Thus the study employed a 3 age groups [Younger children vs. Older children vs. adults] x 2 reading ability levels [Good vs. Poor readers] x (2 [spaced/unspaced] x 3 [high, medium, low character frequencies in start position] x 3 [high, medium, low character frequencies in end position]) design, with repeated measures on the last three factors. The statistical analysis of eye movement data in this experiment, like that in Experiment 2, involved the use of mixed-effects models analysis. As for Experiment 2 the dependent variables were landing site location, first fixation duration and gaze duration.

5.2. Method

5.2.1. Participants

There were two ages (younger, older) x two levels of reading ability (good, poor) – four groups of child participants in this eye movement study. The participants’
latest results of the final examination in Thai subjects were used as the criterion to group them into either good or poor reader group. The results of each subject in the Thai education system are divided into a 4-point system (with an extra grade for failure), where 1 is the lowest and 4 is the highest mark of the passing results, based on the overall score of 100 points. Normally, students whose overall scores for all subjects are less than 20 points will fail that subject; scores between 20 to 39 points will receive grade 1; scores between 40 to 59 points will receive grade 2; scores of 60 to 79 will receive grade 3; and scores higher than 80 points will receive grade 4. In this study, students who received a grading of 3 or 4 for Thai subjects in the previous semester were classified as Good readers and students who received grade 1 and 2 were classified as Poor readers.

Ideally, this study was planned to test the same groups of participants as in the reading time study (Experiment 1) that were first, second, fifth, and sixth grade students. However, this was not possible due to (i) the eye tracker apparatus (EyeLink II) used in this experiment was not fit to test a small child like first and second graders; (ii) the test had to be conducted at the Centre for Research in Speech and Language Processing, Chulalongkorn University, but it was difficult to arrange permission and transportation to take the children from school. Therefore the best substitution for these primary school students was Grade 7 students who did not need a permission from school to participate in the study. Fortunately near the end of the data collection, some of the adult participants gave the consent to run the experiments with their own or friend’s and relative’s children who were study in Grade 6 at the time thus there were two groups of child participants participate in this experiment.
There were eight participants in the older child group (4 good and 4 poor readers) and these were students who were about to finish the second semester of Grade 7 (first year of secondary school) at the time of the test [mean age = 12.7 years (range = 12 – 14 years)]. All of these older children successfully completed the tasks (refer 5.2.4, Procedure). There were 20 participants in the younger group (10 good and 10 poor readers) and they were students who were just starting the first semester of Grade 6 (last year of primary school) at the time of the study. Of these 20 only 15 participants finished the tasks and of these the eye movement data of only 10 participants could be used in the analysis due to the loss of fixation data on the texts, e.g. the record of five participants’ fixation were far beyond the interest area on the screen and could not be retrieved. The final sample of 10 younger children had a mean age of 11.4 years (range = 11 – 12 years), and 5 were good and 5 were poor readers.

The adult participants were the same participants as those in 5.1.1. However, instead of referring to adult participants as Hi-Ed versus P/Mid-Ed groups they were labelled Good and Poor readers respectively so the labels match those of the child participants here.

5.2.2. Materials

There were two sets of materials used in this Experiment 3, the first set [Stimulus Set 1] (easy version of the text) was used for the younger children participants and the second set [Stimulus Set 2] (more difficult version) was used to test the older children group. Stimulus Set 2 was the same set of materials used in Experiment 2 for adults (refer 5.1.2 for details). Stimulus Set 1 also comprised 54 test sentences and 20 practice sentences as well. All manipulated conditions of the text were the same as those in Experiment 2, that is the test sentences were the result of the
combination of spaced/unspaced text x 3 start character frequencies (high, medium, low) x 3 end character frequencies (high, medium, low) x 3 exemplars. Unlike Stimulus Set 2, in which all the stimuli were separated into 3 blocks of 18 trials but still in the same test, the stimuli in Stimulus Set 1 were separated into 4 tests: the first test (Practice) consisted of the practice sentences only (18 sentences), the second (Test 1) consisted of 20 sentences (2 of which were practice sentences always presented at the start), and the third (Test 2) and fourth (Test 3) consisted of 18 sentences each. Every participant did the tests in the same order: Practice, Test 1, Test 2 and then Test 3. The 18 test sentences within each test were presented randomly but according to the frequencies of the start character (refer 4.2.4). The relative frequency of start and end characters were in the same categories as in Experiment 2 (refer details in Section 4.2.2) – with the only difference between these two sets of stimuli being that the words and sentence frames for the younger children consisted of easier materials.

The nine target word conditions of Stimulus Set 1, the easy version, were arrived at by a factorial combination of the frequency of the start and end characters, are shown in Table 41 and Table 42, along with examples. Examples of Stimulus set 2, the more difficult version, can be seen in Section 4.2.2.
Table 41. Examples of sentences with high, medium or low frequency start and end characters used in this Experiment 3 (Unspaced condition)

Target words are in red and underlined here for illustration

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Start Character</th>
<th>End Character</th>
<th>Mean Word Length</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High</td>
<td>4.3</td>
<td>สุดาพา แก่น มาที่ศาลา</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Medium</td>
<td>4.2</td>
<td>มะนีพา ราฆะ มาที่บ้าน</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>Low</td>
<td>4.3</td>
<td>เลอกพา เมฆ มาที่บ้าน</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>High</td>
<td>4.2</td>
<td>สิ่งมักจะถูกเพื่อน ชินชม อยู่เสมอ</td>
</tr>
<tr>
<td>5</td>
<td>Medium</td>
<td>Medium</td>
<td>4.5</td>
<td>วินมักจะถูกเพื่อน โทรศัพท์ อยู่เสมอ</td>
</tr>
<tr>
<td>6</td>
<td>Medium</td>
<td>Low</td>
<td>4.2</td>
<td>กิ๊บ มักจะถูกเพื่อน ซักไซ้ อยู่เสมอ</td>
</tr>
<tr>
<td>7</td>
<td>Low</td>
<td>High</td>
<td>4.2</td>
<td>ดาว มักจะเข้า มาในช่วงเย็นเสมอ</td>
</tr>
<tr>
<td>8</td>
<td>Low</td>
<td>Medium</td>
<td>4.0</td>
<td>นั้นเป็น ผู้ดีที่อยู่ในหมู่บ้านอื่น</td>
</tr>
<tr>
<td>9</td>
<td>Low</td>
<td>Low</td>
<td>4.2</td>
<td>แจ้งเป็นคน ผู้ดีที่สุดในห้อง</td>
</tr>
</tbody>
</table>

Table 42. Examples of sentences with high, medium or low frequency start and end characters used in this Experiment 3 (Spaced condition)

Target words are in red and underlined here for illustration

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Start Character</th>
<th>End Character</th>
<th>Mean Word Length</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High</td>
<td>4.3</td>
<td>สุดา พา แก่น มาที่ศาลา</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Medium</td>
<td>4.2</td>
<td>มะนี พา ราฆะ มาที่บ้าน</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>Low</td>
<td>4.3</td>
<td>เลอก พา เมฆ มาที่บ้าน</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>High</td>
<td>4.2</td>
<td>สิ่งมักจะถูกเพื่อน ชินชม อยู่เสมอ</td>
</tr>
<tr>
<td>5</td>
<td>Medium</td>
<td>Medium</td>
<td>4.5</td>
<td>วินมักจะถูกเพื่อน โทรศัพท์ อยู่เสมอ</td>
</tr>
<tr>
<td>6</td>
<td>Medium</td>
<td>Low</td>
<td>4.2</td>
<td>กิ๊บ มักจะถูกเพื่อน ซักไซ้ อยู่เสมอ</td>
</tr>
<tr>
<td>7</td>
<td>Low</td>
<td>High</td>
<td>4.2</td>
<td>ดาว มักจะเข้า มาในช่วงเย็นเสมอ</td>
</tr>
<tr>
<td>8</td>
<td>Low</td>
<td>Medium</td>
<td>4.0</td>
<td>นั้นเป็น ผู้ดีที่อยู่ในหมู่บ้านอื่น</td>
</tr>
<tr>
<td>9</td>
<td>Low</td>
<td>Low</td>
<td>4.2</td>
<td>แจ้งเป็นคน ผู้ดีที่สุดในห้อง</td>
</tr>
</tbody>
</table>
5.2.3. Apparatus

The same apparatus was used as in Experiment 2, an SR Research Ltd. (Toronto, Ontario, Canada) EyeLink II eye tracker (sampling rate = 500 Hz) coupled with a 21-in. NEC monitor with 1,024 x 768 pixel resolution and 60 Hz refresh rate approximately 50 centimetres from their eyes. As in Experiment 2, each sentence was presented stationary in the centre of the screen in Cordia New font size 22 with 30 seconds time limit after which the next stimulus sentence was automatically presented. To maintain attention, there was a comprehension question for participants to answer after every test sentence.

5.2.4. Procedure

A similar procedure as in Experiment 2 was used. Each participant was tested individually at the CRSLP, Chulalongkorn University, Bangkok, Thailand. For older children participants the same procedure as for the adult participants was used. For the younger children Test 1 (refer Materials), the practice trials, and Test 2 were presented and then the participants were given 5 to 10 minutes break. They then continued with Tests 3 and 4.

After reading a stimulus sentence, participants responded by hitting a key (SPACEBAR) and the comprehension question for that test sentence was then presented on the screen. After participants responded ‘true’ or ‘false’ to the question via the Right Shift (TRUE) and Left Shift (FALSE) keys, the next stimulus sentence was presented on the screen. The participants were required to answer the question before they could move on to the next stimulus trial.
The eye movement data was then processed using the DataViewer program (version 1.9.197) to define area of interest (IA) for each target words and all words in the sentences and to extract the fixation data of each participant. The same two levels of processing were used as in Experiment 2; word and letter levels (refer to Section 4.2.4 for details).

As for Experiment 2, eye movement fixation data of participants were analysed using the lmer program for linear mixed-effects modelling (lme4 package; Bates & Sarkar, 2006) in the R system for statistical computing (R Development Core Team, 2006) (refer details in 4.2).

5.2.5. Hypotheses

The results of Experiment 2 with adults showed that Thai readers’ PVL was at or near the word centre, therefore it is considered possible that the children’s PVL will be at the same position. However, from the results of the spacing and reading ability effects of the reading time experiment (Experiment 1) showing that there was a very strong effect of spaces between words and the reading ability on reading time that decreased as the reading skills of the readers increased (refer 3.3 for details), it is possible to hypothesise that landing site location of young children should be at or near the initial area of the word and that spaces between words should have relatively strong effects on eye movements of children, especially the poor and possible younger readers. Spaces between words should be the primary factor that facilitate eye movement control of the younger child participants, by driving the eyes closer to the PVL. The frequencies of the characters at the start and end position should come into play only when there are no visible visual cues available, i.e., in the unspaced condition. Based on the education and age of child participants, eye movement control
of the older child group should be more similar to those of the adults than to those of
the younger child group.

5.3. Results

The overall comprehension rate of the child participants for the test sentences in
this experiment was good: 96% for older children good readers; 90% for older
children poor readers; 90% for younger children good readers; and 85% indicating
that the participants understood the sentences well.

The same three sets of analysis were conducted as in Experiment 2. In Analysis
Set I (Target Words, Planned Fixed Effects) the independent variables were age group
(age), reading ability level (ability), start character frequencies (startfreq), end
character frequency (endfreq), word frequency (wordfreq), word length (wlength) and
spacing condition (spacing). For Analysis Set II (Target Words, Additional Fixed
Effects) additional fixed effects were included – the relative frequencies of occurrence
of start (first), middle (mid), end character (last) as well as the word frequency
(wordfreq) calculated from the CRSLP Corpus version 2. In Analysis Set III
(Sentences, Additional Fixed Effects), the frequencies were those calculated from the
CRSLP Corpus version 2 only (first, mid, last and wordfreq). In all three Analysis
Sets, the eye movements of child participants above were analysed first, and then the
data of the adults were added for comparison. The same three dependent variables as in
Experiment 2, landing site location, first fixation duration, and gaze duration were
analysed each in a separate mixed-effect model analysis within each analysis set.
5.3.1. Analysis Set I and II (Target Words with Planned and with Additional Fixed Effects)

In these analyses, only the eye movements on the target words were analysed. Mean landing site location, first fixation duration and gaze duration of younger children, older children and adults for target words in Experiment 3 is shown in Table 43. Mean target word length for Stimulus Set 1 (younger children) was 4.22 characters and mean target word length of Stimulus Set 2 (older children and adults) was 4.98 characters.

Table 43. Mean landing site location, first fixation duration and gaze duration of participants in Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Spaced</th>
<th>Unspaced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Younger children</td>
<td>Older children</td>
</tr>
<tr>
<td>Landing Site Location</td>
<td>2.36</td>
<td>2.23</td>
</tr>
<tr>
<td>(characters)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Fixation Duration</td>
<td>293.91</td>
<td>226.00</td>
</tr>
<tr>
<td>(msec.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze Duration</td>
<td>340.66</td>
<td>231.88</td>
</tr>
<tr>
<td>(msec.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Mean target word length in for Stimulus Set 1 was 4.22 characters and mean target word length of Stimulus Set 2 was 4.98 characters.

Table 43 shows that when there were no visible visual cues available (unspaced condition) overall the landing site location of the younger child participants was closer to the beginning of the word than were those of the older children and adults. Mean first fixation duration and gaze duration of older children, as well as the adults, were longer when reading unspaced words. Surprisingly, first fixation duration and gaze duration of younger children group were in the opposite direction, i.e., the durations were longer when reading spaced text. These were the result of a greater number of fixations on the words in the younger child group. As words in the spaced condition
were a clearer target, it should be easier for poor reader younger children to refixate the words more often is they desired. That this was indeed the case is indicated by the greater difference between gaze duration and first fixation in the younger child than the older child group, especially in the spaced condition. The results of the mixed-effect models analyses for the three dependent variables are set out below.

**Landing Site Location**

Details of the analysis of the planned (predetermined) fixed effects (Analysis Set I) for landing site are shown in 0 (IA – children only) and Table 45 (IB – children and adults) and for the additional fixed effects (Analysis Set II) in Table 46 (IIA – children only) and Table 47 (IIB – children and adults). In Analysis Set IA, the analysis revealed that the main effects of spacing (t = -2.113) word length (t = 5.047) were significant. Even though the main effect of age was not significant, there was a significant interaction effect of spacing by age (t= 2.004). In Figure 44 it can be seen that, as would be expected, landing site location was further into the word as the length of the target word increased. Figure 45 shows that children landed their eyes further into the words when reading spaced text and younger children’s landing site on target words in unspaced text was closer to the word-initial area. In essence, older children’s landing site was unaffected by spacing – landing site was relatively optimal in both spaced and unspaced conditions, but for younger children spaced text allowed them to land fixations further into the word.
Table 44. Mixed-effect models results of children participants’ landing site location on target word (Analysis Set IA)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.323</td>
<td>0.177</td>
<td>7.464</td>
</tr>
<tr>
<td>spacing.L</td>
<td>-0.142</td>
<td>0.067</td>
<td>-2.113</td>
</tr>
<tr>
<td>wlength</td>
<td>0.183</td>
<td>0.036</td>
<td>5.047</td>
</tr>
<tr>
<td>age.L:spacing.L</td>
<td>0.189</td>
<td>0.095</td>
<td>2.004</td>
</tr>
</tbody>
</table>

Figure 44. Landing site location as a function of word length
151

Figure 45. **Landing site as a function of age and spacing**

Turning to Analysis Set IB which included the adults’ eye movements data, the results showed that only word length had a significant main effect on landing site ($t = 9.816$). There was still a significant interaction effects between spacing and age group ($t = 2.081$) due to the same cause – only for younger children did spacing facilitate fixations further into the word. As shown in Figure 46 the landing site location of older participants was closer to the word centre when there were no spaces between words available as a visual cue.

In addition, the interaction effect of spacing and end character frequency became significant in this analysis ($t = -2.136$). Figure 47 shows that the effects of end character frequencies were more pronounced in the unspaced condition, that is the higher the end character frequency, the closer was the landing site to the word centre in the unspaced condition. Such an effect was not apparent in the spaced condition.
Table 45.  Mixed-effect models results of child and adult participants’ landing site location on target word (Analysis Set IB)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.03653</td>
<td>0.13679</td>
<td>7.578</td>
</tr>
<tr>
<td>wlength</td>
<td>0.25602</td>
<td>0.02608</td>
<td>9.816</td>
</tr>
<tr>
<td>spacingPresent:age.Q</td>
<td>0.21460</td>
<td>0.10311</td>
<td>2.081</td>
</tr>
<tr>
<td>spacing.L:endfreq.Q</td>
<td>-0.18819</td>
<td>0.08810</td>
<td>-2.136</td>
</tr>
</tbody>
</table>

Figure 46.  Landing site as a function of age and spacing
The results of Analysis Set IIA (additional fixed effects) found that only the main effects of word length was significant ($t = 5.140$). However, when adding the adults’ data in the analysis (IIB), the results showed that in addition to word length ($t = 10.048$), there was a significant main effect of middle character frequency ($t = -2.861$). As can be seen in Figure 48, for all age groups high frequency middle characters pulled the eyes back further to the left of word centre, a similar result to those of the adult participants in Chapter 4. The reason why the higher middle character frequency pulls the eyes back toward the word-initial area is still unclear.

Table 46. Mixed-effect models results of child and adult participants’ landing site location on target word (Analysis Set IIA)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.530</td>
<td>0.298</td>
<td>5.125</td>
</tr>
<tr>
<td>wlength</td>
<td>0.200</td>
<td>0.039</td>
<td>5.140</td>
</tr>
</tbody>
</table>
Table 47. Mixed-effect models results of child and adult participants’ landing site location on target word (Analysis Set IIB)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.280493</td>
<td>0.206684</td>
<td>6.195</td>
</tr>
<tr>
<td>mid</td>
<td>-0.668078</td>
<td>0.233548</td>
<td>-2.861</td>
</tr>
<tr>
<td>wlength</td>
<td>0.256166</td>
<td>0.025493</td>
<td>10.048</td>
</tr>
</tbody>
</table>

Figure 48. Landing site location as a function of middle character frequency

**First Fixation Duration**

The results of mixed-effect models of Analysis Set IA showed in Table 48. The results showed that only the main effect of end character frequency was significant ($t = 2.292$). It also showed a significant interaction effect of end character frequency by word length ($t = -2.472$). Figure 49 shows that first fixation duration of the participants was longer the lower the frequency of the end character with this effect being more
pronounced as words became longer with the effect being ameliorated with short 4-letter words and disappearing completely with shorter 3-letter words. Result of Analysis Set IB shown in Table 49 reveal that only the age of the participants showed a significant effect on first fixation duration (t= -2.42).

Table 48. **Mixed-effects models results of child participants’ first fixation duration on target word (Analysis IA)**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>258.519</td>
<td>22.393</td>
<td>11.544</td>
</tr>
<tr>
<td>endfreq.L</td>
<td>78.126</td>
<td>34.080</td>
<td>2.292</td>
</tr>
</tbody>
</table>

Table 49. **Mixed-effects models results of child and adult participants’ first fixation duration on target word (Analysis IB)**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>238.281</td>
<td>6.561</td>
<td>36.32</td>
</tr>
<tr>
<td>age.L</td>
<td>-23.084</td>
<td>9.938</td>
<td>-2.42</td>
</tr>
</tbody>
</table>
There were no significant effects in Analysis Set IIA, when additional fixed factors were analysed. However, when the adults’ data were included (Analysis Set IIB) the age of the participants showed a significant effect on first fixation duration (t = -2.25) similar to the result of Analysis IB (see Table 50 for details). As can be seen in Figure 50, the younger children had the longest duration first fixation, and there were negligible differences between the older children and adult participants. In addition, Figure 50 shows that first fixation duration for the poor readers was always longer than good readers in every age group, but this was not significant.
Table 50. Mixed-effects models results of child and adult participants’ first fixation duration on target word (Analysis IIB)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>245.40415</td>
<td>21.24073</td>
<td>11.553</td>
</tr>
<tr>
<td>age.L</td>
<td>-22.78656</td>
<td>10.12819</td>
<td>-2.250</td>
</tr>
</tbody>
</table>

Figure 50. First fixation duration as a function of age and ability

Gaze Duration

Mixed-effect models analysis for results of Analysis Set IA is shown in Table 51. There were significant interaction effects of reading ability by spacing (t = -2.004) and start character frequency by end character frequency (t = -2.278).
As shown in Figure 51, the differences between gaze duration of good and poor reader children were greater when reading spaced texts. This might be because good readers made more fixations on the target words before moving on to the next word. Figure 52 shows that, generally speaking, the combination of higher start and higher end character frequencies allowed faster word recognition as shown by shorter gaze duration, although the effect is more linear for start character frequency.

In Analysis Set IB when the adults were added, there were similar results to analyses of first fixation duration – only age was significant (t= 3.06), and there were no significant frequency effects (see Table 52). As shown in Figure 53, gaze duration of younger child was significantly longer than those of older children and adults.
Figure 51. Gaze duration as a function of reading ability and spacing

Figure 52. Gaze duration as a function of start and end character frequencies
Figure 53. Gaze duration as a function of age group

The results of mixed-effects models analysis in Analysis Set IIA showed no significant effects. However, Analysis Set IIB showed significant effects of age ($t = 3.157$) as well as the interaction of spacing x end character frequency ($t = -2.175$) as shown in Table 53 below. As can be seen in Figure 54, gaze duration was marginally longer on target words that end with medium character frequencies when reading spaced texts, although the differences between the three frequency levels were not great. However, when reading unspaced text, gaze duration was clearly longest when reading target word ending with low frequency end characters. It is as if there were extra fixations when readers were unsure whether a word had ended or not (due to a low frequency end character), and shorter when reading words ending in high or medium frequency.
Table 53. Mixed-effects models results of child and adult participants’
gaze duration on target word

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>257.3575</td>
<td>10.6015</td>
<td>24.276</td>
</tr>
<tr>
<td>age.L</td>
<td>42.9229</td>
<td>13.5945</td>
<td>3.157</td>
</tr>
</tbody>
</table>

Figure 54. Gaze duration as a function of spacing and end character frequency

5.3.2. Analysis Set III (Sentences – Additional fixed effects)

In these analyses, eye movements on all the words in all sentences were analysed. There were two parts in this Analysis Set III as well, that is IIIA analysed the children’s eye movements only, while IIIB included the adults’ data in the analyses. Mean landing site location, first fixation duration and gaze duration of younger children, older children and adults are shown in Table 54. The mean word length of sentences in Stimulus Set 1 was 3.32 characters and 3.92 characters for sentences in Stimulus Set 2.
Table 54. Mean landing site location, first fixation duration and gaze duration for words in sentences of participants in Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Spaced</th>
<th>Unspaced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Younger children</td>
<td>Older children</td>
</tr>
<tr>
<td>Landing Site Location</td>
<td>1.97</td>
<td>2.25</td>
</tr>
<tr>
<td>(characters)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Fixation Duration</td>
<td>277.73</td>
<td>227.37</td>
</tr>
<tr>
<td>(msec.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze Duration</td>
<td>282.20</td>
<td>226.69</td>
</tr>
<tr>
<td>(msec.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Mean target word length in for Stimulus Set 1 was 3.32 characters and mean target word length of Stimulus Set 2 was 3.92 characters.

The overall results of Analysis Set III were similar to those of Analysis Set I and II. From Table 54 above, it is cleared that fixation time of participants both first fixation and gaze durations decreased as the age increased. Basically, fixation duration was longer when participants read unspaced text. Landing site location of younger children was further to the left of word centre than those of older children and adults. Older participants landing site was closer to the PVL (left to word-centre) when spaces between words were presented. The results of the mixed-effect models analyses for the three dependent variables are set out below.

**Landing Site Location**

The results of mixed-effect models of Analysis Set IIIA are shown in Table 55 and IIIB in Table 56 below.

In Analysis Set IIIA, there were significant main effects of age group (t = 3.918), start character frequencies (t = 2.346), middle character frequencies (t = 3.009), and word length (t = 5.715) on landing site of children participants. There were also interaction effects of age group by start character frequencies (t = -2.380); age group
and word length (t = -2.925); reading ability and start character frequencies (t = 2.233); spacing by middle character frequencies (t = -2.082); and middle character frequencies and word length (t = -3.212).

Table 55. **Mixed-effect models results of child participants’ landing site location (sentences)**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.204</td>
<td>0.326</td>
<td>0.625</td>
</tr>
<tr>
<td>age.L</td>
<td>0.646</td>
<td>0.165</td>
<td>3.918</td>
</tr>
<tr>
<td>first</td>
<td>0.294</td>
<td>0.125</td>
<td>2.346</td>
</tr>
<tr>
<td>mid</td>
<td>1.719</td>
<td>0.571</td>
<td>3.009</td>
</tr>
<tr>
<td>wlength</td>
<td>0.459</td>
<td>0.080</td>
<td>5.715</td>
</tr>
<tr>
<td>age.L:first</td>
<td>-0.326</td>
<td>0.137</td>
<td>-2.380</td>
</tr>
<tr>
<td>age.L:wlength</td>
<td>-0.091</td>
<td>0.031</td>
<td>-2.925</td>
</tr>
<tr>
<td>ability.L:first</td>
<td>0.259</td>
<td>0.116</td>
<td>2.233</td>
</tr>
<tr>
<td>spacing.L:mid</td>
<td>-0.308</td>
<td>0.148</td>
<td>-2.082</td>
</tr>
<tr>
<td>mid:wlength</td>
<td>-0.483</td>
<td>0.150</td>
<td>-3.212</td>
</tr>
</tbody>
</table>

The results of Analysis Set IIIA are illustrated in Figure 55 to Figure 59 below. Figure 55 shows that higher start character frequency allowed children, especially the older children, to land their eyes closer to the word centre. In addition, Figure 56 shows that landing site location was further into the words as the word length was longer and this again was more pronounced for older children. Start character frequencies became more important in guiding children’s eye movement to land closer to the word centre especially good readers (see Figure 57) and higher middle character frequencies still caused the reverse effects to pull the eyes to land closer to the word-initial area as found in the results of the adults’ eye movements (see Figure 58 and Figure 59).
Figure 55. Landing site as a function of age group and first character frequency

Figure 56. Landing site as a function of age group and word length
Figure 57. Landing site as a function of reading ability and first character frequency

Figure 58. Landing site as a function of middle character frequency and spacing
Turning to Analysis Set IIIB (see Table 56), there were main effects of age ($t = 2.477$ [Linear], $t = -3.177$ [Quadratic]), start character frequency ($t = 2.803$) and word length ($t = 11.768$). The interactions of age x word length as well as of ability x start character frequency were also significant.

With respect to the age x word length interaction, generally speaking, the landing site location of the younger children group was a little bit further to the left of the word centre except for the 7- and 8-character-words. On the other hand, adults’ landing sites were further into the word as word length increased, as shown in Figure 60.
Table 56. Mixed-effect models results of child and adult participants’ landing site location

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.15057</td>
<td>0.10274</td>
</tr>
<tr>
<td>age.L</td>
<td>0.30550</td>
<td>0.12334</td>
</tr>
<tr>
<td>age.Q</td>
<td>-0.34438</td>
<td>0.10839</td>
</tr>
<tr>
<td>first</td>
<td>0.23740</td>
<td>0.08470</td>
</tr>
<tr>
<td>wlength</td>
<td>0.22069</td>
<td>0.01875</td>
</tr>
<tr>
<td>age.Q:wlength</td>
<td>0.08285</td>
<td>0.02253</td>
</tr>
<tr>
<td>ability.L:first</td>
<td>0.13578</td>
<td>0.06686</td>
</tr>
</tbody>
</table>

Figure 60. Landing site as a function of word length and age
Figure 61 shows that high frequency start character frequency assisted both good and poor readers to land further into the word. However, the difference between landing site location of high and low frequency of start character was greater for good readers group, i.e., the good readers were better able to make use of this cue.

![Figure 61. Landing site as a function of reading ability and start character frequency](image)

**First Fixation Duration**

There were no significant results in Analysis Set IIIA for first fixation duration of child participants. However, when adding the adults’ data – the mixed-effect models analyses found that age of participants ($t = -3.10$) and word frequency ($t = -2.20$) had significant effects on first fixation duration (refer to Table 57 for details). First fixation duration of younger children group was always longest and the duration decreased as age increased. As shown in Figure 62 the first fixation of older children group was
almost identical to the adults. In addition, higher word frequency generally reduced first fixation duration, and while the interaction was not significant, this is evident mainly for younger children.

**Table 57. Mixed-effect models results of child and adult participants’ on first fixation duration**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>235.838</td>
<td>5.457</td>
<td>43.22</td>
</tr>
<tr>
<td>age.L</td>
<td>-25.345</td>
<td>8.170</td>
<td>-3.10</td>
</tr>
<tr>
<td>wordfreq.L</td>
<td>-5.444</td>
<td>2.480</td>
<td>-2.20</td>
</tr>
</tbody>
</table>

**Figure 62. First fixation duration as a function of word frequency and age**

**Gaze Duration**

In Analysis Set IIIA (see Table 58), the mixed-effects models analyses showed that only the main effect of word frequency was significant (t = -4.665). It also showed
interaction effects between age group and word frequency ($t = -4.512$) and between reading ability and word frequency ($t = -2.245$). As shown in Figure 63, younger child participants’ gaze duration was always longer than those of older children and the age differences were more pronounced when the word frequency was low. Regarding reading ability and word frequency effects, it seemed that the when the word frequency was low, good child readers fixated words longer than poor readers, but the effects were reversed when word frequency was high (see Figure 64). This may be due to the fact that good readers try to finish the word recognition process on the fixated word before moving on to the next word. If so then for good readers, higher frequency words, more familiar words, will take less time to process resulting in shorter fixation times. On the other hand, poor readers may move their eyes to the next word before word recognition is complete, thus resulting in more refixations.

Table 58. **Mixed-effect models results of child and adult participants’ on gaze duration**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>248.357</td>
<td>12.782</td>
<td>19.431</td>
</tr>
<tr>
<td>lrwfreq.L</td>
<td>-17.646</td>
<td>3.783</td>
<td>-4.665</td>
</tr>
<tr>
<td>age.L:lrwfreq.L</td>
<td>-20.034</td>
<td>4.441</td>
<td>-4.512</td>
</tr>
</tbody>
</table>
Figure 63. Gaze duration as a function of age group and word frequency

Figure 64. Gaze duration as a function of reading ability and word frequency
Gaze duration mixed-effects models analyses (Analysis Set IIIB) showed that age (t= 3.62), space condition of the texts (t= 2.36), and word frequency (t= -5.03) had significant main effects. The results also showed significant interaction effects between age and word frequency (refer details in Table 59).

Table 59. **Mixed-effect models results of child and adult participants’ on gaze duration**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>239.705</td>
<td>6.069</td>
<td>39.50</td>
</tr>
<tr>
<td>age.L</td>
<td>32.994</td>
<td>9.111</td>
<td>3.62</td>
</tr>
<tr>
<td>spacing.L</td>
<td>5.637</td>
<td>2.385</td>
<td>2.36</td>
</tr>
<tr>
<td>wordfreq.L</td>
<td>-12.458</td>
<td>2.477</td>
<td>-5.03</td>
</tr>
<tr>
<td>age.L:wordfreq.L</td>
<td>-17.852</td>
<td>2.959</td>
<td>-6.03</td>
</tr>
<tr>
<td>age.Q:wordfreq.L</td>
<td>-12.788</td>
<td>2.825</td>
<td>-4.53</td>
</tr>
</tbody>
</table>

Figure 65 shows that participants read spaced text faster than unspaced except the younger children group. This proved that spaces between words facilitate skilled readers to fasten word recognition process. The possible explanation why the younger children take longer time fixate words in spaced condition was that they finished word recognition process on the words before move on to the next word since the word boundaries were clear, but in the unspaced condition they might move the eyes to the next words before the process is completed resulting in shorter gaze duration but more regression when reading unspaced texts.
Figure 65.  Gaze duration as a function of spacing and age group

Figure 66 shows, as expected, that the gaze duration of younger children was longest compared to the older participants. Similar to the results for first fixation duration, gaze duration did not differ much between older children and adults. High word frequency facilitated shorter gaze durations at every age and low frequency words had the greatest impact on reading speed of younger children as indicated by longer gaze duration.
5.4. Discussion

The results of the study in this chapter show that the eye movement control of Thai younger children readers is similar to those of the adults only in terms of landing site location, whereas the older children’s eye movements were similar to the adults in all three aspects – landing site location, first fixation and gaze durations. Children also targeted word-centre as a landing site for their eyes when reading, even though the landing site was slightly further to the left of the centre than those of the adults. Generally speaking, landing site location of Thai readers was closer to the OVP (word-centre) as their age and reading skill increased.

Fixation time, both first fixation and gaze durations, of older children was similar to those of the adults – they fixated words in the unspaced condition longer. The results showed that it was the good younger children readers’ fixation time that caused these results not the poor readers. The reasons for this need further
investigation, but a possible explanation is that the visible word boundaries in spaced
text allowed the children, especially good readers, to finish the word recognition
process on the fixated word before moved on to the next word, thus resulting in longer
fixation times. In contrast, when spaces were not available and the children had to scan
for the word boundaries from the text – children may move the eye away from fixated
words before the word recognition process is finished resulting in shorter fixation time
but causing more regressions and thus longer gaze duration.

The results of children’s and adults’ eye movements in this Experiment 3
confirmed that the PVL of Thai readers of three different ages is near word centre.
Both child and adult participants’ landing site location was always close to this PVL.
For normal Thai text (unspaced) the landing site of younger children was a bit further
to the left of the word centre than that for the older children and adults; nevertheless,
their landing site was too far into the word to be designated as being in the word
beginning area. From the results of spacing on landing site location, it can be seen that
spaces between words facilitated word targeting during reading especially for the
younger children. Indeed, there was greater difference between landing site location on
spaced and unspaced target words for younger children than those of the older children
or adults. There was also an interaction between spacing and end character frequency
for landing site location; participants’ eyes landed closer to the word centre except
when reading target words ending with low frequency characters in the spaced
conditions. It seems that reliance on end frequency in the spaced conditions may lead
to uncertainty about word ending in the low end frequency condition resulting in
overshoot.
First fixation duration decreased as the age of participants increased and only the interaction effects of start character frequencies by word length had any effect on first fixation duration in child participants. Basically, lower start character frequencies caused longer first fixation durations except in short words.

Turning to gaze duration, the results on target words showed that the interaction between reading ability and spacing condition as well as the interaction effects of start and end character frequencies was significant. Gaze duration on target words was lower if the target word had higher frequency start and end characters. However, the results of gaze duration regarding the effects of spacing and reading ability seem to be reversed compared to those found in the adults. Good reader child participants’ gaze durations were always longer than poor reader groups and the differences were even greater when there were spaces between words.

As mentioned earlier, the results in this study seem to show that eye movement patterns of native Thai children and adults are similar. There were slight developmental trends in eye movement patterns and eye movement control with respect to the primary features driving landing site – start character frequency, end character frequency and middle character frequency. From these results, it is possible to conclude that end character frequency played a more important role in eye movement control of younger children and these effects decreased as the age (thus reading skills) of the participants increased. On the other hand, the effects of start character frequencies became stronger over age. Spacing appeared to have a very strong effect on the eye movements of younger children pushing the landing site further into the word, but it simply facilitated reduced first fixation and gaze durations of the older children and adults.
Overall, the presence or absence of spaces between words still did not cause any major shift in eye movements of Thai readers. The PVL of both children and adults was still around the word centre area regardless of spacing conditions. However, before it can be confidently concluded that there was no developmental trend for landing site location on eye movements in reading Thai text, eye movements in reading of one more group of participants (much younger, beginning readers, e.g. first and second grade students) need to be investigated. This was not possible here because of the size of the EyeLink II apparatus, but it is hoped that improved technology may overcome this limitation in the near future so that younger children with smaller heads can be tested.
CHAPTER SIX

CONCLUSION, DISCUSSION AND SUGGESTIONS FOR FUTURE STUDIES
6. CONCLUSION, DISCUSSION AND SUGGESTIONS FOR FUTURE STUDIES

The main objective of this thesis was to investigate low-level visual information of distinctive features in Thai orthography that might allow Thai readers, both children and adults to read such an alphabetic scriptio continua language fluently. More specifically the aim was to investigate what features provide visual cues in guiding eye movements of Thai readers, that is their preferred viewing location (PVL), as well as whether Thai readers’ oculomotor control is different when reading spaced versus unspaced texts, silently and aloud. In order to do so, three experiments were conducted. The first experiment (Experiment 1) was a study of reading time and reading accuracy across different age groups testing the effects of relative frequency of occurrence of start and end characters which were selected on the basis of previous studies (Reilly et al., 2003). Experiments 2 and 3 were eye movement studies using the same orthographic features (relative frequency of start and end characters), as the independent variable to systematically manipulate the test stimuli. Experiment 2 investigated eye movements of Thai adults when reading spaced and unspaced text silently and aloud and Experiment 3 studied eye movements of Thai children and adults also reading spaced and unspaced texts but only when reading silently. The overall results of each experiment are described in turn in the following sections.

6.1. Conclusions for Experiments 1, 2 and 3

6.1.1. Experiment 1: Reading time studies

This experiment investigated the effects of relative frequency of characters in the word-start and word-end positions on reading time and reading accuracy of Thai participants. These characters could be either consonants or vowels as long as they
were written in the main horizontal line (refer Section 3.2 for more details). In this experiment, the target words comprising the distinctive features were embedded at or near the centre of the sentence frames. Relative frequency of occurrence of character in the start and end positions of the target words each had three levels of frequency (High, Medium or Low) in each position. The factorial design of was 5 age groups x 2 reading ability levels x (3 start character frequencies x 3 end character frequencies x 2 spacing conditions of text).

There were five age groups of participants, Grade 1, 2, 5, and 6 students plus adults as a control group. Participants in each age group were separated into two subgroups, good and poor readers, depending on their reading ability level (refer details of screening and recruitment process in Section 3.2). Participants read all the stimuli aloud and from recordings of their reading, reading time, error rate and error types were derived. The results of Experiment 1 are summarised in terms of reading time and reading accuracy below.

**Reading Time**

Overall the results showed that Thai participants’ reading time and errors decreased as their age and reading skill increased. The differences between reading time of good and poor readers also narrowed as the age of the participants increased. Participants of all ages read spaced text sentences faster than unspaced sentences even though the advantage of spaces between words most facilitated the reading of younger poor readers, especially 1st and 2nd graders. Higher relative frequencies of characters especially, word-start characters allowed the participants to read words more quickly. The effect of start character frequency contributed to decreased reading time in both spacing conditions of the text, whereas the effects of end character frequency was
stronger in the spaced condition. The effects of start character frequency on reading
time of child and adult participants were similar in this study but the effects of end
character seemed to reverse over age: higher frequency end characters reduced reading
time of younger children but increased reading of the older children and adults. This
may be due to lexical effect since these two groups of participants had a larger lexicon,
the higher the frequency of occurrences of the characters the higher the number of
possible word that ended with that character.

Reading Accuracy

There were three types of errors found in this experiment; phonological,
orthographic, and semantic errors. The results showed that error rate decreased as a
function of age. Participants at every age made more phonological (especially tone)
errors than orthographic or semantic errors. Poor readers always made more errors than
good readers.

In conclusion, the results of Experiment 1 show that younger participants can
read unspaced texts without much difficulty even though spaces between words
facilitate their reading by decreasing their overall reading time. The results showed that
character frequencies, especially those in word-start position, facilitate reading times
of both child and adult participants. Whether such differences are reflected in eye
movements was investigated in Experiments 2 and 3.

6.1.2. Experiment 2: Eye Movements of Native Thai Adults
Reading Silently and Aloud

In the first eye movement experiment the eye movements of 30 native Thai
adults (15 Hi-Ed and 15 P/Mid-Ed readers) reading silently or aloud with unspaced or
spaced text was investigated. The stimuli used in this experiment were sentences containing target words with different relative frequency of characters in the start and end position. The analysis was divided into three sets of analyses depending whether the main focus was on the eye movement data, on the target words, or the sentences and which frequency information was used (refer details in Section 4.2). The three dependent variables were landing site location, first fixation duration and gaze duration. The overall results showed that adult Thai readers have a preferred viewing location at the left of word-centre, similar to English readers. Spaces between words facilitated reading speed, as shown by shorter gaze duration, but did not interfere with oculomotor control of Thai readers, i.e., their eye movements did not change dramatically when reading spaced versus unspaced texts. When reading aloud, participants’ landing site moved slightly further to the left of the PVL and first fixation and gaze duration was also longer when reading aloud. The results of landing site location, first fixation duration and gaze duration are described in more detail below.

**Landing Site Location**

Landing site location of Thai adults is near the centre of the word irrespective of target word conditions or spacing of the text. The results from all three analyses (refer 4.3 for details) showed that mode of reading had significant effects on landing site location [(t= 3.139) in Analysis Set I, (t= 3.081) Analysis Set II, and (t= 4.938) Analysis Set III]. That is when reading silently the participants’ eyes landed further into the words than when reading aloud, and first fixation and gaze duration were longer.

When considering only the landing sites on target words, it was found that the eyes landed further into the word when the word was longer. This clearly indicates that
readers tend to target the word centre, with their landing sites correspondingly shifting as word length increased. Considering the relative frequency of characters, the results showed a significant effect of start character frequency (Analysis Sets I and III) and middle character frequency (Analysis Set II). Higher start character frequencies, as expected, drove the eyes further into the word. In contrast, high frequency middle characters in Analysis Set II pulled the eyes further to the left of the word centre (refer details in Section 4.3). Analysis Set I also showed some significant interaction effects between start character frequency x education level of participants, word length, and end character, but overall higher start character frequency was effective, over these variable in directing eye movements toward the centre of words.

**First Fixation Duration**

The results of all three analyses showed that modes of reading and word frequency had significant effects on first fixation duration. Generally speaking, fixation duration decreased if the word frequency was high and was greater when reading aloud than reading silently. The effect of word frequency interacted with spacing in the word in sentence analysis (Analysis Set III); first fixation duration was shorter when reading high frequency words in spaced than in unspaced text.

**Gaze Duration**

Gaze duration was generally longer when reading aloud than when reading silently. Analysis Set I also showed that the interaction between modes of reading and end character frequency was significant, gaze duration was shorter when reading words ending with higher end character frequencies suggesting that high end character
frequencies helped signal the end of words and so immunized readers against further refixations.

6.1.3. Experiment 3: Eye movements of Thai children and adults reading silently

Eye movements of two groups of native Thai children (refer to 5.2), when reading spaced and unspaced texts silently, showed that landing site of the child participants was also located near the word-centre (PVL) similar to that of adults. Even though the younger children tended to land their eyes further to the left of the PVL, the site was too far into the word to be designated as the word-initial area. Viewing time, as indicated by first fixation duration and gaze duration decreased as the age of the participants increased. Participants’ fixation durations when reading unspaced text were longer than for spaced text except for younger children. The results were discussed in more details below.

Landing Site Location

Word length had a significant effect on landing site location both in target word and words in sentences analyses, i.e. landing site was further into the word as word length increased. The analyses of children’s landing site location on target words in Analysis Set I also showed a significant effect of spacing, but no effect of character frequency. However, when the adults’ eye movements were included, the interaction effects between spacing by end character frequency became significant; the effects of end character frequency appeared to be more pronounced in unspaced conditions with higher frequency end characters allowing participants to land their eyes closer to the
There was always a significant interaction effect of age by spacing condition on landing site location of child readers as the differences between landing site of younger children when reading spaced and unspaced text were greater than those of older children. In Analysis Set II middle characters also had the same effects on landing site location as in Experiment 2; high frequency middle characters pulled the landing site back toward the beginning area of word.

The effects of start character frequency turned out to be significant in the sentence level analysis. In keeping with the results for Experiment 2, landing site was further into the word if the relative frequency of the start characters were high. Middle character also had significant effect on landing site location on the sentence same way as on target word analysis. The interaction effects of reading ability by frequency of start character were also significant, i.e. higher start character frequency allowed good readers to land their eyes further into the word.

**First Fixation Duration**

First fixation duration of younger children was longer than the other two age groups while durations for the older child group was very similar to those of the adults. There was a significant effect of end character frequency as well as interaction effect between end character frequency and word length on first fixation duration of children. The results showed that children’s first fixation duration was longest only when fixating longer (≥ 5 characters) target words. However, all these effects were not apparent when the adults’ data were included in the analysis; then only the effect of age group was significant. The results of Analysis Set III also showed a significant effect of word frequency as well as age group on participants’ first fixation duration.
Gaze Duration

The results of children’s gaze duration in Analysis Set I showed the significant interaction effects of reading ability by spacing and start character by end character frequency, but there were no significant effects in Analysis Set II. When including adults’ data in the analysis, both analyses showed main effect of age group with additional interaction between spacing and end character frequency in Analysis Set II. Turning to gaze duration on the sentence level, the results showed that effect of word frequency was significant both the main effect and interaction effects by age group and spacing. Higher word frequency allowed children readers to recognise the word faster especially those with less reading skills. The interaction effects between age group and word frequency remained significant when adding the adults’ data in the analysis.

6.2. Discussion

The main finding in this thesis was that the preferred viewing location (PVL) of native readers of Thai, an alphabetic scriptio continua language, is near the word centre similar to the PVL of the readers in spaced languages. Since the centre of the word already is the most effective saccadic target in Thai, putting spaces between words in Thai text does not improve the landing site of Thai readers. This suggests that even though there are no spaces to serve as a visual cue in guiding eye movements, Thai readers can use other sources of guidance for eye movements. These cues turn out to be the position-specific frequency of characters especially at the start of words. While end character frequencies have some effects, and while middle character frequencies emerged in Analysis Set II, by far the overriding effect, both when the predetermined factors were used in analysis (Analysis Set I) and when a large sample
of words were analysed thus greatly improving the power of the analysis (Analysis Set III), was of start character frequency.

Mode of reading had a strong effect especially in viewing time (first fixation duration and gaze duration) and landing site location of participants was a bit nearer to word beginning when reading aloud. Huestegge, Radach, Corbic, and Heustegge (2009) also found similar results - that landing site was nearer to the word beginning during both careful reading and reading aloud. First fixation duration as well as gaze duration was longer when participants read aloud (refer Section 4.3 for details).

Spaces between words did not have a major effect on landing site distributions of adult Thai readers. Spaces, however, served to decrease viewing time as shown by shorter fixation durations. The results of adults’ and older children’s fixation durations (both first fixation and gaze duration) showed that duration decreased when spaces were available. The results agree with the reading time study of spacing effects in Thai (Kohsom et al. 1997), who found that reading time was faster when spaces were present. Similarly, the results of a reading time studies in this thesis (Experiment 1) also showed that reading time when reading spaced text was faster across all ages even though the spaced-unspaced difference decreased as the participants’ reading skills increased. Therefore, from the results of three studies (one reading time and two eye movement studies), it is possible to conclude that reading spaced text results in faster reading. This is presumably because of a faster word recognition process due to better word visibility. Together with the landing site results in which there are no effects of spacing, it is therefore possible to conclude that faster reading in spaced text is due to faster word recognition rather than failure to land near the OVP in unspaced text reading.
Turning to the eye movements of child participants, the results also showed that the PVL of children was at or near the word centre even though it was further to the left of the word-centre in the unspaced text condition. Spaces between words had significant effects on landing site location of Thai children especially the younger readers. Spaces helped facilitating the eye guidance to land closer to PVL. The relative frequency of characters had no significant effects on landing site location of the children. However, when analyses were expanded to cover the whole sentence (thus increasing power), the frequency of start and middle characters became a significant factor on children’s landing site location. Higher start character frequency pulled the eyes to land further into the word and closer to the OVP while high middle character still pulled the eye backward to land closer to the initial area in the same manner as that found in the adults’ eye movements.

Turning to the fixation time of the children, the results showed the reverse effects of spaces between words on the fixation time of younger children. Young children tended to fixate longer when spaces were presented. This was true for both first fixation and gaze duration results. The reasons behind this are still unclear and need to be investigated, but one possible explanation was that the children fixated longer when reading spaced text because they completed the word recognition process of the fixated word before moving onto the next word, since word boundaries were clearly visible. On the other hand, they might move their eyes from the fixated word before word recognition is complete to scan for the word boundaries in the unspaced condition, resulting in shorter fixation time and more regressions.

The fact that the effects of spacing ceased to be significant when the readers’ skills increased and also the stronger effects of frequency of characters on children’s
eye movement when reading unspaced text, supports the conclusion that it is the frequency of characters not spaces between words that are critical factors in word targeting and in reading Thai, even though the character frequency effects were relatively stronger for the skilled readers’ eye movements. These results conform to the expectation that the spacing condition of the text is a more low-level visual cue that does not need to be learned while information about the relative frequency of character effects needs to be learned. The character frequency effects here for both children and adults also provide the further support for the role of character frequencies in word targeting (Reilly et al., 2003) and suggest that they are not just a proxy for some other variables.

From the overall results of the eye movement studies in this thesis, it is clear that low-level visual information such as the relative frequency of especially start characters are used as visual cues in word targeting during reading Thai. Spaces between words do not have a significant role in guiding eye movements in reading Thai when the readers have acquired better reading skills. They simply facilitate reading time and allow the readers to process words more quickly. However, from the results that spaces between words have significant effects on landing site location and fixation time especially on young children it can be concluded that there is a developmental trend in eye movement patterns in reading Thai. Less skilled readers rely more on a visible cue such as spaces before turning to the information of the text when these cues are not available. This may due to the fact that literacy lessons in Thai start with spaces text therefore less skilled readers tend to seek the primary visual cue when reading and are able to execute better oculomotor control when they are available.
Models of Reading

As suggested previously in 2.4 it appears that the Glenmore model may potentially provide the best account of eye movements in reading Thai. The results found in Experiments 2 and 3 are similar to the predictions of Glenmore model (refer 2.4.3) for eye movement control in reading scriptio continua languages. Those predictions were i) landing site location should rely on the saliency of the letter and word being fixated and the landing site should be further into the word if frequency of the start and end characters of the word are high, and ii) fixation duration and gaze duration should be associated with word length and word frequency. The results of reading studies with Thai participants reading Thai confirm both these predictions. The results from eye movement studies in this thesis should provide a good focus for extending the modelling of eye movements in reading alphabetic languages. It seems that for the readers of scriptio continua languages, there are no major eye movement shifts even though spaces between words are present in the text. Therefore, the models with fundamental architecture that have interactive activation design principles such as the Glenmore model should allow accommodation of eye movement control during reading both spaced and scriptio continua languages.

6.3. Suggestion for Future Studies

The distinctive features used in these eye movement studies are just one of many low-level visual features in the Thai language. There are some other distinctive features that still need to be investigated with respect to the effects of eye movement patterns and control. For example, Thai allows vowel and tone characters as well as other diacritics to be placed vertically above or below the main horizontal line and it is
of interest to investigate whether the presence or absence of these vertical characters in the text will affect the eye movement patterns of Thai readers or not.

From the fact that there were slightly different effects of spaces on the oculomotor control of the adults and children found here, it appears that there is a developmental trend in eye movements in reading Thai regarding both landing site location and fixation time. However, before this can be confirmed, eye movements in reading of one more age, Grade 1 and 2 students who are beginning readers, should be investigated. The possible role of middle character frequencies found in one set of analyses here need to be investigated in studies in which start, middle and end position-specific frequencies are systematically manipulated (rather than the systematic manipulations of only start and end character frequencies here). With these further studies the effects of the low-level factors involved in reading Thai can be further elaborated. For now, however, it can be concluded that, despite the lack of spaces between words in Thai, Thai readers use at least the word-start start character frequency to direct their eyes to that same position in the word as do readers of spaced texts, just to the left of the centre of the word. Moreover, Thai readers also use the frequency of end characters to identify the word end and thus terminate the word recognition process.

The results of the eye movement studies in this thesis show no major shift in the eye movement control of scriptio continua alphabetic language readers; spaces simply serve to perfect the execution of oculomotor control. Nevertheless, more eye movement studies on reading other scriptio continua alphabetic languages such as other South East Asian languages like Lao and Cambodian/Khmer need to be
investigated before these results can be generalised to all scriptio continua alphabetic texts.
CHAPTER SEVEN

REFERENCES
7. REFERENCES


Biscaldi, Monica, Gezech, Stefan, & Stuhr, Volker. (1998). Poor saccadic control correlates with dyslexia. Neuropsychologia 36, 1189-1202


De Luca, Maria, Borrelli, Marta, Judica, Anna, Spinelli, Donatella, & Zoccolotti, Pierluigi. (2002). Reading words and pseudowords: an eye movement study of development dyslexia. Brain and Language 80, 617-626


Hawelka, Stefan, & Wimmer, Heinz. (2005). Impaired visual processing of multi-element arrays is associated with increased number of eye movements in dyslexic reading. Vision Research 45, 855-863


APPENDICES
Appendix A: Tone realisations and syllable structures

Table A1: Open syllable with long vowel (C$_1$V: or C$_1$VV:)

<table>
<thead>
<tr>
<th>Initial</th>
<th>Tone</th>
<th>Mid level</th>
<th>Low level</th>
<th>Falling</th>
<th>High level</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>High class</td>
<td>-</td>
<td>ถ่อ [pʰ:]</td>
<td>ถ่อ [pʰ:]</td>
<td>-</td>
<td>ถ่อ [pʰ:]</td>
<td></td>
</tr>
<tr>
<td>Mid class</td>
<td>ปอ [pʰ:]</td>
<td>ป ่ อ [pʰ:]</td>
<td>ป  อ [pʰ:]</td>
<td>ป  อ [pʰ:]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low class</td>
<td>ทอ [pʰ:]</td>
<td>-</td>
<td>ท่อ [pʰ:]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A2: Open syllable with short vowel (C$_1$V, C$_1$VV, C$_1$VɁ or C$_1$VVɁ)

<table>
<thead>
<tr>
<th>Initial</th>
<th>Tone</th>
<th>Mid level</th>
<th>Low level</th>
<th>Falling</th>
<th>High level</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>High class</td>
<td>-</td>
<td>ฉะ [ʨʰà]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mid class</td>
<td>จะ [ʨà]</td>
<td>จ ะ [ʨâ]</td>
<td>จ ะ [ʨá]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low class</td>
<td>ค่ะ [kʰâ]</td>
<td>-</td>
<td>คะ [kʰá]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A3: Closed syllable ending with /m, n, ŋ, w, j/ or live syllable (C$_1$VC$_f$[sonorant], C$_1$VVC$_f$[sonorant], C$_1$V: C$_f$[sonorant] or C$_1$VV: C$_f$[sonorant])

<table>
<thead>
<tr>
<th>Initial</th>
<th>Tone</th>
<th>Mid level</th>
<th>Low level</th>
<th>Falling</th>
<th>High level</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>High class</td>
<td>-</td>
<td>ข่าง [kʰà:ŋ]</td>
<td>ข าง [kʰâ:ŋ]</td>
<td>ขาง [kʰǎ:ŋ]</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Low class</td>
<td>คาง [kʰā:ŋ]</td>
<td>-</td>
<td>ค่าง [kʰâ:ŋ]</td>
<td>ค าง [kʰá:ŋ]</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table A4: Closed syllable with short vowel and ending with /p, t, k/ or short dead syllable (C$_1$VC$_f$[stop], C$_1$VVC$_f$[stop])

<table>
<thead>
<tr>
<th>Initial</th>
<th>Tone</th>
<th>Mid level</th>
<th>Low level</th>
<th>Falling</th>
<th>High level</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>High class</td>
<td>-</td>
<td>ขาบ [kʰà:p]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mid class</td>
<td>กาบ [kà:p]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low class</td>
<td>ค ้ บ [kʰá:p]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A5: Closed syllable with long vowel and ending with /p, t, k/ or long dead syllable (C$_1$V: C$_f$[stop], C$_1$VV: C$_f$[stop])

<table>
<thead>
<tr>
<th>Initial</th>
<th>Tone</th>
<th>Mid level</th>
<th>Low level</th>
<th>Falling</th>
<th>High level</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>High class</td>
<td>-</td>
<td>ขาบ [kʰà:p]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mid class</td>
<td>กาบ [kà:p]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low class</td>
<td>คาบ [kʰâ:p]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B1A: Screening Reading Test (Grade 1 and 2)

ข้อสอบวัดผลความสามารถในการใช้ภาษาไทย

1. ที่วัดนี้อากาศร่มรื่น เพราะมีต้น.............อยู่มาก
da. ไซ b. ไทร c. ไซร d. ใทร

2. ฉันชอบปลูก...........เป็นผักสวนครัว
da. กะเพรา b. กระเพรา c. กระเพลา d. กะเพลา

3. คำในข้อใดถูกต้อง
da. พันธ์ผัก b. พรรณผัก c. พันธุ์ผัก d. พันผัก

4. “ว”ในคำใดเป็นตัวสะกด
da. หัว b. กล้า c. แมว d. กลัว

5. ข้อใดใช้สะกด“ไ”ผิด
da. คณิป b. หัวใจ c. ล้าไย

6. “ก-เอ-มา”เขียนเป็นคำได้อย่างไร
da. กอน b. เกอม c. แกน d. เกิน

7. คำว่า “พฤหัสบดี” อ่านว่าอย่างไร

8. “ผะ-หลิด-ตะ-ผน”ข้อใดถูกต้อง
da. ผะหลิดตะผล b. ผลิตผล c. ผลิตตะผล d. ผลิตตะผล

9. “เด็กฉลาดชอบเล่นสนุกอยู่เสมอ”มีคำอักษรนำที่คำมีค่าอักษรนำกี่คำ
da. 3 คำ b. 2 คำ c. 4 คำ d. 1 คำ

10. “รัฐบาล” อ่านออกเสียงกี่พยางค์
da. 3 พยางค์ b. 4 พยางค์
c. 2 พยางค์
d. 1 พยางค์
11. วันนี้อากาศ........... เราจึงช่วยกันเปิดหน้าต่าง
   เปิดหน้าต่าง
   a. ปลอดโปร่ง
   b. อบอวล
   c. ชุ่มชื้น
   d. อบอว
12. ปิดเทอมนี้ชั้นมี...........ได้ไปเที่ยวต่างประเทศ
   ต่างประเทศ
   a. โอกาส
   b. โอกาส
   c. เวลา
   d. โอกาส
13. กล้า...........เสื้อและกางเกงตัวใหม่
   ตัวใหม่
   a. สอด
   b. สวม
   c. ใส
   d. ทิ้ง
14. ช่วงไม่ใช...........ตัดไม้ในการสร้างบ้าน
   บ้าน
   a. ถนน
   b. ต้น
   c. ใส
   d. ทิ้ง
15. ชื่อช่วยแปลสัญลักษณ์.........ใช้รับประทาน
   เรียกชื่อ
   a. สายผม
   b. สายครัว
   c. สายป่า
   d. ข้างบ้าน
16. ผนึกปากให้พื้นดิน..............
   ผนึกปาก
   a. เบียร์บีร์
   b. เหลอะเพล
   c. เหลอะ
   d. แห้งแล้ง
17. เราต้องก้มลงไปดูผู้มีพระคุณไม่คิด............ท่าน
   ท่าน
   a. สอบครู
   b. อบดู
   c. สอบครู
   d. สอบครู
18. แซ่ย่านหนังสืออย่าง...........
   สื่อ:
   a. ชมเก๊ม
   b. ชมกุ๊ม
   c. ชมกิ๊ม
   d. ชมกิ๊ม
19. น้องหลั่ยมหัวเข่า...........
   หัวเข่า
   a. กลกล
   b. กลกล
   c. กลกล
   d. กลกล
20. พ่อแม่พาออกมาไปทำบู phậnที่........
   ทำบู phận
   a. โรงเรียน
   b. สวน
   c. วัด
   d. หมู่บ้าน
21. ข้อใดมีตัวสะกดเป็นมาตราเดียวกัน
   เรียกชื่อ
   a. พิษ, โรค
   b. ภาพ, ก้าง
   c. ทรง, กังวล
22. "ไซ" อ่านออกเสียงพยัญชนะต้น เหมือนคำใด
   a. สร้าง
   b. เสร้า
   c. ทราบ
d. เสา

23. คำใดยังเสียงวรรณยุกต์ได้อย่างอักษรกลาง
   a. ครู
   b. เสือ
c. ตะย
d. ตรง

24. คำใดใช้สระ "ออ" ดังรูป
   a. วัน
   b. พร
c. คด
d. เกิด

25. พยัญชนะชุดใด เป็นอักษรกลาง
   a. ก, ข, ค, ฅ
   b. ป, ม, พ
c. ฅ, ส, ศ, ษ,d. ต, ถ, ท

26. กลุ่มคำในข้อใด เป็นระดับเสียงสั้น
   a. แม่, โขง
   b. ช่วยกินบาง
c. ถมแรงมาก
d. ป่าหินน้ำ

27. คำในข้อใด มีตัวสะกดตามระ แม่
   กับ ทั้งหมด
   a. รูปภาพ
b. ใชคคลาน
c. ป้ากกรรรม
d. ตอบแทน

28. คำในข้อใดยู่ในมาตรา แม่กง
   ทุกคำ
a. งามาย
b. ตะกลอง
c. งวงย์
d. ติ่งตึง

29. ข้อใดมีเสียงวรรณยุกต์ สามัญ เอก โท ตามลำดับ
   เอก โท ตามลำดับ
a. ตาม, หา, พอ
   b. รัต, มั่น, นิภ้น
c. ลิฟนู, ติ่ง, ข้อ
d. หา, พอ, แม่

30. ข้อใดบอกเสียงวรรณยุกต์ของคำ ต่อไปนี้ถูกต้อง "วาดภาพน้ำตกหรือ"
   a. สามัญ สามัญ โท สามัญ จัตวา
b. โท โท โท สามัญ จัตวา
c. สามัญ สามัญ ตรี เอก จัตวา
d. โท โท ตรี เอก จัตวา

ข้อ 31–33 จงเลือกคำที่มีความหมายเข้ากับคำที่กำหนดให้
31. ปีน ..............
   a. โยน
   b. โหน
   c. ไต
   d. ขาม

32. ครู ..............
   a. ร้องเพลง
   b. สอน
   c. ดุ
   d. ตี

33. ออย ..............
   a. ไร
   b. นา
   c. สวน
   d. น้ำ

34. ค้าในข้อใดมีความหมายคล้ายกัน
   a. อย่นแอ – แข็งแกร่ง
   b. ยิ้มแย้ม – เศร้าซึม
   c. สุภาพ – อ่อนน้อม
   d. หัวเราะ – ร้องไห้

35. ค้าในข้อใดไม่เกี่ยวข้องกัน
   a. ไข – เป็ด
   b. ถุงเท้า – กระเป๋า
   c. ซุง – ต้นไม้
   d. น้ำ – ข้าว

36. ค้าคู่ใด มีความหมายตรงข้ามกัน
   a. สวย – นารัก
   b. เร็ว – ไว
   c. ใหญ่โต – กว้างขวาง
   d. กลางวัน – กลางคืน

37. ค้าในข้อใดไม่เข้าพวก
   a. ชาย
   b. ชาย
   c. ชาย
   d. ชาย

38. ค้าใดหมายถึง “วางเป็นระเบียบ”
   a. คัด
   b. จัด
   c. ตัด
   d. ขัด

39. “หนาแน่น” มีความหมายตรงข้ามกับค้าใด
   a. เบาบาง
   b. แออัด
   c. บอบบาง
   d. อึดอัด

40. “เจริญ” มีความหมายตรงข้ามกับค้าใด
   a. พัฒนา
   b. เสื่อม
   c. รุ่งเรือง
   d. ท่าวหน้า
Appendix B1B: Spelling Test (Grade 1 and 2)

ชื่อ.................................... นามสกุล....................................ชั้น.................เลขที่................

1. กระ-ดาด เขียนว่า _________________________
2. ราง-วัน เขียนว่า _________________________
3. วัน-พุด เขียนว่า _________________________
4. สัน-ยา เขียนว่า _________________________
5. บัน-ซี เขียนว่า _________________________
6. ฝน-ละ-ไม เขียนว่า _________________________
7. หน้า-รัก เขียนว่า _________________________
8. วาด-พาบ เขียนว่า _________________________
9. อา-กาด เขียนว่า _________________________
10. มา-ไล เขียนว่า _________________________
11. สุก-ชะ-พาบ เขียนว่า _________________________
12. น้ำ-ทาน เขียนว่า _________________________
13. รุบ-ซัง เขียนว่า _________________________
14. ตัว-เลก เขียนว่า _________________________
15. บูร-จัก เขียนว่า _________________________
16. ยัก เเขียนว่า _________________________
17. จะ-ทรงตุ เขียนว่า _________________________
18. ดวง-จัน เขียนว่า _________________________
19. สบ-คา เขียนว่า _________________________
20. หาด-ชาย เขียนว่า _________________________
ข้อสอบวัดผลความถนัดในการใช้ภาษาไทย

1. ข้อใดมีพยัญชนะดั้นเป็นอักษร
   ควบแท้ 3 อักษรน่า 3
   a. จริงหรือแปลงเป็นป่าสร้อย
      กำลังด่างควง
   b. หมอปล่อยก้ามป้าย
      ตกกระจายเกลื่อน
   c. ผลิตภัณฑ์นี้ซื้อจาก
      ตลาดนัดสนามหลวง
      เพราะถูก
   d. ไสวปลูกหญ้าและขนุนที่
      สนามเพื่อปิดทราย

2. คำควบกล้าข้อใดเขียนไม่ถูกต้อง
   a. ไขวหาง
   b. ขวักไขว
   c. ขวางปา
   d. กระกลาย

3. ข้อใดอย่างออกเสียงไม่ถูกต้อง
   a. คุณวุฒิ อ่านว่า คุณ-วุฒิ
   b. คุณธรรม อ่านว่า คุณ-ธรรม
   c. คุณโทษ อ่านว่า คุณ-โทษ
   d. คุณวิเศษ อ่านว่า คุณ-วิเศษ

4. “พระบรมมหาราชวัง” อ่านว่า
   อย่างไร
   a. พระ-บรม-มหาราช-วัง
   b. พระบรม-มหาราช-วัง
   c. พระบรม-มหาราช-วัง
   d. พระ-บรม-มหาราช-วัง

5. ข้อใดมีอักษรควบกล้ามากที่สุด
   a. ให้ชิงช้าย่างไก่ใน
      อากาศ
   b. ผลัดพลังกลั้งเกลียด
      พลัน
   c. เกลาตอมกล่าฉบับเรียบ
      เบรย์
   d. ใส่ชื่อแทน

6. “จตุพรนั่งฟังอยู่นานก็สัปหงกโงกเงก” ข้อความนี้มีค่าทั้งหมดกี่
   ค่า
   a. 7 ค่า
   b. 8 ค่า
   c. 10 ค่า
   d. 12 ค่า

7. ข้อใดมีจำนวนสะกดที่เท่ากับ
   “จตุพรนั่งฟังอยู่นานก็สัปหงกโงกเงก”
   a. สะพานพุทธยอดฟ้า
   b. สะพานทิพย์พิทักษ์
   c. สะพานพุทธยอดฟ้า
   d. สะพานพุทธยอดฟ้า
8. “ภูมิปัญญาเป็นจังหวัดหนึ่งของประเทศไทย” ข้อความนี้ประกอบด้วยกี่คำ กี่พยางค์
   a. 7 คำ 13 พยางค์
   b. 6 คำ 13 พยางค์
   c. 5 คำ 12 พยางค์
   d. 8 คำ 12 พยางค์

9. ข้อใดเป็นคำประสมทั้งหมด
   a. หัวเราะ, ลูกคิด
   b. แม่น้ำ, ว้าเหว
   c. ห้องแถว, แม่น้ำ
   d. นาฬิกา, อ่อนน้อม

10. คำในข้อใด ย่อได้ถูกต้อง
    a. โฆษณา ย่อว่า โค้ด-
    b. ทิฐิ ย่อว่า ทิ-
    c. ภูมิปัญญา ย่อว่า พู-พิ-
    d. สมภูมิ ย่อว่า สะ-

11. พระสงฆ์ เจ็บป่วย คำที่ข้อเส้นได้ตรงกับคำภาษาพื้นที่คำใด
    a. ประชาชน
    b. อาจารย์
    c. หมอ
    d. อาจารย์

12. การช่วยกันทำงานหลายๆ คน
    a. ลงแขก
    b. ลงนา
    c. ลงแรง
    d. ลงขัน

13. พระภิกษุ 5 .......... กัลลยางก์ ..........

14. ส่วนใดของร่างกายที่อยู่ในพระโยธร
    a. พระพักตร์
    b. พระศอ
    c. พระทนต์
    d. พระเสโท

15. คำว่า “เข้าเฝ้าฯ” เป็นคำย่อของ
    a. เข้าเฝ้าพระบาท
    b. เข้าเฝ้าใต้ฝ่าละอองธุลีพระบาท
    c. เข้าเฝ้าใต้ฝ่าละอองธุลีพระ
    d. เข้า

16. “พระบาทสมเด็จพระเจ้าอยู่หัว
    เสด็จแปร................... ณ พระ
คำถามที่เกี่ยวกับข้อความ:

17. ข้อใดเป็นคำชั้น
   a. วิ่งราว  
   b. คูแขง  
   c. ปิดบัง  
   d. ของแข็ง

18. คำสั่งของพระเจ้าแผ่นดิน ตรงกับข้อใด
   a. พระราชโอวาท  
   b. พระบรมราโชวาท  
   c. พระราชโองการ  
   d. พระบรมราชโองการ

19. คำในข้อใดมีความหมายเหมือนกันหมด
   เหมือนกันหมด
   a. ผกา, บุปผา, พนาลี  
   b. กระบินทร์, พานร, กระบี่  
   c. รากษส, ราพณาสูร, กระบินทร์  
   d. เยาวมาลย์, มาลี, มารศรี

20. ............ มีหน้าที่ประหารชีวิต
   นักโทษ
   a. ต RID  
   b. มีอิน  
   c. เพชรฆาต  
   d. เพชฌฆาต

21. พยัญชนะต้นในข้อใดมีเสียง
   พยัญชนะตัวยังก้านมากที่สุด
   a. ใกล้ เพื่อนมีแรงปุ่ม  
   b. กระดาษจากฝรั่งเศส  
   c. ช้างชื่อด้วยจดหมายกับพี่  
   d. ความซ่อนโหน่งตรง

22. ข้อใดมีคำเป็นมากที่สุด
   a. วัดที่มีธรรมนูญเจ้าบุญเก็บไก่  
   b. วัดที่สุดกรุcohได้จะไม่คิด  
   c. วัดที่เป็นคนดีนี่สิช้างคน  
   d. วัดที่คนขับสามไม่เท่าไหร้

23. ข้อใดมีคำตานมากที่สุด
   a. ศิลปะที่น่ายิ่งดุเดือดจริงๆ  
   b. ศิลปะที่น่าตีศิลปะไป  
   c. ศิลปะที่เป็นคนดีนี่สิช้างคน  
   d. ศิลปะที่คนขับสามไม่เท่าไหร่

24. ข้อใดมีเสียงวรรณยุกต์ โท ตรี โท
   ตรีตามลำดับ
   a. เรื่อง, ย่าง, ดาบ, หน้า
25. คำในข้อใด ใช้พยัญชนะต้นเป็นอักษรดุจดังเป็น

อักษรดุจดังเป็น

(โปรดระบุ)

26. ข้อใดมีเสียง “ทร” ต่างไปจากข้ออื่น

(โปรดระบุ)

27. คำในข้อใด เป็นคำที่มีตัวสะกดตรงแม

(โปรดระบุ)

28. คำในข้อใด จัดว่าเป็นคำพ้องเสียง

(โปรดระบุ)

29. ข้อใด มีเสียงวรรณยุกต์ครบ 5 เสียง

(โปรดระบุ)

30. ข้อใดมีวรรณยุกต์แตกต่างจากข้ออื่น

(โปรดระบุ)

31. ข้อใดมีความหมายต่างไปจากพวก

(โปรดระบุ)

32. ข้อใด หมายถึง “ความสับสนรุนเร้า”

(โปรดระบุ)

33. คำในข้อใด หมายถึง “หญิงสาวสวย”

(โปรดระบุ)

34. “ขวานผ่านซาก” หมายถึงข้อใด

(โปรดระบุ)
35. ข้อใดบอกความหมายข้อคำาไมถูกต้อง
d. พุฒแบบมีชั้นเชิง
a. กฎหมาย = ข้อบังคับ
b. สรณะ = เลขาธิการ
c. โจษ = เล่าสือ
d. ครหา = ดูถูก
d.

36. คำในข้อใด มีความหมายเหมือนกันทุกคน
e. หัวหน้า, มัจฉา, อมรินทร์
b. ภูมิภัณฑ์, ราชสี, อรรภิรมณ์
c. อัปสร, สุรางค์, อรรดิ์
d. อาสัญ, บรรลัย, ตักษัย
d.

37. ข้อใดมีความหมายไม่เข้าพวก
a. มฤคี
b. สมมาส
b. รามพันธ์
c. หมื่น

d. ระวัง
d.

38. คำในข้อใด ที่สับสนที่กันแล้ว
a. ค้าลูก
d. หมื่น
b. ค้าเริ่ม
b. โลหิตรื่น
c. ลูก

d. ลูก

39. บทร้อยกรองข้างต้นกล่าวถึงประเทศไทยในด้านใด
a. การอุตสาหกรรม
d. การเกษตรกรรม
c. การต่างประเทศ
d. การเศรษฐกิจ
d.

40. ข้อใดกล่าวได้ถูกต้องและเหมาะสมที่สุด
a. สินค้าไทยมีคุณภาพด้อยกว่าต่างประเทศ
d. ช่วยกันผลิตช่วยกันซื้อจะไม่ทำให้เงินรับไปต่างประเทศ
c. สินค้าไทยควรขายในประเทศไทยเท่านั้น
d. สินค้าไทยล่าสมัยกว่าต่างประเทศ
d.

ตอบคำถามข้อ 39 – 40 
Appendix B1D: Spelling Test (Grade 5 and 6)

ชื่อ...................................... นามสกุล...................................ชั้น.................เลขที่................

1. ทัน-ทะ-คาด เขียนว่า ______________________________
2. กะ-แสด-กะก้า เขียนว่า ______________________________
3. อา-กาด เขียนว่า ______________________________
4. เลค-สะ-กิต เขียนว่า ______________________________
5. กรอบ-รุบ เขียนว่า ______________________________
6. คำ-صاب เขียนว่า ______________________________
7. สาย-สิน เขียนว่า ______________________________
8. สะ-เหน เขียนว่า ______________________________
9. สิ้น-ชะ-ทาน เขียนว่า ______________________________
10. ชะ-นิล-สาค เเขียนว่า ______________________________
11. พต-ชะ-นา-บุ-กรม เเขียนว่า ______________________________
12. ประ-ชา-ราค เเขียนว่า ______________________________
13. ชะ-เริน-พัด-ชะ-นา เขียนว่า ______________________________
14. รำ-วัณ เขียนว่า ______________________________
15. ประ-กาค เเขียนว่า ______________________________
16. พะ-มะ-หา-กะ-สัด เเขียนว่า ______________________________
17. โป-รำน เเขียนว่า ______________________________
18. กา-นางน เขียนว่า ______________________________
19. อุด-สา-ชะ-ก้า เขียนว่า ______________________________
20. โสก-เล่า เขียนว่า ______________________________
Appendix B2A: Stimuli for shadowing task (Grade 1 and 2)

Word and Now-word lists

<table>
<thead>
<tr>
<th>Word</th>
<th>Non-Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>กางเกง</td>
<td>เกิงกัง</td>
</tr>
<tr>
<td>รูปภาพ</td>
<td>รูปภาพ</td>
</tr>
<tr>
<td>สอาง</td>
<td>ซุ่มอง</td>
</tr>
<tr>
<td>ลายยศ</td>
<td>ฉับปิต</td>
</tr>
<tr>
<td>ประกาศ</td>
<td>ประกิต</td>
</tr>
<tr>
<td>ใบลาน</td>
<td>ใบลาน</td>
</tr>
<tr>
<td>ยา</td>
<td>ยาอิ</td>
</tr>
<tr>
<td>อะไรซับ</td>
<td>อะไรซับ</td>
</tr>
<tr>
<td>แงงวาว</td>
<td>แงงวาว</td>
</tr>
<tr>
<td>ทองกวาว</td>
<td>ทองกวาว</td>
</tr>
</tbody>
</table>

Sentences

ไม่สักเป็นไม่หายากและมีค่าทางเศรษฐกิจสูงมากในปัจจุบัน การพูดและการฟังเป็นกระบวนการที่ต้องการให้สื่อสารที่สำคัญ นักเรียนไม่ควรซื้อในสิ่งที่ได้รับพัฒนาโดยไม่ติดตามคำตัวเลือกที่ยั่งยืน นักเรียนจะเป็นที่รู้จักและจะมีผลเสียในหลายจังหวัดมาก การแสดงความคิดเห็นต้องทำอย่างมีเหตุผลและต้องมีการมีการคิด จุดยืนแนวคิดนิติการและต้องมีการมีการตัดสินใจที่จะตัดสินใจ
พจนานุกรมเป็นหนังสาหรับการเขียนและความหมายของคำ ครูควรจะพิจารณาคำที่มีชื่อเสียงให้ต้องรับฟังเป็นProjectileการสู้ เด็กๆ จึงควรรับประทานมากๆ ภาษาไทยเป็นเครื่องมือสื่อสารหลักที่มีอยู่ในภาษาไทย ประเทศที่มีเรื่องราวของโลกเป็นคำสำคัญที่ใช้ทั่วไปในภาษาไทย
นักเรียนสามารถค้นคว้าเพิ่มเติมในอินเทอร์เน็ตได้ทำการเรียนรู้ที่จะมีการบ่อยๆที่มีการสอนในแผนภูมิของโลก
คอมพิวเตอร์ช่วยให้การทำงานของมนุษย์สะดวกรวดเร็วมากขึ้น

ห้องสมุดนับเป็นแหล่งความรู้ที่สำคัญมากในการศึกษาเล่าเรียน

เทคโนโลยีสมัยใหม่ทำให้การดำเนินชีวิตของคนไทยสะดวกสบายขึ้น
Appendix B2B: Stimuli for shadowing task (Grade 5 and 6)

Word and Non-Word lists

<table>
<thead>
<tr>
<th>Word</th>
<th>Non-Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>บวงสรวง</td>
<td>จวงสรวง</td>
</tr>
<tr>
<td>อาศรม</td>
<td>อธิริส</td>
</tr>
<tr>
<td>ชลมารค</td>
<td>ชาลมอก</td>
</tr>
<tr>
<td>ทัณฑฆาต</td>
<td>คันทมาจ</td>
</tr>
<tr>
<td>ฤดูกาล</td>
<td>ฤดีกอล</td>
</tr>
<tr>
<td>ปลาวาฬ</td>
<td>ปลอวุฬ</td>
</tr>
<tr>
<td>ฐานราก</td>
<td>ธอนรอก</td>
</tr>
</tbody>
</table>

Sentences

การโฆษณาเป็นการประชาสัมพันธ์เพื่อเผยแพร่สินค้าหรือบริการในระดับภูมิภูมิคุณภาพซึ่งมีผลต่อการด้านเศรษฐกิจของมนุษย์ทั้งสิ้น

จังหวัดกาญจนบุรีเป็นจังหวัดทางทิศตะวันตกของประเทศไทย

โรคเอดส์เป็นโรคที่ทำให้ระบบภูมิคุณภูมิถูกลดลงและยังไม่สามารถรักษาให้หายได้

ชาวนาซึ่งมีกิจกรรมค้าขายเป็นหนึ่งในราษฎร์ที่มีอยู่ในประเทศไทย

ธรรมชาติและสิ่งแวดล้อมรอบตัวมีผลต่อการดำเนินชีวิตของมนุษย์ทั้งสิ้น

เจ้าภาพเป็นพระราชนิพนธ์ในพระบาทสมเด็จพระจุลจอมเกล้าเจ้าอยู่หัว

มังกรเป็นสัตว์ในนิทานของจีนมีรูปร่างคล้ายงูแต่ตัวใหญ่กว่าและมีหางที่ต่างจากเสือ

ข้าวหอมมะลิของไทยเป็นข้าวที่มีคุณภาพดีเป็นที่ต้องการของตลาดโลก

คติพจน์เป็นถ้อยคำที่เป็นแบบอย่างยังคงให้ความคิดและเป็นข้อความสั้น ๆ

ชาวบ้านบางระจันมีความกล้าหาญจึงได้รับการสร้างเสียงดังทุกวันนี้

ชาวบ้านบางระจันได้รับเลือกให้เป็นหมู่บ้านประชาธิปไตย

พระเจ้าอยู่หัวทรงพระราชทานพระบรมราโชวาทให้คนไทยรู้จักสามัคคีกัน

นักเรียนต้องเรียนรู้การใช้สุภาษิตคำพังเพียรให้ถูกต้องตามความหมาย
รัฐบาลรณรงค์ให้ประชาชนร่วมแจ้งเบาะแสการค้าขายยาเสพติด
โดยลักษณะทางกายภาพของมันยส์มีขนาดใกลเคียงผลส้ม
ระบบสู่ขยะของยาเสพติดในเมืองแห่งดวงดาวที่ชื่อ "กาแล็กซีทางช้างเผือก"
หุ่นยนต์อาจจะเข้ามาแทนที่เครื่องคอมพิวเตอร์เพื่อการทำงานให้มนุษย์ในอนาคต
กระทรวงศึกษาธิการยังพยายามแก้ปัญหาเรื่องการคัดค้านการเปลี่ยนสถานศึกษาอู่
งานเป็นการเรียนเป็นงานศิลปะที่ต้องใช้ความละเอียดในการทำสูงสุดนั่น.
Appendix C: Additional Reading Time Experiments

In Chapter 3 a reading time experiment was described in which the relative frequency of word-start and word-end characters were manipulated. That study formed the basis of further studies (in Chapters 4 and 5) in which the relative frequency of word-start and word-end characters were investigated in eye movement studies. Two additional reading time studies were also conducted in which factors other than relative frequency of characters were investigated. As these could not be followed up in the thesis by eye movement studies, the results of these are reported here. It is hoped that these too may form the basis of eye movement studies in future investigations. These two studies investigated two other distinctive features of Thai orthography on reading time and reading accuracy of Thai readers (both children and adults) as follows:

1) Initial and final consonant characters with high or low feedback consistency relationships (Feedback Consistency Effects)
2) Consonant classes in live syllables combined with different tone characters (Consonant Class and Tone Realisation Effects)

As for the study reported in Chapter 3, these experiments were also normative studies to determine the importance of these orthographic features on the dependent variables of reading time and accuracy of native Thai readers.

The main objective of these experiments is to investigate the effects of the complexities of the Thai writing system on reading ability of Thai children across age, with adult data as a reference point. In Additional Experiment A (Feedback Consistency Effects Experiment), reading accuracy of participants when reading
words in spaced and unspaced sentence conditions was investigated. This experiment was designed to determine the effects of phoneme-to-grapheme correspondences of Thai consonant characters on reading. Additional Experiment B (Consonant Class and Tone Realisation Effects Experiment) investigated the effects of consonant classes in live syllables (open syllables\textsuperscript{13} or syllables ending with sonorants\textsuperscript{14}) and their combination with different tone characters on reading (refer details in Section 1.3). Participants in these experiments are the same as in Experiment 1 (refer details in 3.1).

For these reading time experiments (Additional Experiments A and B) the dependent variables were reading ability of participants as measured by their reading speed, error rate, and error types.

1. Additional Experiment A: Feedback Consistency Effect

In additional experiment A, feedback consistency effects of Thai consonant characters on reading ability were investigated. Feedback consistency refers to the relationship between the phonemes (sounds) and the graphemes (letters) of a language. A feedback consistent relationship is one in which one phoneme represents only one grapheme, while in a feedback inconsistent relationship one phoneme represents two or more graphemes. The independent variables were spaced versus unspaced sentences containing target words with initial and final consonant characters that had either feedback consistent or feedback inconsistent phoneme-to-

\textsuperscript{13} The syllable consists of initial consonant(s) and vowel without final consonant.

\textsuperscript{14} Syllable ending with the following phoneme /m, n, ŋ, l, r, w, j/
grapheme correspondences. Thus there are four conditions of initial and final consonant characters in this experiment:

Feedback Consistent $C_i^{15}$ Feedback Consistent $C_f^{16}$
($FbConC_i \_\_ FbConC_f$);
Feedback Consistent $C_i$ Feedback Inconsistent $C_f$ ($FbConC_i \_\_ FbInconC_f$);
Feedback Inconsistent $C_i$ Feedback Consistent $C_f$ ($FbInconC_i \_\_ FbConC_f$);
Feedback Inconsistent $C_i$ Feedback Inconsistent $C_f$ ($FbInconC_i \_\_ FbInconC_f$).

Participants read sentences aloud as in Experiment 1 and the same dependent variables as in Experiment 1 were derived, reading time, error rate and error type.

1.1 Materials

There were 40 test sentences comprising: 2 initial consonant types (feedback consistent/inconsistent) x 2 final consonant types (feedback consistent/inconsistent) x 2 writing styles (spaced/unspaced) x 5 exemplars (presented in 5 different blocks). In addition, there were 16 practice sentences (eight spaced & eight unspaced). In the practice session, spaced sentences were always presented before the unspaced ones. In the test session, stimuli were divided into five blocks of eight sentences (consistent/inconsistent x initial/final x spaced/unspaced). Target words were presented in the centre of meaningful and grammatical sentence frames. The

\[15\] $C_i$ represents initial consonant

\[16\] $C_f$ represents final consonant
sentence frames fitted within one line. Examples of stimuli are shown below in Table C1 and C0.

**Table C1:** Examples of words with feedback consistent and/or feedback inconsistent initial and final consonant characters (bolded words) for Additional Experiment A (Unspaced condition)

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Initial Character</th>
<th>Final Character</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con/Con</td>
<td>FbCon</td>
<td>FbIncon</td>
<td>ลุงมีและยายมาอาศัยอยู่ในกระท่อมข้างวัด</td>
</tr>
<tr>
<td>Con/Incon</td>
<td>FbCon</td>
<td>FbIncon</td>
<td>สมพรเกิดวันอังคารจึงขอบสีชมพู</td>
</tr>
<tr>
<td>Incon/Con</td>
<td>FbIncon</td>
<td>FbCon</td>
<td>ฝนตกหนักน้ำท่วมทำให้รถติด</td>
</tr>
<tr>
<td>Incon/Incon</td>
<td>FbIncon</td>
<td>FbCon</td>
<td>สุดาปลูกดอกไมริมทางเดินให้ดูสวยงาม</td>
</tr>
</tbody>
</table>

**Table C2:** Examples of words with feedback consistent and/or feedback inconsistent initial and final consonant characters (bolded words) for Additional Experiment A (Spaced condition)

<table>
<thead>
<tr>
<th>Target Word Condition</th>
<th>Initial Character</th>
<th>Final Character</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con/Con</td>
<td>FbCon</td>
<td>FbIncon</td>
<td>ลุงมีและยายมาอาศัยอยู่ในกระท่อมข้างวัด</td>
</tr>
<tr>
<td>Con/Incon</td>
<td>FbCon</td>
<td>FbIncon</td>
<td>สมพรเกิดวันอังคารจึงขอบสีชมพู</td>
</tr>
<tr>
<td>Incon/Con</td>
<td>FbIncon</td>
<td>FbCon</td>
<td>ฝนตกหนักน้ำท่วมทำให้รถติด</td>
</tr>
<tr>
<td>Incon/Incon</td>
<td>FbIncon</td>
<td>FbCon</td>
<td>สุดาปลูกดอกไมริมทางเดินให้ดูสวยงาม</td>
</tr>
</tbody>
</table>

**1.2 Apparatus**

The stimulus sentences were presented stationary and one at a time on the centre of a 17-in. NEC monitor with a 1,024 x 768 pixel resolution and 60 Hz refresh rate approximately 60 centimetres from their eyes in Cordia New font size 22.
Participants’ reading aloud was recorded via the Adobe Audition 1.5 program and from this, reading time, error frequency and error type were derived. Using SpeechView version 2.0.58, acoustic *.wav file was segmented in order to derive the actual reading time of the target word from the starting point (end point of preceding word including silence), to the end point of the target word as shown in Figure C1.

![Reading time extraction using SpeechView version 2.0.58](image)

**Figure C1:** Reading time extraction using SpeechView version 2.0.58

[CSLU Toolkit – Center for Spoken Language Understanding, OGI School of Science & Engineering, Oregon Health & Science University]

### 1.3 Procedure

Each participant was tested individually in a room with minimal noise interference (at school for the children) or a sound-attenuated room (at CRSLP, Chulalongkorn University, Bangkok for adults). There was a set of 16 practice sentences before the test session of 40 test sentences. Participants read all the stimuli aloud and there was no response time limit. The test stimuli were presented in
random order within each block. After reading a stimulus sentence, participants responded by hitting a key (SPACEBAR) in order for the next sentence to be presented on the screen.

1.4 Hypotheses

It was expected that:

1) First and second graders should make more errors in mapping the feedback inconsistent phoneme-to-grapheme correspondences than fifth and sixth graders (as shown in Figure C2), resulting in increased reading time and increased errors.

![Figure C2: Schematic representation of hypothesis 1](image)

2) Since Thai has more initial consonant phonemes than final ones and many consonants change their phonetic representation when they occur in the final position, with respect to error types, younger children may pronounce the final consonant grapheme by substituting its initial grapheme-phoneme correspondence, thus leading to more errors on this measure.
The same hypotheses were entertained regarding spaced versus unspaced sentences and general level of reading as in Experiment 1 (refer Section 3.2.4, Hypotheses 3, 4, and 5, and Figure 19), that is reading performance should improve over age and should be generally better for reading unspaced sentences over age.

### 1.5 Results

The results of this experiment are presented into two sections; reading time and error results:

**Reading Time**

The overall results of this experiment and are schematically represented in Figure C2 and Figure C3. The results of an ANOVA of the factorial design of 2 writing styles (spaced/unspaced) x [2 initial consonant types (feedback consistent/inconsistent) x 2 final consonant types (feedback consistent/inconsistent) x 5 exemplars (one exemplar in each block)] showed that participants’ reading time significantly decreased over age – reading time was significantly faster by adults than children \[F(1,90) = 292.266, p < .001\]. For children, Grade 1 and 2 participants’ reading time was slower than in Grades 5 and 6 \[F(1,90) = 760.425, p < .001\]. Reading level (good versus poor readers) also had the expected significant effect \[F(1,90) = 1174.766, p < .001\]. The interactions between reading level (good versus poor) and age group (adults versus children) \[F(1,90) = 24.615, p < .001\] as well as within the child participants (Grade 1 and 2 versus Grade 5 and 6) \[F(1,90) = 21.831, p < .001\] were also significant. However, the differences between reading time of Grade 1 versus Grade 2 as well as Grade 5 versus Grade 6 readers were not significant \(p > .05\). Therefore there was, as expected, a general improvement in reading speed across all ages.
In addition, reading time in spaced text was generally faster than in unspaced text [F(1,90) = 6.335, p < .01]. This effect was larger in the 1st and 2nd grade compared with the 5th and 6th grade children [F(1,90) = 19.133, p < .001] and the effect was even further reduced in the adults [F(1,90) = 4.430, p < .05]. Thus there was an advantage for spaced text, but this was most pronounced in 1st and 2nd graders, and less pronounced when the participants’ reading skills improved. These overall results for age, reading proficiency and spacing are very similar to the results found in Experiment 1 (refer Figure C2 and Figure C3).

**Figure C2:** Reading time of good readers across age groups
Focusing on the effects of feedback consistency of the consonant in the initial and final position in this experiment, the main effects of both initial \([F(1,90) = 261.576, p < .001]\) and final \([F(1,90) = 212.948, p < .001]\) were significant. Figure C4 shows the effects of feedback consistency and inconsistency of initial consonant (FbCon versus FbIncon in graphs) on reading time of good versus poor readers (GR versus PR in graphs) in each age group when reading spaced versus unspaced texts and Figure C5 similarly shows the effects of the final consonant. The interaction
between initial and final consonants was also significant \(F(1,90) = 20.950, p < .001\) and this effect also interacted with age groups and reading levels; that is this interaction between initial and final consonant effect caused poor readers to read more slowly than good readers across all age groups, \((F(1,90) = 73.056, p < .001\) for child versus adult; \(F(1,90) = 25.808, p < .001\) for Grade 1 and 2 versus Grade 5 and 6; and \(F(1,90) = 16.080, p < .001\) for good versus poor readers].

**Figure C4:** Initial consonant feedback consistency effect

**Figure C5:** Final consonant feedback consistency effect
**Error Type**

Errors were judged by a phonetically-trained native Thai speaker. Errors in this experiment were categorized into four types, as set out below (refer details in Table C3).

1. Phonological Errors:

There were two types of phonological errors:

- **Tone errors:** participants pronounced the incorrect tone

- **Consonant / Vowel errors:** participants pronounced incorrect consonant or vowel that appeared not to be due to confusability via orthographic similarity (refer Orthographical Errors)

2. Orthographical Errors:

- **Orthographic errors:** participants pronounced the incorrect consonant sound substituting a consonant of similar orthography, such as in the example pairs shown in Experiment 1.

3. Semantic Errors:

There was only one type of semantic error found in this experiment.

- **Lexical substitution errors:** participants uttered (substituted) a word consistent in meaning with the context of the preceding or following words or the sentence rather than the target word.

The mean numbers of errors are given in Table C3 and shown in Figure C6. As expected, participants made less error as their age increased and reading skill improved. Strikingly, older child participants made more semantic errors than younger participants. This may be because as their reading ability increases, the
readers increasingly relied on top-down processing when reading, while younger readers rely mainly on bottom up processes. If this is the case then the context in which the target words were embedded would affect the older participants’ reading processes resulting in more semantic errors since the substituted words would normally had an appropriate meaning for the context. It was also notable that the phonological errors also decreased as the reading skills improved.

For the issue of substituting the phonetic representation of final consonant character with its initial phonetic correspondence, the results showed that only consonant sounds that are similar to those found in the final position in the English language, such as /s/, /ʒ/, or /ʃ/ were substituted. That is instead of pronouncing these characters correctly with Thai /t/, child participants as well as some adults tended to use /s/, /ʒ/, and /ʃ/ sounds instead. Such errors were more likely to be made by child participant good readers group than poor readers. This is probably the result of teaching literacy in both Thai and English at the same time in the Thai school curriculum and this causes the English phonological system to interfere with the grapheme-to-phoneme mapping rules of Thai phonology.

Table C3: Error types and mean numbers of errors of each participant group

<table>
<thead>
<tr>
<th>Error type</th>
<th>Good Readers</th>
<th>Poor Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade 1</td>
<td>Grade 2</td>
</tr>
<tr>
<td>Tone</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>Consonant &amp; Vowel</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Orthography</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Lexical Substitution</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>25</td>
</tr>
</tbody>
</table>


1.6 Discussion

The general reading time results of this experiment were similar to those found in Experiment 1; reading time decreased as a function of age and reading skill. Good readers performed better than poor readers in every age group as indicated by shorter reading time and fewer errors. Regarding spacing, as in Experiment 1, spaces between words facilitated especially reading in younger participants, and especially in poor reader groups. Turning to the focus of this experiment, feedback consistency, the results showed that participants’ reading times were slower when the target word started or ended with a feedback inconsistent consonant suggesting that in such inconsistent orthographic and phonological relationships it takes longer to map the letter to its corresponding sound.
2. Additional Experiment B: Consonant Class and Tone Realisation Effects

Additional Experiment B investigated the effects of consonant class (High class, Mid class and Low class) and tone realisation in live syllables (refer detail in Section 1.3) on reading accuracy in Thai. The independent variables were words containing high, mid or low class initial consonant characters in three different syllable structures (CV:#, CVC_{sonorant}, CV:C_{sonorant}) marked with three tone characters (unmarked [ø], mai ek [่ː] and mai tho[่ː]). The participants read every sentence aloud as in Additional Experiment A.

2.1 Materials

The stimuli in this experiment were 135 test words and 12 practice words. The 135 test words consisted of the factorial combination of 3 classes of initial consonant characters (High, Mid and Low) x 3 syllable structures (CV:#, CVC_{sonorant}, CV:C_{sonorant}) x 3 tone characters (unmarked [ø], mai ek [่ː] and mai tho[่ː]) x 5 exemplars (separated into five blocks). Examples of the stimulus item types are given in Table C4.
Table C4: Examples of words with high, mid and low initial consonant class for Additional Experiment B

<table>
<thead>
<tr>
<th></th>
<th>Unmarked [ø]</th>
<th>mai ek [̂ॅ]</th>
<th>mai tho[ॅ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV:♯ (HC&lt;sub&gt;17&lt;/sub&gt;C&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>ขา[kha:4]</td>
<td>ขา[kha:1]</td>
<td>ขา[kha:2]</td>
</tr>
<tr>
<td>CV.C&lt;sub&gt;[son]&lt;/sub&gt; (HCC&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>ชิ้น[khang4]</td>
<td>ถึ้ง[fang1]</td>
<td>ขิ้น[khan2]</td>
</tr>
<tr>
<td>CV:♯ (MCC&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>ป[pa:0]</td>
<td>ป[pa:1]</td>
<td>ป[pa:2]</td>
</tr>
<tr>
<td>CV.C&lt;sub&gt;[son]&lt;/sub&gt; (MCC&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>ปัน[pan0]</td>
<td>ปัน[pan1]</td>
<td>ปัน[pan2]</td>
</tr>
<tr>
<td>CV:C&lt;sub&gt;[son]&lt;/sub&gt; (MCC&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>ปาน[pa:n0]</td>
<td>ปาน[pa:n1]</td>
<td>ปาน[pa:n2]</td>
</tr>
<tr>
<td>CV:♯ (LCC&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>ค[ka:0]</td>
<td>ค[ka:2]</td>
<td>ค[ka:3]</td>
</tr>
<tr>
<td>CV.C&lt;sub&gt;[son]&lt;/sub&gt; (LCC&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>คัน[khan0]</td>
<td>คัน[khan2]</td>
<td>คัน[khan3]</td>
</tr>
<tr>
<td>CV:C&lt;sub&gt;[son]&lt;/sub&gt; (LCC&lt;sub&gt;i&lt;/sub&gt;)</td>
<td>คาน[ka:n0]</td>
<td>คาง[kha:ng2]</td>
<td>คาง[kha:n3]</td>
</tr>
</tbody>
</table>

As set out in Section 1.3, tone realisation of each tone character depends on the class of the initial consonant it co-occurs with. For example, the tone marker (mai ek [̂ॅ]) which is nominally the tone character for Low tone realization, changes its tone realization to Falling tone when it co-occurs with a Low class consonant (refer Table C5 below).

---

17 HC represents High Class; MC represents Mid Class and LC represents Low Class Consonant
2.2 Apparatus

This experiment used the same apparatus as in Experiment 1 and Additional Experiment A.

2.3 Procedure

Participants were tested in the same ways as in Additional Experiment A. There was a set of 12 practice words before the testing session. Participants read all the stimulus words aloud and there was a response time limit of 4000 millisecond before the next stimulus was presented on the screen. The response time limit was necessary in this experiment because the stimulus was a short monosyllabic word (as opposed to sentences in Experiment 1 and Additional Experiment A) which should not take long to read and the time limit would help forcing the participants to make responses faster. The stimuli were presented in random order within each of the five blocks.
2.4 Hypotheses

On the basis of the nature of literacy teaching in Thailand in which Mid and Low consonant classes are introduced to the children before High class consonant characters (refer Section 1.3) and static tones are taught before contour tones, it is possible that children would make more errors due to variations of consonant class and tones, whereas adults, who should have acquired all the phonological and tonological rules should make more errors according to incongruity of tone realizations in Thai. On this basis then, it is possible to hypothesise that:

1) Children should **make more errors** reading words starting with high class initial consonant characters than mid and low classes since they were less familiar with high consonant class, whereas adults should make more errors reading words starting with low class consonants which have more transparent tone realisations than the other two classes.

2) Adults should make more errors reading words with opaque tone realisation, i.e., in which the nominal tone for a particular tone character is modified for a different tone realisation due to its interaction with the initial consonant.

3) Regarding reading time, participants should read transparent tone realisation words faster and good readers should take less time to read than poor readers across all ages.
2.5 Results

The results of this experiment are presented in two sections similar to the presentation of Additional Experiment A; the reading time results and error results.

Reading Time

The results of an Analysis of Variance (ANOVA) of the factorial design of 3 classes of initial consonant characters (High, Mid and Low) x 3 syllable structures (CV:#, CVC[sonorant], CV:C[sonorant]) x 3 tone characters (unmarked [ø]/Tone 1, mai ek [ŋ]/Tone 2 and mai tho[ŋ]/Tone 3) x 5 exemplars (separated into five blocks) showed that participants’ reading time significantly decreased over age – reaction time for reading the single words in this experiment was significantly faster for adults than children [F(4,90) = 4727.911, p < .001]. For children, Grade 1 and 2 participants’ reading time was slower than in Grades 5 and 6 [F(2,72) = 3526.944, p < .001]. Thus there was, as expected, a general improvement in reading speed across all ages. Turning to reading level of participants, as expected, the reading time of good readers was always faster than that of poor readers in every age group [F(1,90) = 333.955, p < .001]. Moreover, the interaction between age groups and reading level also had a significant effect [F(4,90) = 38.366, p < .001] (refer Figure C7A to C7C).
Figure C7A: Reading time of Grade 1 and Grade 2 good and poor readers

Figure C7B: Reading time of Grade 5 and Grade 6 good and poor readers
Turning to the main factor in this experiment, the results of the consonant frequency and tone realisation effects on reading time are shown in Figures C8A to C8C. The main effect of initial consonant class was significant \[F(2,180) = 43.236, p < .001\], as was the effect of tone realisation \[F(2,180) = 11.896, p < .001\] and syllable structure types \[F(2,180) = 8.162, p < .001\]. There were also a significant interactions between initial consonant classes and age groups \[F(8,180) = 3.782, p < .001\] of tone realisation and age group and initial consonant class by tone realisation \[F(8,180) = 3.254, p = .002\].

Considering further the consonant class and tone realisation interactions over age it can be seen in Figures C8A to C8C, that for the child participants, reading time was almost always highest when reading words containing High class initials with Tone 1 [unmarked]. Reading time for the other initial consonant classes and tone characters were quite varied, although for Grade 2, 5 and 6 participants tended to took longer to read words containing Low class initial with Tone 2 (mai ek [_.ć]) than with Tone 3 (mai tho[_.ć]). These effects may be the results of the opaque tone realisation of these tone characters when they co-occur with High and Low initial consonant classes as shown in Table C60.
Figure C8A: Reading time of Grade 1 and 2 participants

Figure C8B: Reading time of Grade 5 and 6 participants

Figure C8C: Reading time of adult participants
Table C6: Transparency of tone realisations on different consonant classes

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Mid</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1 (ø)</td>
<td>Opaque</td>
<td>Transparent</td>
<td>Transparent</td>
</tr>
<tr>
<td>Tone 2 (_ظر)</td>
<td>Transparent</td>
<td>Transparent</td>
<td>Opaque</td>
</tr>
<tr>
<td>Tone 3 (_ظر)</td>
<td>Transparent</td>
<td>Transparent</td>
<td>Opaque</td>
</tr>
</tbody>
</table>

**Error Type**

Errors in this experiment were judged by a phonetically-trained native Thai speaker as in Experiment 1. Errors were categorized into three types, as set out below.

- **Tone errors**: participants pronounced the incorrect tone sounds
- **Consonant errors**: participants pronounced incorrect consonant sounds, including errors caused by orthographic similarity
- **Vowel errors**: participants pronounced incorrect vowels

The results are shown in Table C7. It can be seen that the number of errors decreased over age and that good readers made less errors than poor readers in every grade. Since there were two levels of difficulty of target words (for Grade 1&2 and Grade 5&6), it was expected that Grade 1 and Grade 5 participants should have similar mean numbers of errors while Grade 2 should similar to Grade 6. However, the results show that, even though the target words for Grade 5 participants were more difficult, they made less error than Grade 2 participants, similar to the results found in Experiment 1 and Additional Experiment A.
Table C7: Error types and mean numbers of errors of each participant groups

<table>
<thead>
<tr>
<th>Error type</th>
<th>Good Readers</th>
<th>Poor Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade 1</td>
<td>Grade 2</td>
</tr>
<tr>
<td>Tone</td>
<td>177</td>
<td>144</td>
</tr>
<tr>
<td>Consonant</td>
<td>54</td>
<td>42</td>
</tr>
<tr>
<td>Vowel</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>259</td>
<td>203</td>
</tr>
</tbody>
</table>

From the graphical representation in 0 it can be seen that there were generally more errors in the opaque tone realization conditions (Tone 1 High Class; Tone 2, Low Class and Tone 3 Low Class – refer Table C6) than the transparent ones for both good and poor readers.

2.6 Discussion

Similar to the findings of Experiment 1 and Additional Experiment A, the results in this Additional Experiment B here showed that reading time decreased over age. Additionally, good reader groups always performed better than poor reader
groups in every grade. Turning to the consonant class effects, as expected, younger participants made more errors reading words with High class consonants while for adult participants the errors were more likely to occur when reading the words starting with High and Low consonant classes. Considering initial consonant class and tone realisation effects, the results showed that for transparent tone realisations, as hypothesised, participants across all groups made fewer errors pronouncing the words comprised of these combinations. On the other hand, the opaque tone realisations, always caused problems for the participants irrespective of the the readers’ age or reading skill (refer Figures C10A to C10C).

**Figure C10A:** Consonant classes and tone realisation effects (Grade 1 and 2)

**Figure C10B:** Consonant classes and tone realisation effects (Grade 5 and 6)
Figure C10C: Consonant classes and tone realisation Effects (Adults)
Appendix D1: Target words used in Experiment 1

Grade 1 and 2 stimuli

<table>
<thead>
<tr>
<th>Item no.</th>
<th>Initial</th>
<th>Final</th>
<th>Spacing</th>
<th>Word</th>
<th>Length</th>
<th>Syllable</th>
<th>Reduced syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H</td>
<td>H</td>
<td>S</td>
<td>แกน</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>H</td>
<td>H</td>
<td>S</td>
<td>การบาน</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>H</td>
<td>H</td>
<td>S</td>
<td>พระ</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td>H</td>
<td>U</td>
<td>ทำงาน</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>H</td>
<td>H</td>
<td>U</td>
<td>ตัวถ้า</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>H</td>
<td>H</td>
<td>U</td>
<td>น้ำตก</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>M</td>
<td>S</td>
<td>ราช</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>M</td>
<td>S</td>
<td>คิดเลข</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>H</td>
<td>M</td>
<td>S</td>
<td>เกศ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>H</td>
<td>M</td>
<td>U</td>
<td>ตั้งเลข</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>H</td>
<td>M</td>
<td>U</td>
<td>ประเทศ</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>H</td>
<td>M</td>
<td>U</td>
<td>พันที่</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>H</td>
<td>L</td>
<td>S</td>
<td>เผย</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>H</td>
<td>L</td>
<td>S</td>
<td>มะโพ</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>H</td>
<td>L</td>
<td>S</td>
<td>เทวิยแฒ</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>H</td>
<td>L</td>
<td>U</td>
<td>ตั้งกญู</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>H</td>
<td>L</td>
<td>U</td>
<td>กอไผ่</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>H</td>
<td>L</td>
<td>U</td>
<td>กองโพ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>H</td>
<td>S</td>
<td>ชื่นชม</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>H</td>
<td>S</td>
<td>ละเล่น</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>M</td>
<td>H</td>
<td>S</td>
<td>ฝีนม้า</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>M</td>
<td>H</td>
<td>U</td>
<td>ถาวร</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>M</td>
<td>H</td>
<td>U</td>
<td>ซือซบ</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>M</td>
<td>H</td>
<td>U</td>
<td>ซุบซิบ</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>M</td>
<td>M</td>
<td>S</td>
<td>โกรธ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>M</td>
<td>M</td>
<td>S</td>
<td>โทษ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>M</td>
<td>M</td>
<td>S</td>
<td>วาดภาพ</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>M</td>
<td>M</td>
<td>U</td>
<td>บางชัด</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>M</td>
<td>M</td>
<td>U</td>
<td>ไอเดีย</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>M</td>
<td>M</td>
<td>U</td>
<td>ช้านานเวลา</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>M</td>
<td>L</td>
<td>S</td>
<td>ช้างใหญ่</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>M</td>
<td>L</td>
<td>S</td>
<td>ซากไม้</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>M</td>
<td>L</td>
<td>S</td>
<td>ยกแม่สิงห์</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>M</td>
<td>L</td>
<td>U</td>
<td>ตามมาจุใจ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>M</td>
<td>L</td>
<td>U</td>
<td>ชัยรูพี</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>36</td>
<td>M</td>
<td>L</td>
<td>U</td>
<td>โจซี</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>L</td>
<td>H</td>
<td>S</td>
<td>จังหวะ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>L</td>
<td>H</td>
<td>S</td>
<td>ฉานา</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>L</td>
<td>H</td>
<td>S</td>
<td>ฉันใจ</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>L</td>
<td>H</td>
<td>U</td>
<td>ฉุยชาย</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>L</td>
<td>H</td>
<td>U</td>
<td>ฟองน้ำ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>L</td>
<td>H</td>
<td>U</td>
<td>ฐานงาน</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>L</td>
<td>M</td>
<td>S</td>
<td>ฮึกหาญ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>L</td>
<td>M</td>
<td>S</td>
<td>ญาติ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>L</td>
<td>M</td>
<td>S</td>
<td>ญาณ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>L</td>
<td>M</td>
<td>U</td>
<td>พระบางศัพท์</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>47</td>
<td>L</td>
<td>M</td>
<td>U</td>
<td>ผีเสื้อ</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>L</td>
<td>M</td>
<td>U</td>
<td>ผู้ยืน</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>ชวาใจ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>ยาเสพติด</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>ฐานไม้ฝา</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>L</td>
<td>L</td>
<td>U</td>
<td>ฐานวัตถุ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>L</td>
<td>L</td>
<td>U</td>
<td>ข้อ</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>L</td>
<td>L</td>
<td>U</td>
<td>พังงู</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Note:**  H = High, M = Medium, and L = Low character frequency

S = Spaced text, and U = Unspaced text
<table>
<thead>
<tr>
<th>Item no.</th>
<th>Initial</th>
<th>Final</th>
<th>Spacing</th>
<th>Word</th>
<th>Length</th>
<th>Syllable</th>
<th>Reduced syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H</td>
<td>H</td>
<td>S</td>
<td>กรรมการ</td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>H</td>
<td>H</td>
<td>S</td>
<td>เทศกาล</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>H</td>
<td>H</td>
<td>S</td>
<td>ทัณฑฆาต</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td>H</td>
<td>U</td>
<td>เนื้อซุน</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>H</td>
<td>H</td>
<td>U</td>
<td>หลั่งเสียด</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>H</td>
<td>H</td>
<td>U</td>
<td>กระโถน</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>M</td>
<td>S</td>
<td>เจ้าawaii</td>
<td>8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>M</td>
<td>S</td>
<td>คริสต์มาส</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>H</td>
<td>M</td>
<td>S</td>
<td>เซิ่งอรรถ</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>H</td>
<td>M</td>
<td>U</td>
<td>ทองอัฐ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>H</td>
<td>M</td>
<td>U</td>
<td>เทคนิค</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>H</td>
<td>M</td>
<td>U</td>
<td>เครื่องเทศ</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>H</td>
<td>L</td>
<td>S</td>
<td>เรือกลไฟ</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>H</td>
<td>L</td>
<td>S</td>
<td>ปล่องไฟ</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>H</td>
<td>L</td>
<td>S</td>
<td>ทฤษฎี</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>H</td>
<td>L</td>
<td>U</td>
<td>จ้ำมกิจ</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>H</td>
<td>L</td>
<td>U</td>
<td>พายกราฟ</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>H</td>
<td>L</td>
<td>U</td>
<td>เซียมซี</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>H</td>
<td>S</td>
<td>ชโลทร</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>H</td>
<td>S</td>
<td>ล้อเล็ก</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>M</td>
<td>H</td>
<td>S</td>
<td>ถีนมา</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>M</td>
<td>H</td>
<td>U</td>
<td>กระจก</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>M</td>
<td>H</td>
<td>U</td>
<td>ชวดรา</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>M</td>
<td>H</td>
<td>U</td>
<td>ซุปซิบ</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>M</td>
<td>M</td>
<td>S</td>
<td>โจติรัส</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>M</td>
<td>M</td>
<td>S</td>
<td>โจษ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>M</td>
<td>M</td>
<td>S</td>
<td>วาดนิพพาน</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>M</td>
<td>M</td>
<td>U</td>
<td>บังชย</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>M</td>
<td>M</td>
<td>U</td>
<td>ไม่ให้</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>M</td>
<td>M</td>
<td>U</td>
<td>ข้ามานั้น</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>M</td>
<td>L</td>
<td>S</td>
<td>ข้ายุทธ</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>M</td>
<td>L</td>
<td>S</td>
<td>ข่องไฟ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>M</td>
<td>L</td>
<td>S</td>
<td>ยกแม่</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>M</td>
<td>L</td>
<td>U</td>
<td>ตุ๋วอ้วน</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>M</td>
<td>L</td>
<td>U</td>
<td>ข้ากับ</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>M</td>
<td>L</td>
<td>U</td>
<td>โจเซ</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>L</td>
<td>H</td>
<td>S</td>
<td>ชัย</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>L</td>
<td>H</td>
<td>S</td>
<td>จันทร์</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>L</td>
<td>H</td>
<td>S</td>
<td>จับป่วย</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>L</td>
<td>H</td>
<td>U</td>
<td>ชัยราย</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>L</td>
<td>H</td>
<td>U</td>
<td>พ่อหน่า</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>L</td>
<td>H</td>
<td>U</td>
<td>จ่าเหลา</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>L</td>
<td>M</td>
<td>S</td>
<td>อิศหาญ</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>L</td>
<td>M</td>
<td>S</td>
<td>ญาติศ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>L</td>
<td>M</td>
<td>S</td>
<td>ญาณ</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>L</td>
<td>M</td>
<td>U</td>
<td>ชราวัส</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>47</td>
<td>L</td>
<td>M</td>
<td>U</td>
<td>ภูติภัส</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>L</td>
<td>M</td>
<td>U</td>
<td>ถูกไฟ</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>ชาวใส่</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>สายเสีย</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>ฐานไม้ฝัก</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>L</td>
<td>L</td>
<td>U</td>
<td>สายอยู่</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>L</td>
<td>L</td>
<td>U</td>
<td>ษัฏ</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>L</td>
<td>L</td>
<td>U</td>
<td>พ่อแห่ง</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D2: Sentences used for adults in Experiment 2 and older child in Experiment 3

1 ศักดิ์ชาย เป็น กรรมการ ที่ ทำ หน้าที่ ได้ ดีมาก
2 ประชาชน ร่วมงาน ออกพรรษา อย่าง ล้นเหล่า
3 ภาษาไทย ใช้ ทัณฑฆาต ก้าบ ถูก ไม่ ออกเสียง
4 ร้านนิโอเบลทำขนมเม็ดขนุนได้รับรางวัลชนะเลิศ
5 สูญญศัตรูศาสตร์เกี่ยวกับสุคติเพื่อใช้สอนนักเรียน
6 ข้างต้นผสมใช้กระป๋องเป็นอาหารในกระบวนการ
7 หลวงตาจีน เป็น เจ้าอาวาส ที่ มี ชื่อเสียง น้า นับถือ
8 ชาววิเศษ ร่วมงาน ศิริมงคล อย่าง สนุกสนาน
9 นักเขียน ใช้ เชิงอรรถ ใน การอธิบาย ที่มา ข้อมูล
10 ศูนย์คณิตศาสตร์แก้ หลักสูตรเพื่อใช้สอนนักเรียน
11 ร้านนิโอเบลทำขนมเม็ดขนุนได้รับรางวัลชนะเลิศ
12 ชาววิเศษ ร่วมงาน ศิริมงคล อย่าง สนุกสนาน
13 นักเขียน ใช้ เชิงอรรถ ใน การอธิบาย ที่มา ข้อมูล
14 ศูนย์คณิตศาสตร์แก้ หลักสูตรเพื่อใช้สอนนักเรียน
ข่าว นี้ เป็น เรื่อง ล่าลือ ของ ผู้คน ได้ ภาคกลาง

นาย ก่าหนา เมื่อมา เพื่อ ใช้ ปลูก ข้าว พันธุ์ใหม่

หมู่บ้านนี้เป็นของทางที่สร้างเพื่อใช้เป็นศูนย์พักพัน

นักเรียนใหม่ชื่อชฎาธารมาจากต่างจังหวัด

นักเรียนใหม่ชื่อชฎาธารมาจากต่างจังหวัด

ญาณีเป็นคนซุบซิบเรื่องไม่ดีของเพื่อนให้คนอื่นฟัง

พนักงานใหม่ชื่อซิมมี่ ถูก ย้าย ออกจาก ขาดาน หลายคน เป็นคนที่สร้างความหุ้นส่วนที่สร้างความหุ้นส่วน

การวิจัยพบว่ามีนักศึกษาเรื่องการผ่าตัดเปลี่ยนถ่ายหัวใจมาก

นักเขียนที่ดีควรเวนช่องว่างใน การเขียน หนังสือ

นักเขียนที่ดีควรเวนช่องว่างใน การเขียน หนังสือ

ข่าวฆาตกรรมเป็นเรื่องโจษของชาวบ้านหลายหมู่บ้าน

ที่ดีควรเวนช่องว่างใน การเขียน หนังสือ

ที่ดีควรเวนช่องว่างใน การเขียน หนังสือ

อาคารแห่งใหม่เป็นของดุษฎีที่จัดเป็นศูนย์กลางการเงิน
42 ตัวราชตามจับฆาตได้ที่บริเวณชายแดนประเทศไทย
43 ตนเป็นคนอีกฝ่าย จึงมักจะมีเรื่องเสมอ
44 ก้าวหน้าดี จัดงานรวบรวมญาติที่อยู่ดังจังหวัด
45 วิเศษชำนาญคนเป็นญาติผู้มีญาณหยั่งรู้อนาคตได้
46 เล่าปูนได้รับการนับถือเป็นแม่รำวงที่ดีที่สุดประจ ajoutจังหวัด
47 โรคฝีดาษเป็นโรคระบาดร้ายแรง
48 ไม่อดีตชาวบ้านควรพินัยฤๅษีว่าเป็นผู้ทรงคือทรงธรรม
49 งามตาเป็นคนฉาวโฉ่จึงไม่มีใครอยากคบหา
50 ช่างไม่ทำฐานไม่ไผ่สำหรับตั้งเครื่องเล่นในงานวัด
51 ไม่พจนานุกรมคำว่าถามเป็นคำปรารถน์ที่สุทธิใช้แล้ว
52 ความหมายของคำว่าษัทธิตามพจนานุกรม
53 ความหมายของคำว่าฟื้นฟูศีลเรียกหรือเดชขึ้น
54 ความหมายของคำว่าพิฉัตรศึกทรัพย์หรือเดชขึ้น
Appendix D3: Sentences for younger child participants in Experiment 3

1 สุดา พา แก่น มา ที่ ศาล
2 ขัน ช่วย ทำงานบ้าน ทุกวัน
3 ใกล้บ้าน พระ มี สุนัข
4 อาเป็นคนทำงานที่มีชื่อเสียง
5 ชายต้องเขียนคำาคาเพื่อสอนเด็ก
6 ที่จังหวัดนั้นน้ำตกหลายที่สวยงาม
7 มาณี พา ราช มา ที่ บ้าน
8 ของ ช่วย คิดเลข ทุกวัน
9 แถว บ้าน เกศ มี รังนก
10 หญิงต้องเขียนตัวเลขเพื่อใช้เป็นโจทย์
11 ผู้คนของประเทศแถบนี้ยิ้มแย้มใส
12 ผูกกางเกงพื้นที่เพื่อใช้ปลูกต้นไม้
13 เอก พา เมฆ มา ที่ ป่า
14 ข้างบ้าน มะไฟ มี ชิงช้า
15 แม่ ช่วย เหวี่ยงแห ทุกวัน
16 ผู้ใหญ่บ้านเป็นคนดังกล่าวที่ยุติธรรม
17 ที่ติดแบบนี้ก็อยู่ไกลจากมาก
18 ข่าวผ่านเขียนกวาดเพื่อใช้ทำงานงาน
19 สัม มักจะ ถูก เพื่อน ซื้อเสริม อยู่ เสมอ
20 ข่าว นี้ เป็น เรื่อง ล่าสุด ของ ผู้คน
21 คุณครู กำหนด มิณมา เพื่อ ใช้ ปลูก ข้าว
22 บ้านหลังนี้เป็นของอาจารย์ที่ซื้อให้คุณแม่
23 นักเรียนใหม่ซื้อโลหะมาจากต่างจังหวัด
24 นิทเป็นคนชุบชิมเรื่องของเพื่อนๆ
25 วิน มักจะ ถูก เพื่อน โจ้ อยู่ เสมอ
26 ข่าว นี้ เป็น เรื่อง ใจซ ของ ชาวบ้าน
27 นิด เป็น คน วาดภาพ ที่ มี ราคา แพง
28 บ้านหลังนี้เป็นของบรรษัทที่เพิ่งเข้ายามา
29 เล็กใหม่ที่ซื้อติดส查明จากกรุงเทพมหานคร
30 จอมเป็นคนขายนายเรื่องสะสมของเก่า
31 นักเรียน คน เทม ช่องไฟ ใน การเขียน หนังสือ
32 กิ๊บ มักจะ ถูก เพื่อน ซักไซ เลย เสมอ
33 ข่าว นี้ เป็น เรื่อง ข่มขู่ ของ นักข่าว
34 ตีกหลังนี้เป็นเรื่องดุริณ์ที่จะแฉ่งให้ค้า
35 คุณครูคนใหม่ซื้อชุดเก้าданจากโรงเรียนชื่อดัง
36 นักเเท้ติดโซเซเพราะชอบดื่มเหล้ามาก
37 ลุงพง เป็น คน เล่น ฆ้อง ที่ เก่ง มาก วัน ถ้า ได้ อย่างดี
ดาว มักจะ เข้า ผาน ไป ใน ช่วง เย็น เสมอ

ทินกร ทำงาน ช้าใจ จึง ได้รับ การยอมรับ

กบชอบร่าฉุยฉายมากเป็นพิเศษ

ฟอนสาวไหมเป็นการฟอนรำประจุภาคเหนือ

ตาวจูจับฆาตกรได้ที่ชายแดน

กลา เป็น คน อิทธิยาจึง มักจะ มี เรื่อง

นากี เป็น ม_naที่อยู่ ใน หมู่บ้าน ชื่น

ปิจิต เป็น ผู้ มี ญาณ หยั่งรู้ อนาคต

ลุงปุ เป็น ฆราวาสที่ดีประจุตาบล

โรคฝีดาษเป็นโรคระบาดอย่างแรง

ในอดีตพระฤๅษีเป็นที่เคารพของชาวบ้าน

ยายเมา เป็น คน ฉาวโฉ่จึง ไม่มีใครคบ

แจง เป็น คน ฮาเฮที่สุด ใน ห้อง

ชางไม่ทำฐานไม่ไผ่ สำหรับ ด้วง เครื่องเล่น

ในฐานะผู้บุกค้ำว่ามาจากเป็นคำโบราณที่เลิกใช้แล้ว

ความหมายของคำว่าษัฑคือหกตามพจนานุกรม

ความหมายของคำว่าฟอนเพื่อเปรียบสร้างหรือดีชื่น
Appendix E1: R Scripts used to analyse Experiment 1

library(lme4)
library(reshape)
library(ggplot2)
library(MASS)
source("myFuncs2.R")

RT.df <- read.csv("ReadingTimeDataNew.csv")
RT.df$Ability <- ordered(substr(RT.df$Group,1,1),
c("G", "P"),
labels=c("good", "poor"))
RT.df$Gender <- ordered(substr(RT.df$Group,2,2),
c("M", "F"),
labels=c("male", "female"))
RT.df$SylCov <- RT.df$Syllable
RT.df$Syllable <- ordered(RT.df$Syllable,c("1", "2", "3"))
RT.df$Word <- makeWordIndex(RT.df$Grade, RT.df$Cond)
RT.df$Grade <- ordered(RT.df$Grade, c("first","second","fifth", "sixth", "Adult"))

# Create an item ID
RT.df$Item <- RT.df$Item + 54
RT.df[RT.df$Grade="first" | RT.df$Grade="second", c("Item") ] <-
RT.df[RT.df$Grade="second", c("Item") ]-54

Freq.df <- read.csv("wordFreq.csv")
RT.df <- merge(RT.df, Freq.df, by=c("Item"))
RT.df$lwfreq <- log(RT.df$wfreq+1)
RT.df$lfcfreq <- log(RT.df$fcfreq+1)
RT.df$llcfreq <- log(RT.df$lcfreq+1)

summary (RT.df)

with (RT.df,
tapply(RT, list(Grade, Syllable), mean, na.rm=TRUE))

with (RT.df,
tapply(RT, list(Grade, Ability, Syllable), mean, na.rm=TRUE))
with (RT.df,
  tapply(RT, list(Grade, Initial), mean, na.rm=TRUE))

with (RT.df,
  tapply(RT, list(Ability, Grade, Initial), mean, na.rm=TRUE))

with (RT.df,
  tapply(RT, list(Grade, Final), mean, na.rm=TRUE))

with (RT.df,
  tapply(RT, list(Ability, Grade, Final), mean, na.rm=TRUE))

with (RT.df,
  tapply(RT, list(Spacing, Grade), mean, na.rm=TRUE))

with (RT.df,
  tapply(RT, list(Spacing, Ability), mean, na.rm=TRUE))

p <- qplot(Grade, RT, data = RT.df,
        stat="summary", fun.y = "mean", geom="bar", position="dodge",
        fill=Syllable) +
        opts(title="Number of Syllable")
print (p)

# define contrast matrices
c.fdiff1 <- matrix(
  c(  4/5, -1/5, -1/5, -1/5, -1/5,
      3/5,  3/5, -2/5, -2/5, -2/5,
      2/5,  2/5,  2/5, -3/5, -3/5,
      1/5,  1/5,  1/5,  1/5, -4/5),
      5, 4,
  dimnames = list( c("1st","2nd","5th","6th","Adult"),
                   c(".1st<>2nd",".2nd<>5th",".5th<>6th",".6th<>Adult")))
c.fdiff1
c.contrasts(RT.df$Grade) <- c.fdiff1

# define a forward difference contrast matrix
c.fdiff2 <- matrix(
  c(  2/3, -1/3, -1/3,
      1/3,  1/3, -2/3),
      3, 2,
  dimnames = list(c("H","M","L"),
                   c(".H<>M",".M<>L")))
c.fdiff2
c.contrasts(RT.df$Initial) <- c.fdiff2
c.contrasts(RT.df$Final) <- c.fdiff2
c.fdiff3 <- matrix(
  c(1/2, -1/2),
  2, 1,
  dimnames = list(c("S","U"), c(".S<>U")))
c.fdiff3
contrasts(RT.df$Spacing) <- c.fdiff3
c.fdiff4 <- matrix(
  c(1/2, -1/2),
  2, 1,
  dimnames = list(c("G","P"), c(".G<>P")))
c.fdiff4
contrasts(RT.df$Ability) <- c.fdiff4

#Plot 1
p <- qplot(Grade, RT, data = RT.df,
  stat="summary", fun.y = "mean", geom="bar", position="dodge",
  fill=Initial) + facet_wrap (~Syllable) +
  opts(title="Reading time")
print (p)

#########################################
#          Latest Models         #
#########################################
print (RT.m1 <- lmer(
  RT ~ SylCov + lwfreq + (fcfreq + lcfreq + Grade + Ability + Spacing) +
  (1|Subj) + (1|Item),
  data=RT.df),
  cor=FALSE)

#try interaction
print (RT.m2 <- lmer(
  RT ~ SylCov + lwfreq + (fcfreq + lcfreq + Grade + Ability + Spacing)^2 +
  (1|Subj) + (1|Item),
  data=RT.df),
  cor=FALSE)

#clean up
print (RT.m3 <- lmer(
  RT ~ SylCov + lwfreq + (fcfreq + lcfreq + Grade + Ability + Spacing) +
     fcfreq:lcfreq + fcfreq:Ability + lcfreq:Spacing +
     Grade:Ability + Grade:Spacing + Ability:Spacing +
  (1|Subj) + (1|Item),
  data=RT.df),
  cor=FALSE)
cor=FALSE)

anova (RT.m3, RT.m1)

#Plot 1 (fcfreq:Ability)

p <- qplot(Ability, RT, data = RT.df,
           stat="summary", fun.y = "mean", geom="bar", position="dodge",
           fill=Final) +
           opts(title="Reading Time")

print (p)
library(lme4)
library(reshape)
library(ggplot2)
source("myFuncs.R")

# Read in data files for letter level
Letters.Silent.Junior.df <- read.csv("KidLetter.csv")
Letters.Silent.Adult.df$age <- "adult"
Letters.Silent.Child.df$age <- "child"
Letters.Silent.Junior.df$age <- "junior"
Letters.Silent.Adult.df$mode <- "silent"
Letters.Silent.Child.df$mode <- "silent"
Letters.Silent.Junior.df$mode <- "silent"

Letters.Aloud.Adult.df$age <- "adult"
Letters.Aloud.Adult.df$mode <- "aloud"

Letters.df <- rbind(Letters.Silent.Adult.df,
                     Letters.Silent.Child.df,
                     Letters.Silent.Junior.df,
                     Letters.Aloud.Adult.df)
Letters.df$mode <- factor(Letters.df$mode)
Letters.df$age <- factor(Letters.df$age, levels=c("junior","child","adult"), ordered=TRUE)
# Remove question data
Letters.df <- Letters.df[makeGlobalFixIndex(Letters.df$item) ==
Letters.df$CURRENT_FIX_INDEX,]

# Clean up names to conform for merger
Letters.df <- rename(Letters.df, c(item = "sent",
CURRENT_FIX_INTEREST_AREA_INDEX = "fixIA",
CURRENT_FIX_INDEX = "nfix", RECORDING_SESSION_LABEL = "subjn"))

summary(Letters.df)

# Merge in condition variable
Letters.df <- merge(Letters.df, read.csv("BenCondition.csv"), by="sent")

# Distinguish spacing conditions
Letters.df$spacing <- ordered(substr(Letters.df$condition,3,3),
c("S", "U"), labels=c("Present", "Absent"))

# Temporarily split the df to allow merger of different LIW data
Letters.spaced.df <- Letters.df[Letters.df$spacing=="Present",]

# Load in text information data frame
Text.df <- rbind(read.csv("Ben_Sentences3b.csv"), read.csv("Ben_Sentences_Kid3.csv"))
Text.df <- subset (Text.df, select=c("sent", "word", "wlength", "wordcode", "target", "wfreq", "pfirst", "pmid", "plast","avgBigram"))
Text.df$logfreq <- log10(1+Text.df$wfreq)
Text.df$lrwfreq[Text.df$logfreq<3] <- "low"
Text.df$lrwfreq[Text.df$logfreq>=3] <- "high"
Text.df$lrwfreq <- factor(Text.df$lrwfreq, levels=c("low", "high"), ordered=TRUE)
Text.df$avgBigram <- factor(trunc(Text.df$avgBigram), order=TRUE)
Text.df$avgBigram[Text.df$avgBigram<3] <- 3
Text.df$avgBigram[Text.df$avgBigram>4] <- 4
Text.df$avgBigram <- factor(Text.df$avgBigram, ordered=TRUE, labels=c("low","high"))

LIWunspaced.df <- makeUnspacedLIWDF(Text.df$sent, Text.df$word, Text.df$wlength)
LIWspaced.df <- makeSpacedLIWDF(Text.df$sent, Text.df$word, Text.df$wlength)
Letters.spaced.df <- merge(Letters.spaced.df, LIWspaced.df, by.x=c("sent","fixIA"), by.y=c("sent","char"))
Letters.unspaced.df <- merge(Letters.unspaced.df, LIWunspaced.df, by.x=c("sent","fixIA"), by.y=c("sent","char"))
Letters.df <- rbind(Letters.spaced.df,Letters.unspaced.df)

Words.Silent.Adult.df <- read.csv("AdultSilentWordNewSeg2.csv")
Words.Silent.Child.df <- read.csv("KidAdultWordNewSeg2.csv")
Words.Silent.Junior.df <- read.csv("KidWord.csv")
Words.Silent.Adult.df$age <- "adult"
Words.Silent.Child.df$age <- "child"
Words.Silent.Junior.df$age <- "junior"

Words.Silent.Adult.df$mode <- "silent"
Words.Silent.Child.df$mode <- "silent"
Words.Silent.Junior.df$mode <- "silent"

Words.Aloud.df <- read.csv("AdultAloudWordNewSeg2.csv")
Words.Aloud.df$age <- "adult"
Words.Aloud.df$mode <- "aloud"

# Combine the aloud and silent data
#

Words.df <- rbind(Words.Aloud.df, Words.Silent.df)

Words.df$mode <- factor(Words.df$mode)

Words.df$age <- factor(Words.df$age)

Words.df$age <- factor(Words.df$age, levels=c("junior","child","adult"), ordered=TRUE)

Words.df$ability <- ordered(substr(Words.df$RECORDING_SESSION_LABEL,1,1), 
    c("P", "G"),
    labels=c("Poor", "Good"))

Words.df <- Words.df[makeGlobalFixIndex(Words.df$item) ==
   Words.df$CURRENT_FIX_INDEX,]

Words.df <- rename(Words.df, c(
    RECORDING_SESSION_LABEL = "subjn",
    item = "sent",
    CURRENT_FIX_DURATION = "fixdur",
    CURRENT_FIX_INDEX = "nfix",
    CURRENT_FIX_INTEREST AREAS = "word",
    CURRENT_FIX_INTEREST AREA RUN ID = "fixIARunID",
    PREVIOUS_SAC DIRECTION = "sacdir"
))

# Summarise it

summary(Words.df)

# Load Ben's target word frequency data

Words.Freq.df <- rbind(read.csv("FrequencyWords.csv"),
   read.csv("FrequencyWords2.csv"))

Words.Freq.df <- subset(Words.Freq.df,select=c(sent,nw,ns,ne))

Words.Freq.df$lwfreq <- factor(trunc(log10(1+ Words.Freq.df $nw)), ordered=TRUE)

Words.Freq.df$lssfreq <- factor(trunc(log10(1+ Words.Freq.df $ns)), ordered=TRUE)

Words.Freq.df$lsefreq <- factor(trunc(log10(1+ Words.Freq.df $ne)), ordered=TRUE)
Words.Freq.df $lwfreq[Words.Freq.df $lwfreq==2]<-1
Words.Freq.df $lwfreq <- factor(Words.Freq.df $lwfreq, labels=c("Low","High"))

Words.Freq.df $lsfreq <- factor(Words.Freq.df $lsfreq, labels=c("Low","Med","High"))

Words.Freq.df $lefreq[Words.Freq.df $lefreq>4]<-4
Words.Freq.df $lefreq <- factor(Words.Freq.df $lefreq, labels=c("Low","Med","High"))

# Merge these various files together
Words.df <- merge(Words.df, Words.Freq.df, by="sent")
Words.df <- merge(Words.df, Text.df, by=c("sent","word"))

Words.df <- merge(Words.df, Letters.df, by=c("age","mode","subjn","sent","nfix"))

Words.df <- subset(Words.df, select=c(subjn, sent, fixdur, nfix, word.x, word.y, wordcode, fixIARunID, sacdir, target, lwfreq, lsfreq, lefreq, LIW, wonl, wlength.x, lrwfreq, logfreq, pfirst, pmid, plast, avgBigram, ability, mode, spacing, target, age))

Words.df$wlength <- Words.df$wlength.x
Words.df$wlength.x <- NULL
Words.df$target.1 <- NULL

Words.df$fixIndex <- makeFixIndex(Words.df$nfix, Words.df$word.x)

Words.df <- Words.df[ (Words.df$sacdir=="RIGHT" | Words.df$sacdir=="LEFT") & !is.na(Words.df$sacdir),]

Words.df <- Words.df[Words.df$word.x>1 & Words.df$word.x<Words.df$wonl,]
Words.df <- Words.df[Words.df$word.x==Words.df$word.y,]

Words.df <- Words.df[!is.na(Words.df$pfirst),]

Words.ff.df <- subset(Words.df,

  sacdir=="RIGHT" &
  fixIndex==1 &
  fixIARunID==1 &
  plast != -1)

Words.ff.df <- rename(Words.ff.df, c(fixdur="ffixdur"))

Words.ff.df$canon <- 2

Words.ff.df[ Words.ff.df$pfirst > mean(Words.ff.df$pfirst) &
             Words.ff.df$pmid>mean(Words.ff.df$pmid) &
             Words.ff.df$plast>mean(Words.ff.df$plast),]$canon <- 3

Words.ff.df[ Words.ff.df$pfirst < mean(Words.ff.df$pfirst) &
              Words.ff.df$pmid < mean(Words.ff.df$pmid) &
              Words.ff.df$plast < mean(Words.ff.df$plast),]$canon <- 1

Words.ff.df$canon <- factor(Words.ff.df$canon)

Words.ff.df$rcanon <- relevel(Words.ff.df$canon, ref="2")

with (Words.ff.df,
        tapply(ffixdur, list(spacing, mode, age), mean))

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
        tapply(LIW, list(spacing, mode, age), mean))

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
        tapply(LIW, list(mode, age), mean))
with (Words.ff.df[Words.ff.df$LIW!=0,],
       tapply(LIW, list(spacing, age), mean))

# Landing site analysis - adults

print (ls.m1 <- lmer(
       LIW ~ (mode + ability + spacing + pfirst + pmid + plast + wlength )^2 +
             (1|subjn) + (1|sent) + (1|wordcode),
       data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",],
       cor=FALSE)

# omit interactions
print (ls.m2 <- lmer(
       LIW ~ (mode + ability + spacing + pfirst + pmid + plast + wlength )+
             (1|subjn) + (1|sent) + (1|wordcode),
       data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",],
       cor=FALSE)

# Best model - note the pfirst effect now
print (ls.m3 <- lmer(
       LIW ~ (mode + pfirst + wlength )+
             (1|subjn) + (1|sent) + (1|wordcode),
       data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",],
       cor=FALSE)

# Try the canonical variable - nothing happening
print (ls.m3 <- lmer(
       LIW ~ (mode + wlength + rcanon) +
             (1|subjn) + (1|sent) + (1|wordcode),
       data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",],
       cor=FALSE)
data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",], cor=FALSE)

# Avg bigram measure - nothing happening
print (ls.m4 <- lmer(
    LIW ~ (mode + wlength + avgBigram) +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",], cor=FALSE)

p <- ggplot(Words.ff.df, aes(x=wlength, y=LIW, fill=mode))
p <- p + geom_bar(stat="summary",position="dodge", fun.y="mean")
p <- p + coord_cartesian(ylim=c(0,4))
p <- p + scale_y_continuous()
p <- p + opts(title =
    "landing site as a fn of wlength and mode")
print(p)

##############################################
# Landing site analysis - including children
# Looking at age effects (silent only)

print (ls.m1 <- lmer(  
    LIW ~ ( age + ability + spacing + pfirst + pmid + plast + wlength )^2
    +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",], cor=FALSE)

# keep significant interactions
print (ls.m2 <- lmer(
  LIW ~ (  
    age + ability + pfirst + pmid + wlength ) + 
    age:wlength + ability:pfirst + pfirst:pmid + 
    (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",]),
  cor=FALSE)

# Best model
print (ls.m3 <- lmer(
  LIW ~ (  
    age + ability + pfirst + wlength ) + 
    age:wlength + ability:pfirst + 
    (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",]),
  cor=FALSE)

# Looking at mode effects (adults only)
print (ls.m4 <- lmer(
  LIW ~ (  
    mode + ability + spacing + pfirst + pmid + plast + wlength 
  )^2 + 
    (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",]),
  cor=FALSE)

# delete interactions since none are significant
print (ls.m5 <- lmer(
  LIW ~ (  
    mode + ability + spacing + pfirst + pmid + plast + wlength ) + 
    (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",]),
  cor=FALSE)
# best model
print (ls.m6 <- lmer(  
LIW ~ (    mode + pfirst + wlength ) +  
   (1|subjn) + (1|sent) + (1|wordcode),  
data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult",],  
cor=FALSE)

# First fixation analysis
# Looking at age effects (silent only)

print (ff.m1 <- lmer(  
       ffixdur ~ (age + ability + spacing + lrwfreq)^2 +  
       (1|subjn) + (1|sent) + (1|wordcode),  
data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",],  
cor=FALSE)

# omit interactions
print (ff.m2 <- lmer(  
       ffixdur ~ (age + ability + spacing + lrwfreq) +  
       (1|subjn) + (1|sent) + (1|wordcode),  
data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",],  
cor=FALSE)

# Best model - Only age & freq is significant
print (ff.m3 <- lmer(  
       ffixdur ~ (age + lrwfreq )+  
       (1|subjn) + (1|sent) + (1|wordcode),  
data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",],  
cor=FALSE)

# First fixation analysis - mode effects
print (ff.m4 <- lmer(
    ffixdur ~ (mode + ability + spacing + lrwfreq)^2 +
    (1|subj) + (1|sent) + (1|wordcode),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult"],
    cor=FALSE))

# Best model
print (ff.m5 <- lmer(
    ffixdur ~ (mode + spacing + lrwfreq) +
    spacing:lrwfreq +
    (1|subj) + (1|sent) + (1|wordcode),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age=="adult"],
    cor=FALSE)

###########################
# Gaze duration analysis
Words.gd.df <- Words.df[ Words.df$fixIARunID==1 & Words.df$pfirst!=-0,]

Words.gd.df <- subset(Words.gd.df,
    select=c(subjn, sent, wordcode, fixdur, wlength, lwfreq, lrwfreq, logfreq,
    fixIndex, spacing, lsfreq, lefreq, ability, pfirst, pmid, plast, mode, target, age, LIW))

Words.gd.melt.df <- melt.data.frame(Words.gd.df,
    id=c("subj","sent", "wordcode", "wlength", "lwfreq","lrwfreq","fixIndex",
    "spacing","lsfreq", "lefreq", "ability","logfreq","pfirst", "pmid", "plast", "mode", "target",
    "age", "LIW"),
    measures=c("fixdur"))

Words.gd.df <- cast(Words.gd.melt.df,
    subjn + sent + wordcode + lwfreq + wlength + spacing + lsfreq+ lefreq + ability +
    logfreq + pfirst + pmid + plast + lrwfreq + mode + target + age +
    LIW ~ variable,
    sum)
Words.gd.df <- Words.gd.df[!is.na(Words.gd.df$subjn), ]
Words.gd.df <- rename(Words.gd.df, c(fixdur="gazedur"))
Words.gd.df$ability <- ordered(Words.gd.df$ability, c("Poor", "Good"))
Words.gd.df$lsfreq <- ordered(Words.gd.df$lsfreq, c("Low", "Med", "High"))
Words.gd.df$lefreq <- ordered(Words.gd.df$lefreq, c("Low", "Med", "High"))
Words.gd.df$lwfreq <- ordered(Words.gd.df$lwfreq, c("Low", "High"))
Words.gd.df$spacing <- ordered(Words.gd.df$spacing, c("Present", "Absent"))
Words.gd.df$mode <- factor(Words.gd.df$mode, c("silent", "aloud"))
Words.gd.df$age <- factor(Words.gd.df$age, c("adult", "child", "junior"), ordered=TRUE)
Words.gd.df$lrwfreq <- factor(Words.gd.df$lrwfreq, c("low", "high"), ordered=TRUE)

with (Words.gd.df,
       tapply(gazedur, list(spacing, mode), mean))

with (Words.gd.df,
       tapply(gazedur, list(age, mode), mean))

# Create the first model - age effects

print (gd.m1 <- lmer(
   gazedur ~ (age + ability + spacing + lrwfreq)^2 +
               (1|subjn) + (1|sent) + (1|wordcode),
   data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$mode=="silent".],
   cor=FALSE)

# Best model

print (gd.m2 <- lmer(
   gazedur ~ (age + spacing + lrwfreq) +
              age: lrwfreq +
              (1|subjn) + (1|sent) + (1|wordcode),
   data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$mode=="silent".],
   cor=FALSE)

with (Words.gd.df,
       tapply(gazedur, list(spacing, mode), mean))

with (Words.gd.df,
       tapply(gazedur, list(age, mode), mean))

# Create the first model - age effects

print (gd.m1 <- lmer(
   gazedur ~ (age + ability + spacing + lrwfreq)^2 +
               (1|subjn) + (1|sent) + (1|wordcode),
   data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$mode=="silent".],
   cor=FALSE)

# Best model

print (gd.m2 <- lmer(
   gazedur ~ (age + spacing + lrwfreq) +
              age: lrwfreq +
              (1|subjn) + (1|sent) + (1|wordcode),
   data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$mode=="silent".],
   cor=FALSE)
data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$mode=="silent",],
cor=FALSE)

# mode analysis
print (gd.m3 <- lmer(gazedur ~ (mode + ability + spacing + lrwfreq)^2 +
    (1|subjn) + (1|sent) + (1|wordcode),
data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age=="adult",],
cor=FALSE)

# Best model
print (gd.m4 <- lmer(gazedur ~ (mode + ability + spacing + lrwfreq) +
    ability:lrwfreq +
    (1|subjn) + (1|sent) + (1|wordcode),
data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age=="adult",],
cor=FALSE)

p <- ggplot(Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",], aes(x=age, y=LIW, fill=spacing))
p <- p + geom_bar(stat="summary",position="dodge", fun.y="mean")
p <- p + coord_cartesian(ylim=c(0,3))
p <- p + scale_y_continuous()
p <- p + opts(title =
"landing site as a fn of spacing and age")
print(p)
Appendix E3: R Scripts used to analyse Experiment 3

```r
library(lme4)
library(reshape)
library(ggplot2)
source("myFuncs.R")

# Read in data files for letter level
Letters.Silent.Junior.df <- read.csv("KidLetter.csv")
Letters.Silent.Adult.df$age <- "adult"
Letters.Silent.Child.df$age <- "child"
Letters.Silent.Junior.df$age <- "junior"
Letters.Silent.Adult.df$mode <- "silent"
Letters.Silent.Child.df$mode <- "silent"
Letters.Silent.Junior.df$mode <- "silent"

Letters.df <- rbind(Letters.Silent.Adult.df,
                     Letters.Silent.Child.df,
                     Letters.Silent.Junior.df)

Letters.df$age <- factor(Letters.df$age, levels=c("junior","child","adult"), ordered=TRUE)

Letters.df <- Letters.df[makeGlobalFixIndex(Letters.df$item) ==
                          Letters.df$CURRENT_FIX_INDEX,]
```
Letters.df <- rename(Letters.df, c(
  item = "sent",
  CURRENT_FIX_INTEREST_AREA_INDEX = "fixIA",
  CURRENT_FIX_INDEX = "nfix",
  RECORDING_SESSION_LABEL = "subjn")
)

summary(Letters.df)

# Merge in condition variable
Letters.df <- merge(Letters.df, read.csv("BenCondition.csv"), by="sent")

# Distinguish spacing conditions
Letters.df$spacing <- ordered(substr(Letters.df$condition,3,3),
  c("S", "U"),
  labels=c("Present", "Absent"))

# Temporarily split the df to allow merger of different LIW data
Letters.spaced.df <- Letters.df[Letters.df$spacing=="Present",]

# Load in text information data frame
Text.df <- rbind(read.csv("Ben_Sentences3b.csv"), read.csv("Ben_Sentences_Kid3.csv"))
Text.df <- subset (Text.df,
  select=c("sent", "word", "wlength", "wordcode", "target",
           "wfreq", "pfirst", "pmid", "plast","avgBigram"))

Text.df$logfreq <- log10(1+Text.df$wfreq)
Text.df$lrwfreq[Text.df$logfreq<3] <- "low"
Text.df$lrwfreq[Text.df$logfreq>=3] <- "high"
Text.df$lrwfreq <- factor(Text.df$lrwfreq, levels=c("low","high"), ordered=TRUE)

Text.df$savgbigram <- factor(trunc(Text.df$savgbigram), order=TRUE)
Text.df$savgbigram[Text.df$savgbigram<3] <- 3
Text.df$savgbigram[Text.df$savgbigram>4] <- 4
Text.df$savgbigram <- factor(Text.df$savgbigram, ordered=TRUE, labels=c("low","high"))

LIWunspaced.df <- makeUnspacedLIWDF(Text.df$sent, Text.df$word, Text.df$wlength)
LIWspaced.df <- makeSpacedLIWDF(Text.df$sent, Text.df$word, Text.df$wlength)
Letters.spaced.df <- merge(Letters.spaced.df, LIWspaced.df, by.x=c("sent","fixIA"), by.y=c("sent","char"))
Letters.unspaced.df <- merge(Letters.unspaced.df, LIWunspaced.df, by.x=c("sent","fixIA"), by.y=c("sent","char"))
Letters.df <- rbind(Letters.spaced.df,Letters.unspaced.df)

Words.Silent.Adult.df <- read.csv("AdultSilentWordNewSeg2.csv")
Words.Silent.Child.df <- read.csv("KidAdultWordNewSeg2.csv")
Words.Silent.Junior.df <- read.csv("KidWord.csv")
Words.Silent.Adult.df$age <- "adult"
Words.Silent.Child.df$age <- "child"
Words.Silent.Junior.df$age <- "junior"

Words.Silent.Adult.df$mode <- "silent"
Words.Silent.Child.df$mode <- "silent"
Words.Silent.Junior.df$mode <- "silent"
Words.Silent.df <- rbind( Words.Silent.Adult.df,
Words.Silent.Child.df,

Words.Silent.Junior.df)

Words.df$age <- factor(Words.df$age)

Words.df$age <- factor(Words.df$age, levels=c("junior", "child", "adult"), ordered=TRUE)

Words.df$ability <- ordered(substr(Words.df$RECORDING_SESSION_LABEL, 1, 1),
c("P", "G"),
labels=c("Poor", "Good"))

Words.df <- Words.df[makeGlobalFixIndex(Words.df$item) == Words.df$CURRENT_FIX_INDEX,]

Words.df <- rename(Words.df, c(
  RECORDING_SESSION_LABEL = "subjn",
  item = "sent",
  CURRENT_FIX_DURATION = "fixdur",
  CURRENT_FIX_INDEX = "nfix",
  CURRENT_FIX_INTEREST_AREAS = "word",
  CURRENT_FIX_INTEREST_AREA_RUN_ID = "fixIARunID",
  PREVIOUS_SAC_DIRECTION = "sacdir"
))

# Summarise it

summary(Words.df)

# Load Ben's target word frequency data
Words.Freq.df <- rbind(read.csv("FrequencyWords.csv"),
                    read.csv("FrequencyWords2.csv"))

Words.Freq.df <- subset(Words.Freq.df, select=c(sent, nw, ns, ne))

Words.Freq.df $lwfreq <- factor(trunc(log10(1 + Words.Freq.df $nw)), ordered=TRUE)

Words.Freq.df $lsfreq <- factor(trunc(log10(1 + Words.Freq.df $ns)), ordered=TRUE)

Words.Freq.df $lefreq <- factor(trunc(log10(1 + Words.Freq.df $ne)), ordered=TRUE)

Words.Freq.df $lwfreq[Words.Freq.df $lwfreq==2]<-1

Words.Freq.df $lwfreq <- factor(Words.Freq.df $lwfreq, labels=c("Low","High"))


Words.Freq.df $lsfreq <- factor(Words.Freq.df $lsfreq, labels=c("Low","Med","High"))


Words.Freq.df $lefreq[Words.Freq.df $lefreq>4]<-4

Words.Freq.df $lefreq <- factor(Words.Freq.df $lefreq, labels=c("Low","Med","High"))

# Merge these various files together

Words.df <- merge(Words.df, Words.Freq.df, by="sent")

Words.df <- merge(Words.df, Text.df, by=c("sent", "word"))

Words.df <- merge(Words.df, Letters.df, by=c("age", "mode", "subjn", "sent", "nfix"))

Words.df <- subset(Words.df,
                    select=c(subjn, sent, fixdur, nfix, word.x, word.y,
                              wordcode, fixIARunID, sacdir, target, lwfreq, lsfreq, lefreq,
                              LIW, wonl, wlength.x, lwfreq, logfreq, pfirst, pmid, plast,
                              avgBigram, ability, mode, spacing, target, age))
Words.df$wlength <- Words.df$wlength.x
Words.df$wlength.x <- NULL
Words.df$target.1 <- NULL
Words.df$fixIndex <- makeFixIndex(Words.df$nfix, Words.df$word.x)

Words.df <- Words.df[ (Words.df$sacdir=="RIGHT" | Words.df$sacdir=="LEFT") & !is.na(Words.df$sacdir),]

Words.df <- Words.df[Words.df$word.x>1 & Words.df$word.x<Words.df$wonl,]

Words.df <- Words.df[Words.df$word.x==Words.df$word.y,]

Words.df <- Words.df[!is.na(Words.df$pfirst),]

Words.ff.df <- subset(Words.df, sacdir=="RIGHT" &
                      fixIndex==1 &
                      fixIARunID==1 &
                      plast != -1)

Words.ff.df <- rename(Words.ff.df, c(fixdur="ffixdur"))

Words.ff.df$canon <- 2

Words.ff.df[ Words.ff.df$pfirst > mean(Words.ff.df$pfirst) &
              Words.ff.df$pmid>mean(Words.ff.df$pmid) &
              Words.ff.df$plast>mean(Words.ff.df$plast),]$canon <- 3

Words.ff.df[ Words.ff.df$pfirst < mean(Words.ff.df$pfirst) &
Words.ff.df$pmid < mean(Words.ff.df$pmid) &
Words.ff.df$plast < mean(Words.ff.df$plast),]$canon <- 1

Words.ff.df$canon <- factor(Words.ff.df$canon)

Words.ff.df$rcanon <- relevel(Words.ff.df$canon, ref="2")

# Tables to see what we've got.

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
    tapply(ffixdur, list(spacing, mode, age), mean))

with (Words.ff.df,
    tapply(ffixdur, list(ability, spacing, age), mean))

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
    tapply(LIW, list(spacing, mode, age), mean))

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
    tapply(LIW, list(ability, spacing, age), mean))

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
    tapply(LIW, list(mode, age), mean))

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
    tapply(LIW, list(age, wlength), mean))

with (Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
    tapply(LIW, list(age, wlength, spacing), mean))

with (Words.ff.df[Words.ff.df$LIW!=0,],
    tapply(LIW, list(ability, spacing, age), mean))
tapply(LIW, list(spacing, age), mean))

# Landing site analysis - children

# Looking at age effects (silent only)

print (ls.mC1 <- lmer(
  LIW ~ (age + ability + spacing + pfirst + pmid +
        plast + wlength + lwfreq ) +
      (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult",],
  cor=FALSE)

# try interaction

print (ls.mC2 <- lmer(
  LIW ~ (age + ability + spacing + pfirst +
        pmid + plast + wlength )^2 +
      (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult",],
  cor=FALSE)

# clean up

print (ls.mC3 <- lmer(
  LIW ~ (age + ability + spacing + pfirst + pmid + plast + wlength ) +
        age:pfirst + age:wlength + ability:pfirst +
        spacing:pmid + pmid:wlength +
      (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult",],
  cor=FALSE)
#Best model

print (ls.mC4 <- lmer(
  LIW ~ (age + ability + spacing + pfirst + pmid + wlength) +
    age:pfirst + age:wlength + ability:pfirst +
    spacing:pmid + pmid:wlength +
    (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult",],
  cor=FALSE)

#target word only

print (ls.mC5 <- lmer(
  LIW ~ (age + ability + spacing + pfirst +
    pmid + plast + wlength + lrwfreq )+
    (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes" 
    & Words.ff.df$sage!="adult",],
  cor=FALSE)

print (ls.mC6 <- lmer(
  LIW ~ (age + ability + spacing + pfirst +
    pmid + plast + wlength + lrwfreq )^2 +
    (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes" 
    & Words.ff.df$sage!="adult",],
  cor=FALSE)
Ben's freq

print (ls.mBC1 <- lmer(
    LIW ~ (age + spacing + lsfreq + lefreq + wlength )^2 +
    (1|subjn) + (1|sent),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes"  
        & Words.ff.df$age!="adult",],
    cor=FALSE)
)

#clean up - best model

print (ls.mBC2 <- lmer(
    LIW ~ (age + spacing + lefreq + wlength ) +
    age:spacing + spacing:lefreq +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes"  
        & Words.ff.df$age!="adult",],
    cor=FALSE)
)

print (ls.mBC3 <- lmer(
    LIW ~ (age + spacing + lsfreq + lefreq + wlength )^2 +
    (1|subjn) + (1|sent),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes",],
    cor=FALSE)

###################
# Plot Pfirst-age

Words.ff.df$pfirstF <- cut(Words.ff.df$pfirst, 
  c(0, mean(Words.ff.df$pfirst), max(Words.ff.df$pfirst)), 
  labels=c("low","high"))

Words.ff.df <- Words.ff.df[!is.na(Words.ff.df$pfirstF),]

p <- ggplot(Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult",], 
  aes(x=pfirstF, y=LIW, fill=age))

p <- p + geom_bar(stat="summary",position="dodge", fun.y="mean")

p <- p + coord_cartesian(ylim=c(0,4))

p <- p + scale_y_continuous()

p <- p + opts(title =
  "landing site as a fn of age and start letter frequency")

print(p)

# Plot - age/wlength

p <- ggplot(Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult",], 
  aes(x=wlength, y=LIW, fill=age))

p <- p + geom_bar(stat="summary",position="dodge", fun.y="mean")

p <- p + coord_cartesian(ylim=c(0,4))

p <- p + scale_y_continuous()

p <- p + opts(title =
  "landing site as a fn of age and word length")

print(p)

############################
# First fixation analysis

# Create the first model

# Looking at age effects (silent only)

print (ff.m1 <- lmer(  
  ffixdur ~ (age + ability + spacing + lrwfreq)^2 +  
  (1|subjn) + (1|sent) + (1|wordcode),  
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult",],  
  cor=FALSE)

#check Target only - Nothing

print (ff.m2 <- lmer(  
  ffixdur ~ (age + ability + spacing + lrwfreq)^2 +  
  (1|subjn) + (1|sent) + (1|wordcode),  
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes"  
  & Words.ff.df$age!="adult",],  
  cor=FALSE)

###################################

#check Ben’s freq - target only

print (ff.mB3 <- lmer(  
  ffixdur ~ (age + ability + spacing + lsfreq + lefreq + wlength + lwfreq)^2 +  
  (1|subjn) + (1|sent) + (1|wordcode),  
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes"  
  & Words.ff.df$age!="adult",],  
  cor=FALSE)
#clean up

print (ff.mB4 <- lmer(
    ffixdur ~ (ability + lefreq + wlength + lwfreq) +
    ability:lwfreq + lefreq:wlength +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes" & Words.ff.df$age!="adult",],
    cor=FALSE)

#Best model

print (ff.mB5 <- lmer(
    ffixdur ~ (lefreq + wlength) + lefreq:wlength +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$target=="yes" & Words.ff.df$age!="adult",],
    cor=FALSE)

########################################################

#Plot FF1 - Lefreq/wlength

p <- ggplot(Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$age!="adult" & Words.ff.df$target=="yes",],
    aes(x=wlength, y=ffixdur,fill=lefreq))

p <- p + geom_bar(stat="summary",position="dodge", fun.y="mean")

p <- p + coord_cartesian(ylim=c(0,300))

p <- p + scale_y_continuous()

p <- p + opts(title =
"first fixation duration as a fn of end freq and word length")

print(p)

#############################################################################

# Gaze duration analysis

# Select a subset of data from the dataframe for the first fixation analysis

Words.gd.df <- Words.df[ Words.df$fixIARunID==1 & Words.df$pfirst!=-0,]

Words.gd.df <- subset(Words.gd.df,

   select=c(subjn, sent, wordcode, fixdur, wlength, lwfreq, lrwfreq,
            logfreq, fixIndex, spacing, lsfreq, lefreq, ability,
            pfirst, pmid, plast, mode, target, age, LIW))

Words.gd.melt.df <- melt.data.frame(Words.gd.df,

   id=c("subjn","sent", "wordcode", "wlength","lwfreq","lrwfreq","fixIndex",
        "spacing","lsfreq", "lefreq", "ability","logfreq",
        "pfirst", "pmid", "plast", "mode", "target", "age", "LIW"),

   measures=c("fixdur"))

Words.gd.df <- cast(Words.gd.melt.df,

   subjn + sent + wordcode + lwfreq + wlength + spacing + lsfreq+ lefreq + ability +
   logfreq + pfirst + pmid + plast + lrwfreq + mode + target + age +
   LIW ~ variable,

   sum)

Words.gd.df <- Words.gd.df[!is.na(Words.gd.df$subjn),] 

Words.gd.df <- rename(Words.gd.df, c(fixdur="gazedur"))

Words.gd.df$ability <- ordered(Words.gd.df$ability, c("Poor", "Good"))

Words.gd.df$lsfreq <- ordered(Words.gd.df$lsfreq, c("Low", "Med", "High"))

Words.gd.df$lefreq <- ordered(Words.gd.df$lefreq, c("Low", "Med", "High"))

Words.gd.df$lwfreq <- ordered(Words.gd.df$lwfreq, c("Low", "High"))
Words.gd.df$spacing <- ordered(Words.gd.df$spacing, c("Present", "Absent"))

Words.gd.df$mode <- factor(Words.gd.df$mode, c("silent", "aloud"))

Words.gd.df$age <- factor(Words.gd.df$age, c("adult", "child", "junior"), ordered=TRUE)

Words.gd.df$lrwfreq <- factor(Words.gd.df$lrwfreq, c("low", "high"), ordered=TRUE)

with (Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$target=="yes",],
      tapply(gazedur, list(spacing, age), mean))

with (Words.gd.df,
      tapply(gazedur, list(ability, spacing, age), mean))

# Create the first model - age effects

print (gd.m1 <- lmer(
    gazedur ~ (age + ability + spacing + lrwfreq)^2 +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age!="adult",],
    cor=FALSE)

#Best model

print (gd.m2 <- lmer(
    gazedur ~ (age + ability + lrwfreq) +
    age:lrwfreq + ability:lrwfreq +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age!="adult",],
    cor=FALSE)
#check target word - NOTHING

print (gd.m3 <- lmer(
   gazedur ~ (age + ability + spacing + lrwfreq)^2 +
            (1|subjn) + (1|sent) + (1|wordcode),
   data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age!="adult"
                    &Words.gd.df$target=="yes",],
   cor=FALSE)

#################################################################

#try Ben's freq - target only

print (gd.mB3 <- lmer(
   gazedur ~ (age + ability + spacing + lsfreq + lefreq + lwfreq)^2 +
            (1|subjn) + (1|sent) + (1|wordcode),
   data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age!="adult"
                    &Words.gd.df$target=="yes",],
   cor=FALSE)

#Best model

print (gd.mB4 <- lmer(
   gazedur ~ (ability + spacing + lsfreq + lefreq) +
            ability:spacing + lsfreq:lefreq +
            (1|subjn) + (1|sent) + (1|wordcode),
   data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age!="adult"
                    &Words.gd.df$target=="yes",],
   cor=FALSE)
print (gd.mB5 <- lmer(
    gazedur ~ (age + ability + spacing + lsfreq + lefreq + lwfreq) +
    (1|subjn) + (1|sent) + (1|wordcode),
    data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$target=="yes",],
    cor=FALSE)

#Plot GD - lrwfreq&age
p <- ggplot(Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age!="adult",],
    aes(x=lrwfreq, y=gazedur, fill=age))
p <- p + geom_bar(stat="summary",position="dodge", fun.y="mean")
p <- p + coord_cartesian(ylim=c(0,400))
p <- p + scale_y_continuous(breaks=(1:10)*50)
p <- p + scale_y_continuous()
p <- p + opts(title =
    "gaze duration as a fn of word freq and age")
print(p)

#Plot GD - lrwfreq&age
p <- ggplot(Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$age!="adult",],
    aes(x=lrwfreq, y=gazedur, fill=age))
p <- p + geom_bar(stat="summary",position="dodge", fun.y="mean")
p <- p + coord_cartesian(ylim=c(0,400))
p <- p + scale_y_continuous(breaks=(1:10)*50)
p <- p + scale_y_continuous()
p <- p + opts(title =
    "gaze duration as a fn of word freq and age")
print(p)

#C.fdiff <- matrix(c(2/3, -1/3, -1/3,
#                    1/3,  1/3, -2/3),
#                  3, 2,
#                  dimnames = list(c("Junior","Child","Adult"),
#                                   c("J<>C","C<>A")))

c.fdiff <- matrix(
    c(2/3, -1/3, -1/3,
      1/3, 1/3, -2/3),
    3, 2,
    dimnames = list(c("Junior","Child","Adult"),
                    c("J<>C","C<>A")))

c.fdiff

contrasts(Words.ff$age) <- c.fdiff
summary(Words.ff.df)

# LIW

# Best model

print (ls.m3b <- lmer(  
  LIW ~ ( age + ability + pfirst + wlength ) +  
  age:wlength + ability:pfirst +  
  (1|subjn) + (1|sent) + (1|wordcode),  
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",],  
  cor=FALSE)

# FF

print (ff.m1b <- lmer(  
  ffixdur ~ (age + ability + spacing + lrwfreq)^2 +  
  (1|subjn) + (1|sent) + (1|wordcode),  
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",],  
  cor=FALSE)

# omit interactions

print (ff.m2b <- lmer(  
  ffixdur ~ (age + ability + spacing + lrwfreq) +  
  (1|subjn) + (1|sent) + (1|wordcode),  
  data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",],  
  cor=FALSE)

# Best model - Now only freq is significant

print (ff.m3b <- lmer(  
  ffixdur ~ lrwfreq +  
  (1|subjn) + (1|sent) + (1|wordcode),
data=Words.ff.df[Words.ff.df$LIW!=0 & Words.ff.df$mode=="silent",],
cor=FALSE)

# GD

print (gd.m1b <- lmer(
  gazedur ~ (age + ability + spacing + lrwfreq)^2 +
          (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$mode=="silent",],
cor=FALSE)

clean up

print (gd.m1b <- lmer(
  gazedur ~ (age + lrwfreq) +
          age:lrwfreq +
          (1|subjn) + (1|sent) + (1|wordcode),
  data=Words.gd.df[Words.gd.df$LIW!=0 & Words.gd.df$mode=="silent",],
cor=FALSE)