A developmental exploration of Chinese reading in a population of early readers: from eye movement control to textual coherence

Xi Fan

A dissertation submitted for the degree of

Doctor of Philosophy

Department of Computer Science

Faculty of Science and Engineering

National University of Ireland Maynooth

February 2020

Head of Department: Dr. Joseph Timoney

Supervisor: Prof. Ronan G. Reilly

Acknowledgements

I would like to express my sincere gratitude to my supervisor Professor Ronan Reilly for the patient guidance, encouragement and advice he has provided throughout my time as his student. I have been extremely lucky to have a supervisor who is my role model in every respect.

I must also express my gratitude to Ms Chen for her support and encouragement, without her help I probably wouldn't have started this journey. I would also like to thank Dr. Joseph Timoney, head of the Department of Computer Science, who supported me with a budget for travelling to conferences. My sincere thanks also goes to my fellow lab mates for the helpful discussions and for all the fun we have had in the last four years. In particular I would like to thank Ms Niamh Dootson for the suggestions she made in reference to Chapter 5 of the thesis. A very special gratitude goes to John and Pat Hume scholarship scheme for the financial support it provided me.

I thank the enlightened leaders in Guangzhou Medical University for giving me the chance to pursue my PhD degree and to Huilai Ying Nei Primary school for allowing me to collect data in the school.

Last but not the least, I am grateful to my parents, who have provided me moral and emotional support in my life. I am also grateful to my other family members and friends who have supported me along the way.

Abstract

The research described in this thesis explores the developmental characteristic of Chinese reading in a population of young readers with different reading ability. There are two in-depth explorations described: (1) an exploration of the processes underlying eye movement control in Chinese reading; and (2) the use of semantic similarity measures based on distributed representations of words, sentences, and paragraphs to assess the impact of supra-lexical constraints on eye-movements in beginning readers of Chinese. The main results show that the most likely account of processes underlying eye movement control in Chinese reading is a two-factor process whereby the character is the main driver for longer saccades and that the word plays a role in shorter ones. A small-scale extension to the Glenmore model of eye movement control in reading is proposed to account for the saccade targeting patterns of readers. It provides an integrated account of the dynamic interaction of the two factors and demonstrates that a model architecture facilitating a dynamic interaction between top-down lexical and orthographic constrains and bottom-up visual inputs is suited to account for Chinese readers' eye movement. Results also showed that text similarity measures have a significant impact on the moment-to-moment processing of words in reading. Consequently, these factors need to be incorporated into any realistic model of Chinese reading. An additional study is described where a character confusion matrix was generated that can be used as a resource for both pedagogical and psycholinguistic studies of Chinese reading and other studies interested in character recognition.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation for the research

Effective teaching of reading is a strategic priority for all governments. However, large-scale reading intervention programmes aimed at improving the teaching of reading have had limited success. The billion dollar Reading First programme in the US, for example, failed to demonstrate any improvements in reading comprehension measures (Gamse et al., 2008). The final evaluation report concluded that:

> The study finds, on average, that after several years of funding the Reading First program, it has a consistent positive effect on reading instruction yet no statistically significant impact on student reading comprehension (Gamse et al., 2008, p. xviii).

In a further evaluation of four intervention programmes, James-Burdumy et al. (2009) found that rather than helping to improve reading comprehension, paradoxically the interventions had had the opposite effect:

> Reading comprehension test scores were not statistically significantly higher in schools using the selected reading comprehension curricula than in control schools. In fact, students' reading comprehension test scores were sta-

> tistically significantly lower in treatment schools than in control schools (James-Burdumy et al., 2009, p. xxxii).

Several researchers have argued that the root of the problem is the failure of the interventions to take account of the complex nature of the interaction between the intervention and the developing child. These are referred to as child-by-instruction interactions (Connor et al., 2009; Lonigan & Phillips, 2009). For example, the association between vocabulary size and literacy has been well established (Anderson & Freebody, 1981), but efforts to increase vocabulary have not lead to sustained and generalisable comprehension gains. It appears that vocabulary size or indeed any other snapshot-like measure of reading attainment must be interpreted as part of a dynamic and adaptive model of the reader and more research in reading of individual readers or finely sorted groups of readers rather than aggregated assessments are needed.

Given much of our understanding of the reading process is built on alphabetic writing systems, which most likely provide a biased perspective on how the reader's brain works. An exploration of reading in a widely-used logographic writing system, namely Chinese, on young readers with different reading ability would be an effective complement to the science of teaching and learning to read at a fine-grained group level.

Moreover, If we are to understand a brain process as complex as reading, not only to understand how cognitive and perceptual processes (e.g., visual information processing, word recognition, attention, and oculomotor control) work together to perform the complex task of reading, we also need to be able to model it computationally to some degree. Therefore, the ultimate goal of the thesis is to study reading in Chinese in a population of young readers using the empirical eye movement data and computational modelling tools. Both a broader understanding of the reading process and an ability to frame that understanding as a computational model are two complementary parts that are important not only in the reading domain, but also in all complex cognitive processes.

1.2 Research questions

This thesis focuses on a detailed study of Chinese reading in a population of young readers. The overall goal of the research is to provide a picture of the changes that occur in the visual aspects of the reading process as readers increase their reading skills. An ancillary goal has been to develop a set of conceptual and analytic tools to support the study of beginning readers. These tools involve techniques for the analysis of various viewing time measures, the application of a new approach to assessing text coherence, broadly construed, and the use of computational models to integrate and understand the processes underlying saccadic control in Chinese reading.

More specifically, the questions addressed are:

- *What are the characteristics of Chinese readers at different points in the reading ability spectrum from poor to good and how do they change from 4th to 5th grade?*
- *What factors affect eye movement control in reading Chinese?*
- *How can we use computational models to explore Chinese reading?*
- *Can we measure the impact of supra-lexical influences on readers' viewing times?*
- *Can we use the theoretical insights afforded by the research and the model to pinpoint interventions for improving reading of Chinese young readers?*

1.3 Outline of the thesis

Chapter 2 provides background on eye movements and reading research, reviews some basic characteristics of eye movements in reading alphabetic writing systems and

Chinese.

Chapter 3 explores the developmental characteristics of eye movements in Chinese reading among a population of young readers with different reading ability at 4th and 5th grade.

Chapter 4 explores the processes underlying eye movement control in Chinese reading based on the data collected in the experiment described in the previous chapter. Various proposals to explain the underlying mechanisms involved in eye movement control are examined and the chapter concludes that the most likely account is of a two-factor process whereby the character is the main driver for longer saccades and that the word plays a role in shorter saccade.

Chapter 5 describes the construction of a discrimination matrix for the 3,500 most frequently used Chinese characters based on saccade latency. Based on the subset of data collected in the experiment a convolutional network was used to build a model to predict the data for the remaining comparisons. The aim was to generate a confusion matrix that could be used as a resource for both pedagogical and psycholinguistic studies of Chinese reading and other studies relating to character recognition.

An extension to the Glenmore model of eye movement control in reading (Reilly & Radach, 2006) is proposed in chapter 6 to account for the saccade targeting patterns of readers of Chinese writing system. It provides an integrated account of the dynamic interaction of the two factors described in chapter 4, and demonstrates that a model architecture facilitating dynamic interaction between top-down lexical and orthographic constrains and bottom-up visual inputs is best suited to account for Chinese readers' eye movements.

Chapter 7 describes the use of semantic similarity measures based on distributed representations of words, sentences, and paragraphs (so-called "embeddings") to assess the impact of supra-lexical constraint on eye-movement from early readers of Chinese. In addition, a corpus-based measure of surprisal was used to assess the impact of local word predictability. Results indicated that the text similarity measures have a significant impact on the moment-to-moment processing of words in reading. Consequently, the factors need to be incorporated into any realistic model of Chinese reading.

Chapter 8 highlights the key findings of the research described in this thesis. It also proposes a way of building a computational model of reading in Chinese by the utilisation of results from the research.

Chapter 2

Eye Movements in Reading

2.1 Eye movement characteristics

2.1.1 Basic eye movement events

Fixations and saccades are two basic events during eye movements in reading. The eye is assumed to remain relatively still during a fixation and to make a rapid, ballistic movement during a saccade. Fixation duration and saccade length are the most widely used eye movement measures in the reading literature (Rayner, 1998).

A fixation is a brief period of time when the eye is relatively static in one spatial location in the visual field. In the case of English reading, the average fixation duration for skilled readers is between 200 and 250ms. However, there is significant variability in average fixation durations both between and within subjects. The fixation duration of the same reader when reading a passage can range from 100 to 500ms (Rayner, 1978). In the case of Chinese reading, the average fixation duration of skilled reader in various studies ranges from about 230ms to 279ms (Inhoff & Liu, 1998; Yan et al., 2006).

Saccades refer to the movement of the eye from one fixation location to the next. Saccades are ballistic in nature, so once the saccade is launched, the eye cannot change direction until the saccade lands. A typical saccade lasts for between 20 to 40 ms and during this time vision is suppressed so that no new information is acquired (Rayner et al., 2001). Saccades account for approximately 10% of reading time. In reading English, the average saccade extent is between eight to nine character spaces (or about 2° of visual angle). Again, saccade lengths are also highly variable and can range from 2 to 18 characters for a single reader in a single passage of text (Rayner, 1978, 1998).

2.1.2 Eye movement measures

Using eye-tracking technology, researchers can measure the eye movement characteristics of readers during the reading process and use the eye movement data to understand the nature of human vision and the underlying hypothesised cognitive process involved in decoding the written word into a meaning representation. There are different types of measures of fixations and saccades that provide a distinct insight into cognitive and perceptual processes.

However, as noted by Inhoff and Radach (1998), in spite of the increasing popularity and methodological promise of eye-movement measures, there are currently no generally agreed standards that define basic oculomotor events. Researchers have to select appropriate variables to capture desired effects. McConkie et al. (1991) pioneered the analysis of children's eye movements at the word level, including fine-grained analyses of saccade landing sites within words, and the first quantitative analyses of relations between eye-movement parameters and psychometric reading assessments. This was also one of the first studies in which the now common decomposition of viewing times into initial fixation duration, gaze duration and re-reading time was applied to research on developing readers. The idea of this decomposition is to delineate the time course of word processing into time intervals that reflect early orthographic and lexical processing (initial fixations), full lexical access (re-fixation times) and time spent on integration of word meanings into representations at the sentence and text level (rereading times) (see Radach & Kennedy, 2004, Vorstius et al., 2014, for a recent discussion of these measures). The analyses of word viewing times in the thesis will be in line with this logic. The following are measures for which there is reasonable consensus for the indexing of perceptual and cognitive processes and will be applied in the study described here.

First fixation duration First fixations comprise the time of the first fixation on a target word, provided that word is fixated during the first pass of text reading (i.e., the initial reading of text consisting of all forward fixations) (Inhoff & Radach, 1998). Inhoff (1984) initially defined the measure. It can reflect the early orthographic and lexical processing (Inhoff & Radach, 1998). There is evidence that linguistic factors such as orthographic properties, phonological properties, lexical properties, metaphorical status and contextual constraints have significant effects on First fixations (Inhoff, 1984; Inhoff et al., 1984; Inhoff & Rayner, 1986; Inhoff & Topolski, 1994; Lima & Inhoff, 1985; Pollatsek et al., 1992; Rayner & Duffy, 1986). However, it has been criticised for confounding instances in which the target receives a single fixation and instances in which it receives multiple fixations (O'Regan, 1990, 1992). In this thesis, first fixation duration is defined as the duration of the first fixation on the word during the first pass (Vorstius et al., 2014).

Refixation duration Refixation durations refer to summed duration of additional fixations within the current pass and prior to an exit from the word (Inhoff & Radach, 1998). There is evidence that the probability of refixation is modulated by lower-level visual factors such as landing position in a word. The probability of a refixation is less if the initial fixation lands near the optimal viewing position, which is defined as the position near the middle of the word (O'Regan, 1992; Rayner, 1998). Other evidence suggests that the probability and the location of refixations are influenced by linguistic factors such as the morphemic properties of words (Hyönä & Pollatsek, 1998). However, refixations have often been filtered out by eye movement control theories (Engbert et al., 2005; Reichle et al., 2003; Reilly & Radach, 2006). A refixation in the present document is the summed duration of additional fixations within first pass reading.

Re-reading duration Re-reading durations refer to the sum of durations spent in re-reading a target word after first pass reading (Inhoff & Radach, 1998). Relatively few studies have examined eye movement control for re-reading compared with studies of reading during first pass sentence reading. It is considered to reflect general comprehension difficulty (White et al., 2017). Rereading duration may reflect linguistic processing difficulty related to text comprehension (Clifton et al., 2007; Just & Carpenter, 1980)) or where readers fail to identify previously read words (Bicknell & Levy, 2010). Re-reading in this study is the the sum of durations spent re-reading a target word after the first pass word reading.

Saccade latency Saccade latency refers to the period related to the planning and execution of a saccade, which requires a minimum of 150-175 ms (Rayner et al., 1983; Salthouse et al., 1981). In this study, saccade latency is used to measure Chinese character discriminability. Latency in this context is the time between when the visual information is available and when the eye moves, i.e., the delay between perception and action (see chapter 3 for more detailed discussion).

saccade movement measures Besides saccade latency, other types of saccadic measures are reported as well: (a) saccade length; (b) landing positions of saccades; and (c) launch distance of saccades. Saccade length, landing position and launch distance are usually reported in character spaces in alphabetic writing systems and half character spaces in Chinese. In this thesis, they are reported in 0.1 character spaces.

Skipping rate Not all words are fixated during reading. A word is considered skipped when there are no fixations on it or in the space prior to it during first pass, right-directed reading. There is, however, ample evidence that words are processed even though they may have been skipped (Fisher & Shebilske, 1985). Evidence has shown that the skipping rate of target words is mainly determined by target word length and the launch distance of a rightward saccade to the target word. The target word is more likely to be skipped when it is short or when the launch distance of a incoming saccade is nearer to its beginning (McConkie et al., 1994; Vitu et al., 1995). There is also evidence that skipping rate is modified by linguistic factors, with high frequency words more likely to be skipped than low frequency words (Inhoff & Topolski, 1994).

Regression rate Eye movements that lead to the re-reading of text are referred to as regressions. As noted earlier, there is considerable evidence that regressions reflect readers' difficulty in text comprehension (Schotter et al., 2014; Vauras et al., 1992; White et al., 2017). It is also found that regressive saccades are more likely to occur following longer forward saccades (Vitu et al., 1998). In this study, regression rate is computed for all the regressive saccades including saccades that return to a word being read or a previously fixated word.

2.2 Eye movements in reading alphabetic writing systems

Eye movements in reading alphabetic language is usually studied in terms of when the eyes move and to where the eyes move (Aslin & Shea, 1987; Becker & Jürgens, 1979; Rayner & McConkie, 1976).

2.2.1 When the eyes move

The processing time for a fixated word is associated with a number of variables related to word identification factors and high-level factors such as syntactic, semantic, pragmatic, and world knowledge (Clifton et al., 2007). Word processing effects on fixation duration are well-established. Many studies have demonstrated that the length of time readers fixate a word is significantly modulated by the word's cultural frequency, with more time spent fixating low-frequency words (Altarriba et al., 1996; Inhoff & Rayner, 1986; Juhasz & Rayner, 2003; Raney & Rayner, 1995). There are also frequency effects associate with the words prior to and after the fixated word, which are known as spill-over effects (Rayner & Duffy, 1986) and parafoveal preview effects (Chace et al., 2005; Kennison & Clifton, 1995). To be specific, a low-frequency preceding word tends to inflate the processing time of the fixated word and a valid preview of a high-frequency word prior to its fixation may facilitate its processing time. Beside word frequency, word length, predictability, familiarity and lexical ambiguity are also found to impact on when readers move their eyes (Boston et al., 2008; Chaffin et al., 2001; Just & Carpenter, 1980; Kliegl et al., 2004; Rayner & Duffy, 1986; Rayner & Well, 1996).

How fixation duration is affected by syntactic, semantic and discourse factors is less well understood and the research focus has mainly been on syntactic ambiguity, syntactic anomaly or syntactic complexity etc. (Boland & Blodgett, 2001; Frisson et al., 2005; Hyönä & Vainio, 2001). Frazier and Rayner (1982) discovered that longer fixation durations were associated with the first fixation in the region of the sentence that disambiguate the sentence, suggesting that the human sentence-parsing mechanism operates in a rather systematic fashion, immediately computing the structural consequences of fixated material for the analysis of preceding material. Clifton et al. (2007) suggest high-level variables that affect sentence interpretation are much more complex and calls for the development of more explicit theories than existing ones of how syntactic, semantic, pragmatic, and real-world knowledge guide language comprehension. A major challenge in this area is to develop quantitative measures of text complexity at a variety of levels that can be used to predict eye movement behaviour. See chapter 7 for a detailed discussion of the attempt to extend research on the supra-lexical effects on Chinese reading.

2.2.2 Where the eyes move

In fluent reading, the question of where the eyes move can be specified as to which word and to which position within the word the eyes move. Findings from alphabetic reading literature suggest that word boundary spacing is the primary influence in deciding where to move next. Neither removal of spaces between words nor filling them seriously decreases saccade length (McConkie & Rayner, 1975; Pollatsek & Rayner, 1982; Rayner, 1986). Therefore, most theories of eye movement control during the reading of alphabetic scripts assume that the word is the default saccade target unit facilitated by spaces between words, even though there is no consensus on which word to move to next among different theories. Research has also indicated that the length of the current fixated word and the word to the right of the fixation are important cues in deciding where to move next by modulating the saccade length. (O'Regan, 1980; O'Regan, 1979). Although there is variability in where the eyes land on a word, considerable evidence suggests that somewhere between the beginning and the centre of a word tend to be the preferred viewing position (PVL) (Rayner, 1979; Rayner et al., 1998; Vitu et al., 1995). Low-level visual information such as spaces between words and word length appear mainly to influence where the eyes move. However, the role of other high-level properties in influencing where to look next is less clear (Rayner, 1998). Note that the eyes do not always move forward word by word along the text, they also regress, refixate the current word, or skip words.

Neither space nor any other notation is used to indicate word boundaries in written Chinese. What factors influence where the eye fixates and if a PVL exists in written Chinese is still an open question. A possible saccadic targeting strategy in reading Chinese will be discussed in chapter4.

2.2.3 Developmental changes and individual difference in eye movements

There are consistent research findings that as a reader's reading skill increases, fixation duration decreases, saccade length increases, the number of fixations decreases, and the frequency of regressions decreases (Buswell, 1922; McConkie et al., 1991). The reading ability effect is observed either in developmental changes of a reader (Rayner, 1985) or the individual difference among readers (Gilbert, 1959a, 1959b). However, it has been found that readers of similar reading ability can have considerable variability in their eye movements. Gilbert (1959b) compared the eye movement pattern of two poor readers. One reader read at the rate of 186 words per minute making 80 fixations per 100 words. The other reader read at 176 words per minute but made 53 fixations per 100 words. It was also found that children in their first year of reading tend to send their eyes to the middle of a word which is similar to adult reading in alphabetic scripts. (McConkie et al., 1991).

2.3 Eye movements in reading Chinese

2.3.1 Major feature of written Chinese

Unlike alphabetic writing systems such as English, the most noticeable property of written Chinese is that texts are formed by monospaced, square-shaped character units. A character is the basic perceptually prominent unit of Chinese and represents a single syllable, the fundamental unit of meaning. There are 3,500 most frequently used characters in mainland China (Ministry of Education of the People's Republic of China, 1988). The 3,500 characters differ visually in their number of strokes (1-24, with 78% characters has 6-13 strokes), number of radicals (1-8, with 78% characters consist of 2-3 radicals) and the manner of construction (integral, top-bottom, left-right, half-surround, surround, with 58% left-right structure and 23% top-bottom structure) (Chinese character information dictionary, 1988).

It is also important to note that characters can be used independently as a word or be combined with other characters to form a word. Approximately 12% of Chinese words comprise single characters, 74% two characters, 8% three characters, 6% four or more characters and less than 0.3% longer than four characters (Modern Chinese Frequency Dictionary, 1986). As many characters are polysemic characters, the meaning of a Chinese character in text can be context-dependent. For example, "花" has several meanings itself (e.g., "flower", "colourful", "fake", "vague", or "to spend"), it can also join other characters to form words of very different meanings such as "花生" meaning "peanut", "花甲" meaning "sixty years", "花招" meaning "tricks", "花眼" meaning "blurred eyesight" and "花销" meaning "expense". Thus, context is particularly important in helping to disambiguate the meaning of a character (Chen, 1992). However, even though Chinese words are generally less ambiguous than characters, since no spaces or other notation are used to indicate word boundaries, there is sometimes disagreement on word segmentation.

2.3.2 Research on eye movements in reading Chinese

Recent studies of eye movements in reading Chinese focus on how variables related to word identification factors influence the decision of when to move the eyes. Word spacing is found to only facilitate word recognition in Chinese reading for poor readers such as students with learning difficulties (Bai et al., 2015). Bai et al. (2008) investigated native Chinese readers' eye movements as they read text in four conditions: normal unspaced text, text with spaces between words, text with spaces between characters that gave rise to non-words, and text with spaces between every character. The results showed that spaces between words neither hindered nor facilitated reading. Shen et al. (2010) investigated third graders with the same condition and found a similar result, i.e., spaces between words neither hindered nor facilitated reading for third graders. They also found that compared with third graders who have good study performance, third graders with poor study performance rely more on the text's visual information. Similar to findings in alphabetic writing systems, factors such as word frequency, familiarity, predictability and semantic transparency are found to be important factors (Rayner et al., 2005; Yan et al., 2006; Zhang & Yan, 2005).

There is a limited amount of research investigating how the decision about where to move the eyes is affected. Hongxia et al. (2014) investigated character stroke effect on fixation location. They found that when the first character of two character words tend to get more fixation if it has greater number of strokes. Subsequent studies also observed similar stroke effect (Li et al., 2019; Ma & Li, 2015). However, according to Zang et al. (2013)'s study on symmetrical one character and two character words, the manner of character structure may not have any effect on fixation location. Word frequency, predictability and plausibility are also found to have no effect on the first fixation location of target words (Yen et al., 2008; Zang et al., 2012). Research has also indicated that reading ability and age may not affect the fixation landing position in Chinese reading. Zang et al. (2013) examined third graders and adults' eye movement behaviour when reading word-spaced and normal Chinese text. The results showed that the overall fixation location pattern are very similar for third graders and adults reading spaced and unspaced text. Bai et al. (2011) tested dyslexic 5th graders together with their age-matched group and reading ability-matched group under both the word spaced and normal unspaced text, they found no significant difference between all three groups under the two presentation conditions in fixation landing positions. Given the complexity of written Chinese and based on current research, there is currently no consensus on how readers of Chinese select saccade targets. Whether saccades target a character or a word is still an open question.

Chapter 3

Developmental Characteristic of Eye Movements of Young Readers with Different Reading Ability

3.1 Background

There is converging evidence of reading development of children of different age in alphabetic scripts: as children became older, on average fixation durations decrease, saccade lengths increase, the number of fixations decreases, and the frequency of regressions decreases (Blythe & Joseph, 2011; Blythe et al., 2009; Joseph et al., 2009; McConkie et al., 1991). The trajectories of reading development of children is the same when children of different ages read either age-appropriate text (Blythe et al., 2006; McConkie et al., 1991) or the same text (Blythe & Joseph, 2011; Blythe et al., 2009). This research has also found that basic eye movement pattern during reading of children around 11 years old is very close to that of adults.

Feng et al. (2009) conducted a cross-language investigation that compared the development of eye movements of English and Chinese speakers when they read stories in their native language. Participants were third-grade, fifth-grade, and undergraduate English and Chinese students. The results showed that the trajectory of reading development is similar for both languages: mean fixation durations decreased significantly with age for both American and Chinese children; the length of forward saccades increased with age and was longer for Chinese than for English readers; there were no age difference in the proportion of regression fixations.

Given the limited amount of research that has focused on children and the developmental aspects of reading in Chinese, before an in-depth analysis is presented of where and when Chinese young readers move their eyes, the general characteristics of their eye movement data combining cross-sectional and longitudinal comparisons are described in what follows.

3.2 Experiment design and method

3.2.1 Participants

The study involved two large-scale longitudinal data collections at Huilai Yingnei primary school in Guangdong Province, China. Data was collected from the same cohort of students in 2017 and 2018. Raven's Standard Reasoning Test, normed for China (Zhang & Wang, 1989), and the Literacy Test for the Primary School Students (Wang & Tao, 1993) were administered to all 768 grade 4 students attending the school as intelligence and literacy tests, respectively. The results of the Chinese language end-of-term exam, which was administered by the school just prior to the experiment, was also used as a participant selection criterion. Forty eight students with the lowest literacy test scores and below average Chinese term exam scores were selected as the "poor" reader group. The "good" reader group came from 20 students of similar age and intelligence level but with the highest literacy test scores and above average Chinese term exam scores. The "average" reader group comprised 25 students of similar age and intelligence level but with literacy test scores within

Table 3.1: Group comparison of the selection criterion and the number of participants each year

Selection Criterion	Poor	Avg	Good
Age	9.8 (0.5)	9.7 (0.4)	9.8 (0.5)
Literacy Test	793.1 (226.8)	1198.3 (73)	1937.1 (122.7)
Chinese language end-of-term exam	48.1 (16.9)	70.4 (4)	89 (6.9)
Raven's Standard Reasoning Test	43.3 (3.2)	43.5 (2.5)	45.4 (4.3)
2017	45	23	20
2018	41	21	19

+/− 0.3 standard deviation (SD) of the mean and with Chinese term exam scores within +/− 0.5 SD of the mean. Finally, all of the students in the study had normal or corrected vision and had not participated in any prior eye movement studies or similar reading tests.

A total of 93 participants completed the first experiment in 2017, 88 out of the 93 participants completed the second experiment in 2018. Each experiment took between 30 and 40 minutes per participant. However, due to poor data quality, and participants being transferred to other schools, data from 88 Grade 4 students and 81 of the same cohort in Grade 5 were used in the final analyses. Table 3.1 shows the group comparison of the selection criterion and the number of participants each year. They were asked to read texts on a computer screen while their eye movements were being monitored.

3.2.2 Materials

Reading material were a translation of age appropriate short stories from the Florida Instruction for Assessment in Reading (FAIR) toolkit (Florida Department of Education, 2009). Five and six stories were used in the first and second data collections, respectively. Each story was presented in multiple paragraphs (3-4 paragraphs), with each paragraph consisting of 5-7 lines (Stenner et al., 2007). Word length ranged from 1 to 5 characters with a total of 4610 characters, 592 unique. Table 3.2 shows the detail of the material. Stories were presented on the screen one by one. Each story was followed by three comprehension questions. Paragraphs were displayed in black on a light grey background using a 19.5-inch flat-panel monitor. Display resolution was set to 1024 × 768 pixels with a refresh rate of 60 Hz. Texts were presented in Xinhei font at 30 px, left aligned, and double-spaced. Viewing distance was adjusted to 68 cm. At this distance, each character subtended approximately 1° of visual angle laterally. Viewing was binocular and eye movements of the participants' right eye were recorded using the EyeLink 1000 (SR Research Ltd., 2010), with a sampling rate of 500Hz.

Table 3.2: (Details of reading materials)

Grade	**Title**	**No. of paragraphs**	**lines per paragraph**	**sents per paragraph**
4	垃圾桶害虫	3	5,6,5	8,8,4
4	卡拉赢得了比赛	4	5,5,5,6	8,8,7,8
4	牧羊犬	3	6,7,7	8,8,9
4	假期	3	6,6,6	7,6,5
4	王蛇	3	5,5,5	9,5,8
5	垃圾桶害虫	3	5,6,5	8,8,4
5	建城堡	3	5,5,5	6,8,7
5	体操运动	3	6,6,5	8,7,5
5	喂长颈鹿	4	6,6,5,7	10,9,8,9
5	如何做橡皮泥	3	5,6,6	10,9,11
5	丛林花园	3	6,5,5	7,6,8

3.2.3 Procedure

Participants were seated in front of the presentation monitor and received directions for the upcoming task on the display screen in front of them. Participants received identical directions for the reading task, instructing them to read every text so that they understood its meaning and were able to answer comprehension questions. They were advised to read silently. It was also explained that this was not a reading test or contest in order to make them feel more comfortable about the situation. Participants read four three-line practice trials from one story before the reading experiment, to familiarise them with the calibration routine and eye tracking procedures. A 9-point calibration was performed at the beginning of each story. For some participants extra calibrations were needed during the experiments due to head movements. Mean average position error in an accuracy validation routine did not exceed 0.33° of visual angle. A drift-check before every paragraph ensured accuracy between calibrations. If the drift check showed a deviation from more than 0.33° of visual angle, an additional calibration was performed. These settings have proven to produce accurate and reliable data in multiple reading studies across different laboratories (see Inhoff & Radach, 1998, McConkie et al., 1991 for detailed discussions of methodological issues). Children could take breaks between tasks or before calibrations, if necessary. Reading was self-paced and children pressed a mouse button to signal that they were done with a trial. The next paragraph or the comprehension question appeared immediately following the mouse press.

3.3 Data pre-processing

Fixation in which participants blinks, fixations on the first and last words of each line, the first and last fixations of each trial and switch line fixations were excluded from analyses. Extreme fixation duration values less than 80 ms and of greater than 800 ms were discarded (Inhoff & Radach, 1998). Saccades located exactly at the word boundaries were also excluded. Since landing site data involving the first pass fixations in a word resulting from progressive saccades were primarily of interest, all other saccade data were excluded from analysis. Extreme saccade values greater than seven character space were discarded (Inhoff & Radach, 1998) (account for 11% of the first pass progressive saccade). A total of 221,464 fixations and 71,377 saccades contributed to the analysis. Note that the Jieba Chinese text segmentation algorithm was used to do the word segmentation (Sun, 2013). Given there is sometimes disagreement on word boundaries in Chinese, the Modern Chinese Word Dictionary (2007) was used as an arbitrator in determining the precise length of words in the experimental materials.

3.4 Analysis

3.4.1 Developmental analysis

A global analysis was carried out first in order to make developmental comparisons between the eye movements of readers when they are in different grade. Given it is still unclear if a character or a word is a saccade target, reading speed and skipping rate were calculated at both character and word level. Table 3.3 summarises the means and standard deviations of first fixation durations (FFD) , refixation durations (RFD) , re-reading durations (RRD), saccade lengths , words per minute, characters per minute, proportion of skipped words, proportion of skipped characters, and probability of regression for readers in different grade are summarised. As can be seen from the table, there is a steady decline in the mean and standard deviations of FFD, RFD and RRD as readers progress from grade 4 to grade 5. There is also an increase in saccade length, words per minute, characters per minute, and proportion of skipped words and characters, which suggest that grade 5 readers made longer saccades, read faster and skipped more words and characters. Generally, Table 3.3 shows children clearly benefiting from the year of teaching in terms of their ability to process more rapidly the words they fixate, make longer saccades, read faster and skip more words and characters, which is consistent with previous research (Buswell, 1922; Gilbert, 1959a, 1959b; McConkie et al., 1991; Rayner, 1985). Although the probability of regression slightly increases when readers progress from grade 4 to grade 5, the increase is not significant ($F(2, 87) = 1$, $p > .05$). This in agreement with the finding of Feng et al. (2009) an approximately similar rate of around 20% for 3rd graders, 5th graders and college students.

Table 3.3: Approximate mean (and standard deviation) of FFD, RFD, RRD, saccade length, words per minute, characters per minute, proportion of skipped words, proportion of skipped characters and probability of regression for readers in grades 4 and 5

	Grade 4	**Grade 5**
FFD(ms)	254(135)	241 (113)
RFD (ms)	341 (239)	320 (232)
RRD (ms)	450 (353)	438 (349)
Saccade length (character)	2.4 (1.6)	2.7 (1.7)
Words per minute	181 (82)	213 (102)
Characters per minute	304 (138)	362 (173)
Proportion of skipped words	0.33 (0.12)	0.38 (0.12)
Proportion of skipped characters	0.60 (0.07)	0.64 (0.07)
Probability of regression	0.28 (0.07)	0.32 (0.06)

Figure 3.1 below, shows the decomposition of viewing time as a function of reading ability group and grade. We can see that the readers classified as poor and average show a decline in FFD, RFD, and RRD across grade, while those classified as good show a slight increase in re-reading times.

Figure 3.2 is a scatter plot of the individual saccades of readers overlaid with the mean duration for each group-by-grade combination. We can see that each group

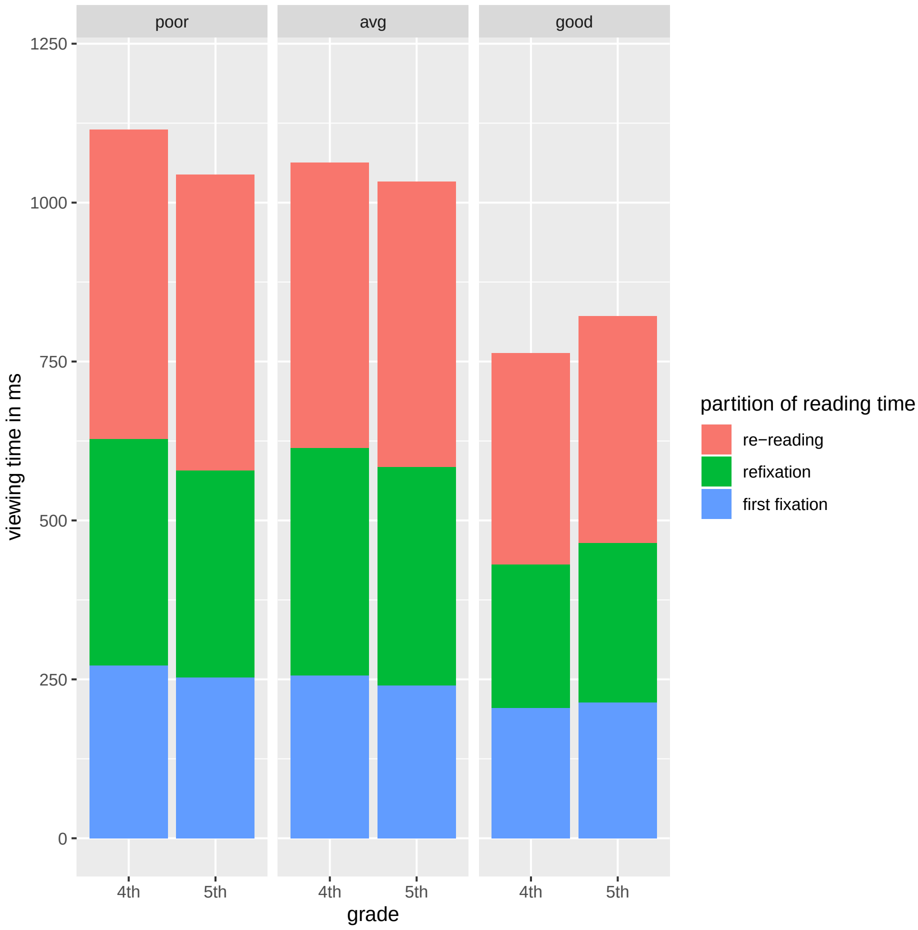


Figure 3.1: Decomposition of viewing times as a function of grade and reading ability

of readers made longer saccades as they progressed from grade 4 to 5, but average readers and poorer readers maintained a very similar mean saccade length.

Figure 3.3 shows the mean of reading speed at character and word level as a function of grade and reading ability. We can see that readers' reading speeds, when converted into words per minute or characters per minute, were increasing for poor, average, and good readers, respectively. Poor and average readers read faster as they progressed from grade 4 to 5 while good readers read at similar speed from grade 4 to 5.

Similar to Figure 3.3, Figure 3.4 shows the mean of skipping rate at character and word level with group-by-grade combination. There is a slight increase in word and

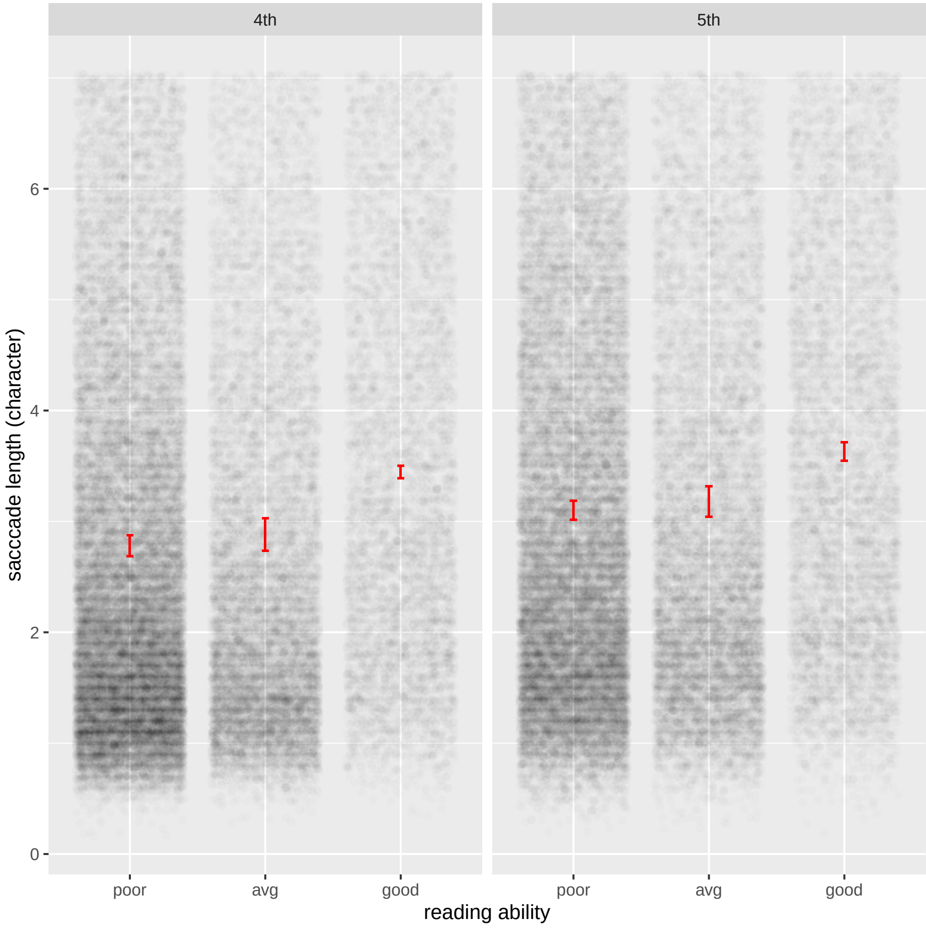


Figure 3.2: Saccade length as a function of grade and reading ability

character skipping rate from grade 4 to 5.

Figure 3.5 is the average probability of regression for each group-by-grade combination. We can see a steady increase from grade 4 to 5 for each group of readers. Poor and average readers have very similar pattern of probability of regression while good readers regressed more. It may reflect a change from a dominance of decoding-related effort towards allocating more cognitive resources to comprehension in the case of good readers.

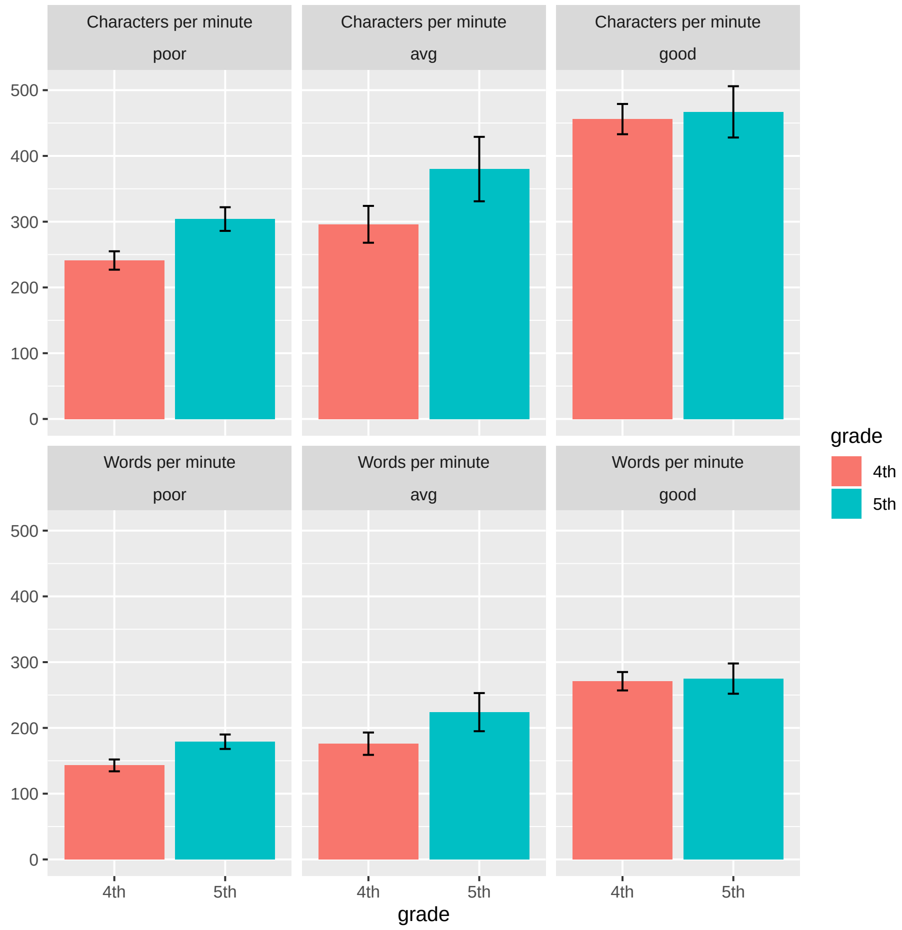


Figure 3.3: Reading speed as a function of grade and reading ability

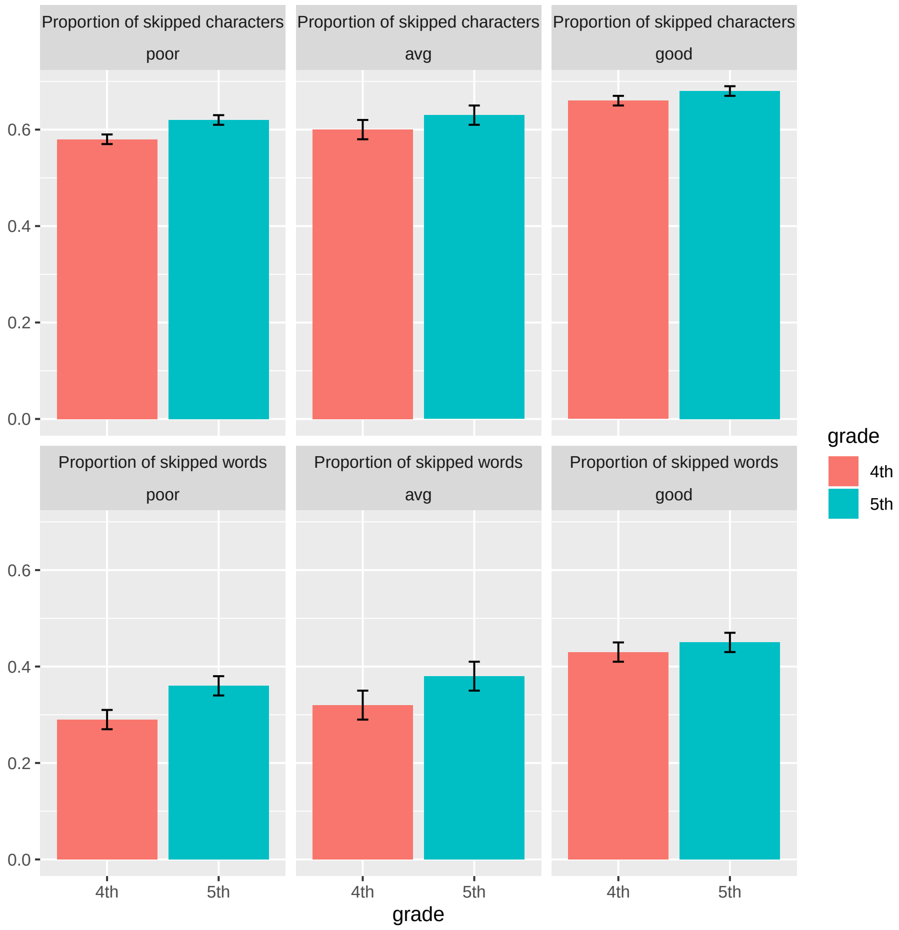


Figure 3.4: Skipping rate as a function of grade and reading ability

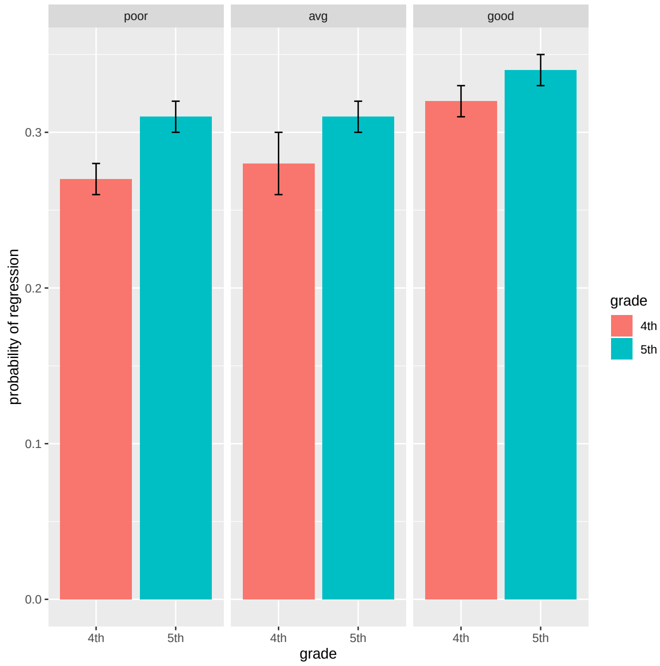


Figure 3.5: Probability of regression as a function of grade and reading ability

3.5 Conclusion

The preliminary analysis of the data corpus yielded few surprises and confirmed the alignment of the pattern of readers' learning and developmental progression with those of readers of other writing systems. This, in turn, confirmed the validity of using the key viewing time metrics of first fixation duration (FFD), refixation duration (RFD), and re-reading duration (RRD).

Overall, the corpus of data is a valuable research resource in itself. In addition, a range of complementary IQ and language ability measures were collect, which provide valuable context for further analyses of the eye movement data.

Chapter 4

Eye Movement Control in Reading Chinese: A Matter of Strength of Character

4.1 Background

While we have a reasonable understanding of the principles involved in saccade targeting for spaced alphabetic writing systems, such as English, how these generalise to an unspaced, non-alphabetic system such as Chinese is still the subject of debate. Most theories of eye movement control during the reading of alphabetic scripts assume that the word is the saccade target. The existence of a preferred viewing position (PVL) has been used to argue that saccades are sent to words as the default target. While the word centre is considered as the optimal viewing position (OVP) based on evidence from isolated word recognition (O'Regan, 1992; O'Regan & Lévy-Schoen, 1987; O'Regan et al., 1984; Vitu et al., 1990), saccade landing positions within words during continuous reading usually show a pronounced peak somewhere between the beginning and the centre of the word. This peak is referred to as the preferred viewing location or PVL (McConkie et al., 1988; McConkie et al., 1989; Rayner, 1979). In the case of alphabetic writing systems, the eye's targeting mechanism appears to make use of boundary information provided by inter-word spaces. However, in the case of an unspaced writing system such as Chinese, whether the saccade target selection is word-based or character-based and whether there is a PVL are still open questions (Li et al., 2011, 2015; Li et al., 2009; Yan et al., 2010; Yang & McConkie, 1999). Compared to most alphabetic writing systems the most noticeable property of written Chinese is that texts are formed by monospaced, square-shaped characters, where no spaces or other notation is used to indicate word boundaries. A character is the basic perceptually prominent unit of Chinese and represents a single syllable, the fundamental unit of meaning. Individual characters can be combined together to form words. Twelve percent of Chinese words comprise single characters, 74% two characters, 8% three characters, and 6% four or more characters (Modern Chinese Frequency Dictionary, 1986).

Several studies have manipulated word boundary spacing in alphabetic writing systems to explore the role they play in fluent reading. In a study by Rayner et al. (1998), the reading rate for spaced writing systems decreased by approximately 50% when spacing was removed and the average fixation duration during the reading of normally spaced text increased to around 340 ms from 250 ms. Furthermore, the PVL of readers shifted towards the beginning of the word (Rayner et al., 1998; Res, 1996; Spragins et al., 1976). This suggests that readers of spaced writing systems may be able to identify word beginnings in unspaced text and adapt their oculomotor strategies accordingly. These studies demonstrate that while the absence of spaces for readers of spaced texts does not make reading impossible, it changes the PVL to the presumably less efficient word-initial position. On the basis of the shift in PVL, when readers of spaced writing systems read unspaced text it might be expected that readers of inherently unspaced scripts would also have PVLs near word beginnings. Based on the same reasoning, it is also possible that if spaces were inserted between words in Chinese text, readers of Chinese might perform better, as measured by shorter viewing times and a shift in PVL from word beginning towards word centre.

To investigate this possibility Bai et al. (2008) recorded native Chinese readers' eye movements as they read text under four conditions – normal unspaced text, text with spaces between words, text with spaces between characters that gave rise to non-words, and text with spaces between every character. They found that spaces between words neither hindered nor facilitated reading. Nonetheless, spacing between each character and spacing that gave rise to non-words caused readers to take longer to read, suggesting that it disrupted word processing.

In one of the first studies of saccade targetting in Chinese reading, Yang and McConkie (1999) found a more-or-less flat distribution of landing sites across the characters of a word. Tsai and McConkie (2003) found that both the properties of near-foveal words and character complexity influenced saccade targetting. It seems clear, therefore, that in reading Chinese, while words influence saccade decisions, they are not the sole source of influence nor even the most important source. More broadly, there is evidence from studies of Japanese reading that the more complex Kanji characters tended to attract a greater number of fixations than the visually simpler Hiragana characters (Kajii et al., 2001). These findings also echo earlier work of Vitu (1991) on the importance of visual gravity in modulating saccade targeting and saccade lengths during the reading of alphabetic scripts. In Chinese reading, there's an obvious hierarchy of interacting influences at work in guiding the eyes: at the lowest level we have the relative visual weight and complexity of the characters, which is augmented by the emerging influence over time of lexical affects through successful segmentation and word identification.

More recently, there's been a shift in research focus to the role of word segmentation in the eye guidance process. For example, Yan et al. (2010) found that landing site distributions appear to peak at the word centre when words receive single fixations and at the word beginning when there are multiple fixations. They designed a baseline, non-lexical simulation that assumed Chinese readers use a fixed-amplitude saccadic strategy. Because the resulting simulation did not account well for observed word-skipping probabilities nor for the dependency of fixation probabilities on word length and word frequency, they proposed that Chinese readers do not use a fixed-amplitude strategy but rather a word-based saccade targeting one. They proposed a dual strategy where Chinese readers: (1) target the centre of a word when they can determine the boundary of a word, and (2) target the beginning when they cannot.

However, Li et al. (2011) argued that the concentrations of landing sites at the beginning of a word may be an artifact of how the landing site distributions are calculated. When intra-word refixations were included, the distribution of fixations landing on the word was comparable across the characters of a word. They also found PVL curves similar to Yan et al.'s by using a simulation that assumed Chinese readers saccade a constant distance with some variance. Given that the simulation did not make any assumption that the eyes move to a specific position within a word they argued that Yan et al.'s findings do not necessarily support word-based saccade targeting and suggest that saccade targeting in reading Chinese might involve a combination of character-based and word-based targeting, contingent on word segmentation. In addition, Li et al. (2014) systematically characterized the ways in which the eye movement in reading Chinese is sensitive to word and character properties and suggest that reading is as reliant on words in Chinese as in other writing systems despite the very dissimilar script.

Wei et al. (2013) proposed a processing-based strategy for saccade target selection in which readers try to identify as many characters as possible to the right of the fixation, and then move their eyes beyond the identified characters. As a consequence, the easier the processing of the fixated region, the longer the outgoing saccade. Moreover, a more controlled empirical study involving a boundary paradigm Li et al. (2015) supported the account of a processing-based strategy with evidence that when more information about a word is obtained in parafoveal vision the longer the subsequent outgoing saccade. There are also some recent results that support a dynamic adjustment strategy (or processing-based strategy) where a character tends to be skipped if it is processed in parafoveal vision (Li et al., 2015; Liu et al., 2017; Liu & Reichle, n.d.; Liu et al., 2016; Liu et al., 2019; Liu et al., n.d.).

To sum up, there is a number of different theoretical accounts of how readers of Chinese select saccade targets. Some researchers have failed to find a PVL in Chinese reading, suggesting that saccade targets are possibly character-based (Tsai & McConkie, 2003; Yang & McConkie, 1999). However, Yan et al. (2010) have argued that saccade targeting is based on ongoing word segmentation, where readers move their eyes to a PVL if segmentation is successful, but move their eyes to the word beginning if not. Other researchers have proposed that Chinese readers do not move their eyes to specific locations but that the properties of the fixated word instead affect the outgoing saccade length, with longer saccades launched from words that are easier to process (Li et al., 2015; Liu et al., 2017; Liu & Reichle, n.d.; Liu et al., 2016; Liu et al., 2019; Liu et al., n.d.; Wei et al., 2013).

To further explore this question, this chapter contributes evidence from the corpus-based study of Chinese reading described in the previous chapter.

4.2 Results

The direction and length of saccades are usually assumed to reflect cognitive and perceptual processes (Inhoff & Radach, 1998; Rayner & Pollatsek, 1981). One way of exploring the impact of lexical information on saccade targetting is to systematically partition the landing site data on the basis of launch distance. If lexical factors come into play, It is expected to see their impact on more proximal launches, since the word is more likely to fall within the higher resolution region of the reader's perceptual span. This type of analysis is illustrated below in Figure 4.1. The graph shows empirical landing site distributions as a function of launch distance for two-character words. Note that landing positions are measured in tenths of a character. As the site of saccade launches approaches a word, we see what appears to be more accurate targeting of word centres; the landing site distributions on the target word gradually become more peaked with the centre moving towards the word centre. Another feature of note is the fairly discrete transition in distribution shape for launch distances closer than -0.8 (the launch distance measure assumes the start of the target word to be zero), which suggests that saccade targetting may be coming under the control of lexical factors within the one-character launch distance range.

What mechanism might underlie longer launches that are not influenced by lexical factors? One way of exploring this question is to use a random saccade model as a baseline mechanism, which effectively progresses the eye by the same average saccade length but where saccades are generated by randomly permuting the original saccades so that the sequence of word lengths is preserved (Kliegl, 1981; McDonald et al., 2005; Yan et al., 2010). This randomisation alters the assignment of fixations to words in a sentence and since the model uses the original empirical data, inheriting its distributional characteristics, it should only disrupt the impact of lexical factors. In Figure 4.2, below, the landing site distributions for a permuted version of the empirical data are plotted in figure 4.1. There is, surprisingly, a high degree of similarity between the permuted and actual data, even for near launches. The major difference is that the peak of the fitted curves for near launches (top-left panel in figure 4.1 and figure 4.2) is significantly higher in the actual data compared to the permuted data ($z=-2.67$; $p<0.01$), where differences in the regression estimates for the quadratic term were tested. Launches from further to the left do not show a similar pattern.

To provide an overview of the two previous figures Figure 4.3, below, plots the non-zero peak locations of the distribution for near launch distances for both the

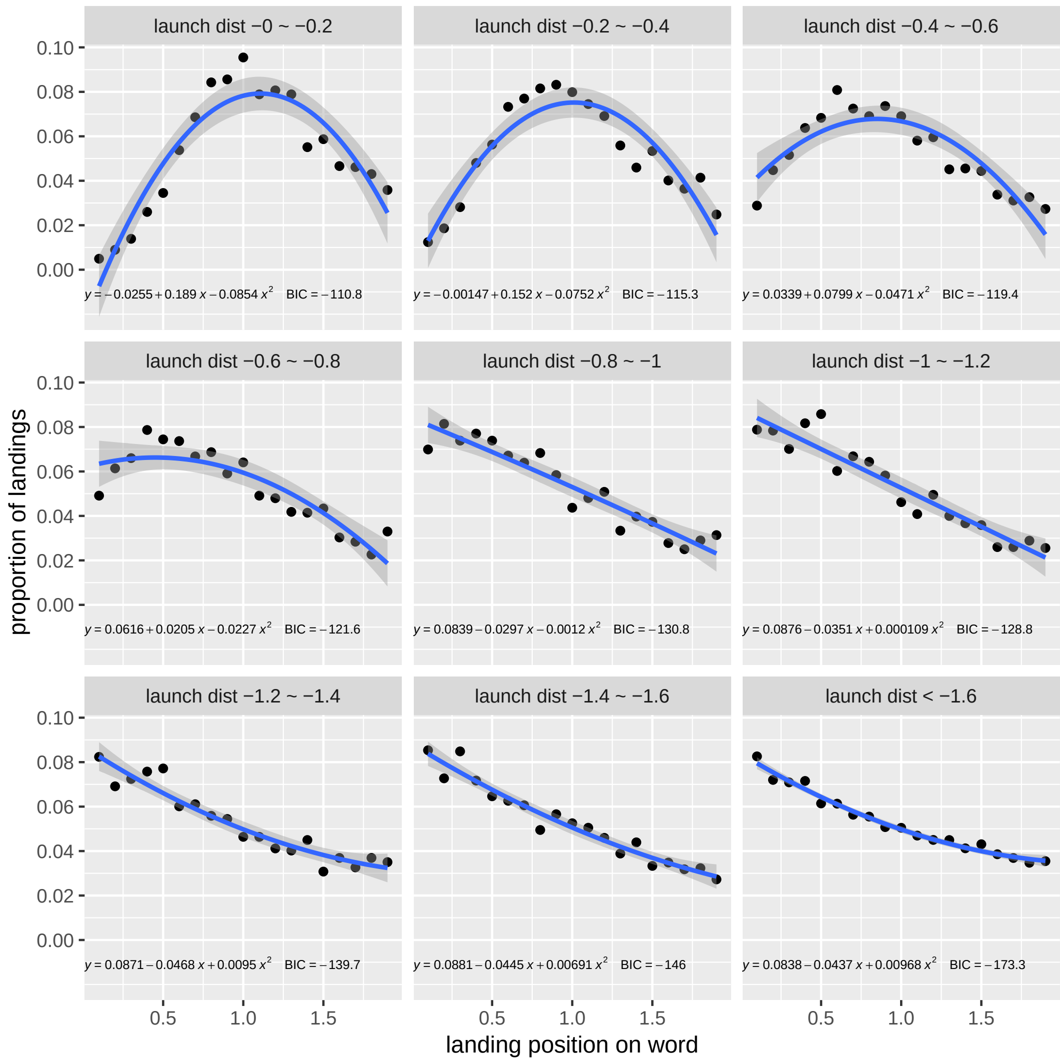


Figure 4.1: Landing site distributions on two-character words as a function of launch distance measured from the beginning of the word

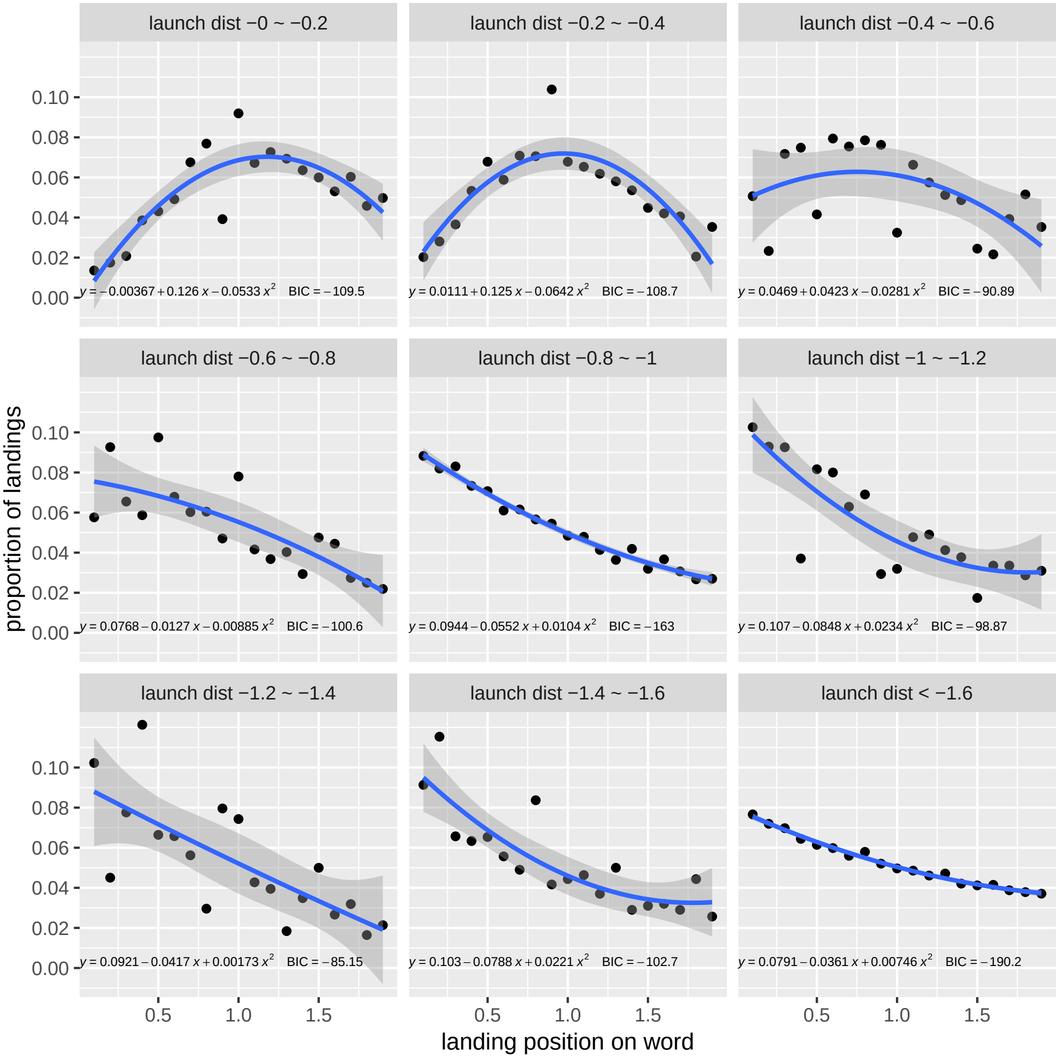


Figure 4.2: Landing site distributions as a function of launch distance using randomly permuted empirical saccades

actual landing site data and those involving permuted saccades. While the transition to non-zero peaks is delayed in the permuted data, the slopes of the rightward shift of distribution peaks are not significantly different from each other.

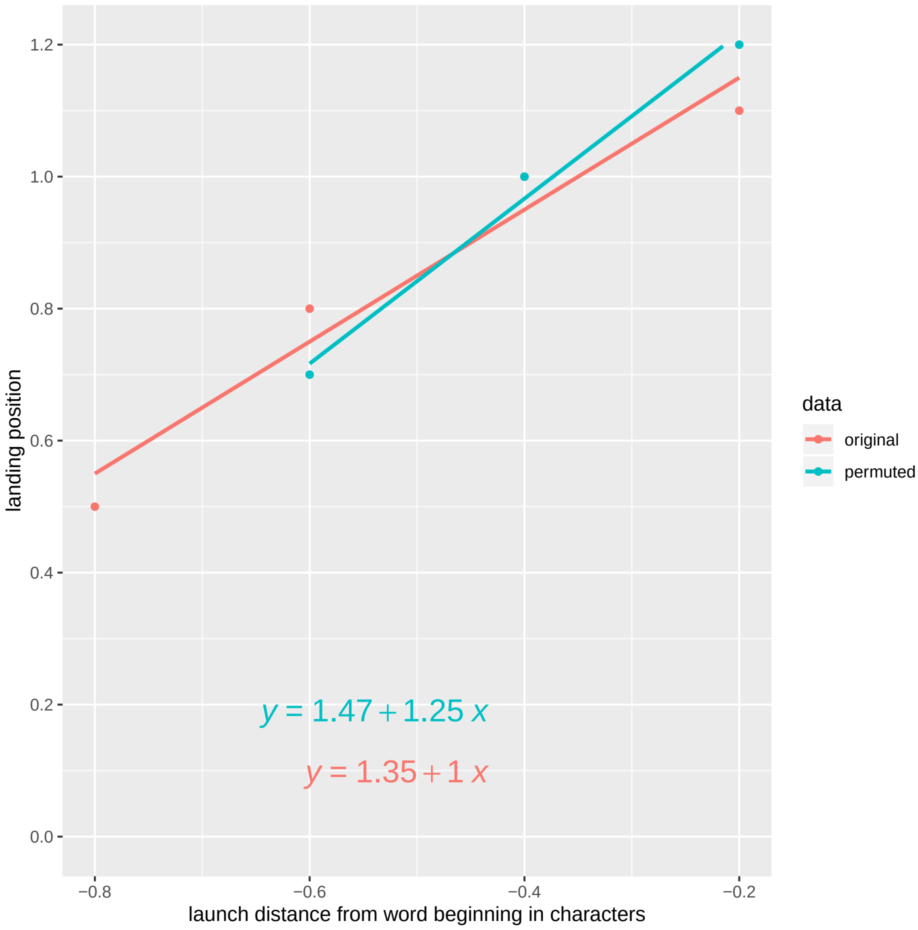


Figure 4.3: Peak of landing site distributions on words of length two as a function of launch distance with respect to word beginning

Taken overall, the preceding analysis suggests that lexical factors play a role in saccade targetting only when the launch site is close to the target word. This is despite the empirical data appearing to show a robust word-centre targetting tendency.

As well as lexical factors influencing incoming saccades, albeit in a limited way, there is evidence that the lexical properties of the fixated word (e.g., word frequency) can also influence the length of the outgoing saccade. Liu et al. (2016) found a relationship between word frequency and outgoing saccade length, where the higher the frequency, the longer the saccade. Therefore, saccade length data as a function of the frequency of the fixated word was examined. In addition, the duration of the launch fixation was looked at and these combined data are graphed in Figure 4.4.

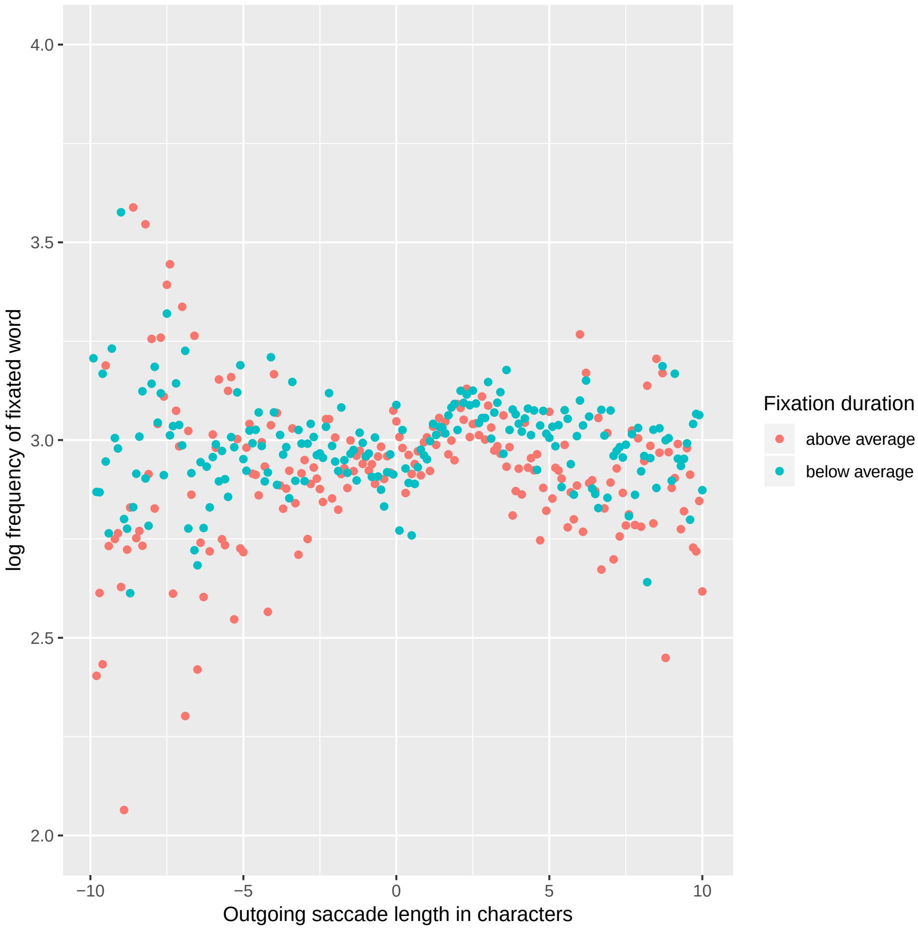


Figure 4.4: Length of leftward and rightward outgoing saccades for words of length two, as a function of fixated word frequency and duration

The pattern of saccade lengths is quite striking. We can see a clear linear effect of log word frequency on saccade lengths for short-range, rightward saccades within a narrow range of two to three characters. It appears that the higher the frequency of the fixated word, the longer the saccade within this range. However, this pattern reverses for longer saccades, showing an inverse and more variable relationship between saccade length and fixated word frequency. The pattern of fixation duration across the range of saccade lengths is, as one would expect, a function of word frequency;

higher frequency words give rise to shorter fixations.

The pattern in Figure 4.4 suggests there might be two distinct mechanisms at work in saccade control: (a) a high-level, lexically driven one, most likely related to the short-range targetting of words, and (b) a lower level, perhaps default, mechanism where other factors come into play. What this low-level mechanism might be is unclear, but a preliminary proposal will be made below.

As discussed earlier, Liu et al. (2016) have proposed a dynamic adjustment model of saccadic control where saccade extent varies dynamically as a function of the availability of lexical information. However, the work of Vitu (1991) and Kajii et al. (2001) cited earlier suggest another possible mechanism: a visual centre-of-gravity effect, where the most visually salient object in the right parafovea is the most attractive target for a saccade.

If we use the stroke count of characters as a proxy for visual gravity, we should expect to see these characters or their neighbourhood being fixated preferentially. Indeed, the data in Figure 4.5 support this view, suggesting that the default saccade targetting mechanism in Chinese is to target regions containing the most visually complex character in the right parafovea. Figure 4.5 shows the average number of strokes of the fixated character following saccades of different length. The stroke number peaks for the fixated character or the one immediately following.

Further evidence for a visual gravity effect in saccade targeting can be seen in Figure 4.6, below. It illustrates fixation locations for sets of four-character combinations of varying character complexity. Note that these are not words, but character sequences that can cross word boundaries. Six different sequence types: SSSC, SSCC, SSCA, SCAA, CAAA, and CCAA were analysed, where "S" (simple) is a character comprising five or less strokes, and "C" (complex) a character comprising 10 or more strokes. The graphs in Figure 4.6 (a) and (b) clearly show a landing site preference for characters with more strokes. When intra-sequence saccades are also included, the

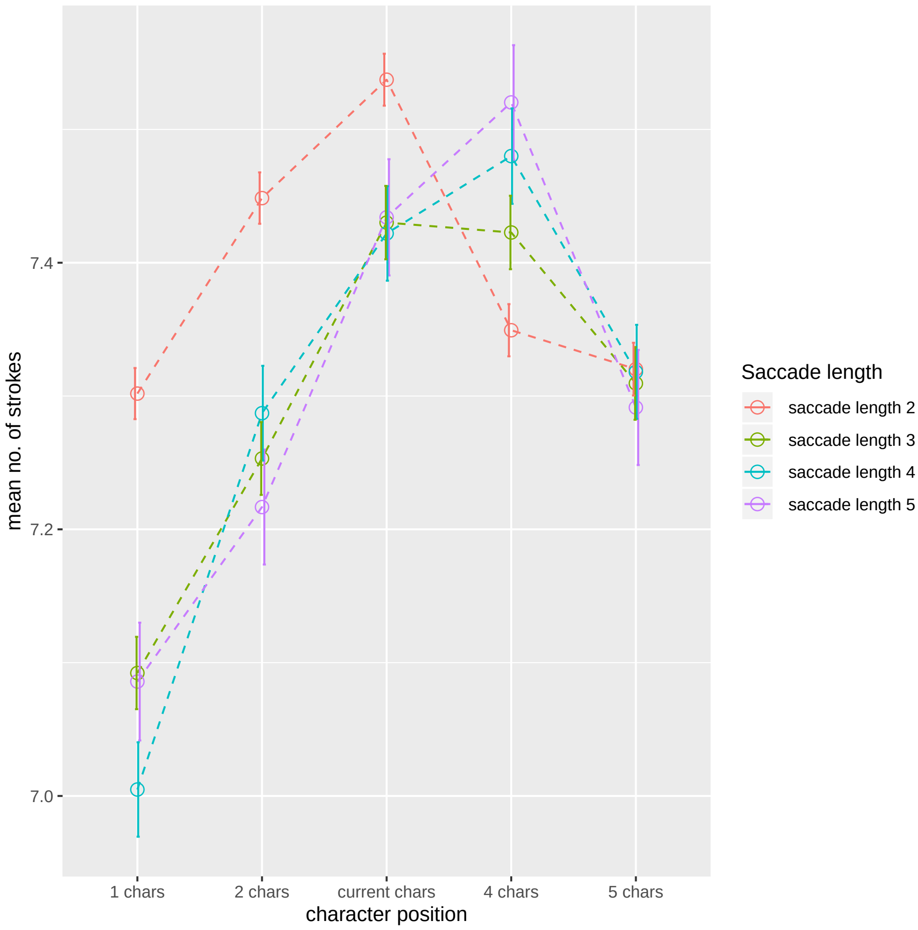


Figure 4.5: Mean number of strokes for the currently fixated character and the two preceding and two following characters as a function of saccade length

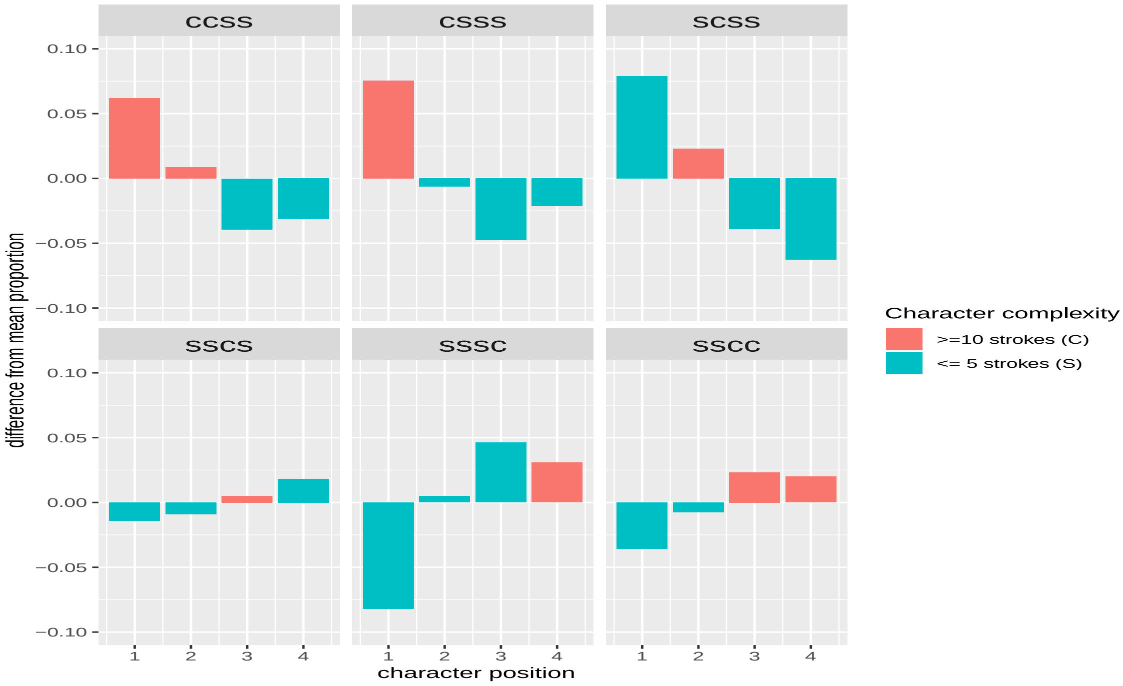
preference for the most complex character dominates. Note that the graphs represent deviations from average or expected fixation counts.

Figure 4.7 also shows a similar tendency, but this time in the context of actual two-character words. Similarly to the previous example, two–character words were classified into four sequence types based on the stroke density of their characters – simple versus complex. We can see in Figure 4.7 that in all cases fixations on the more complex character of the character pair exceeded the mean proportion of fixations for that character position.

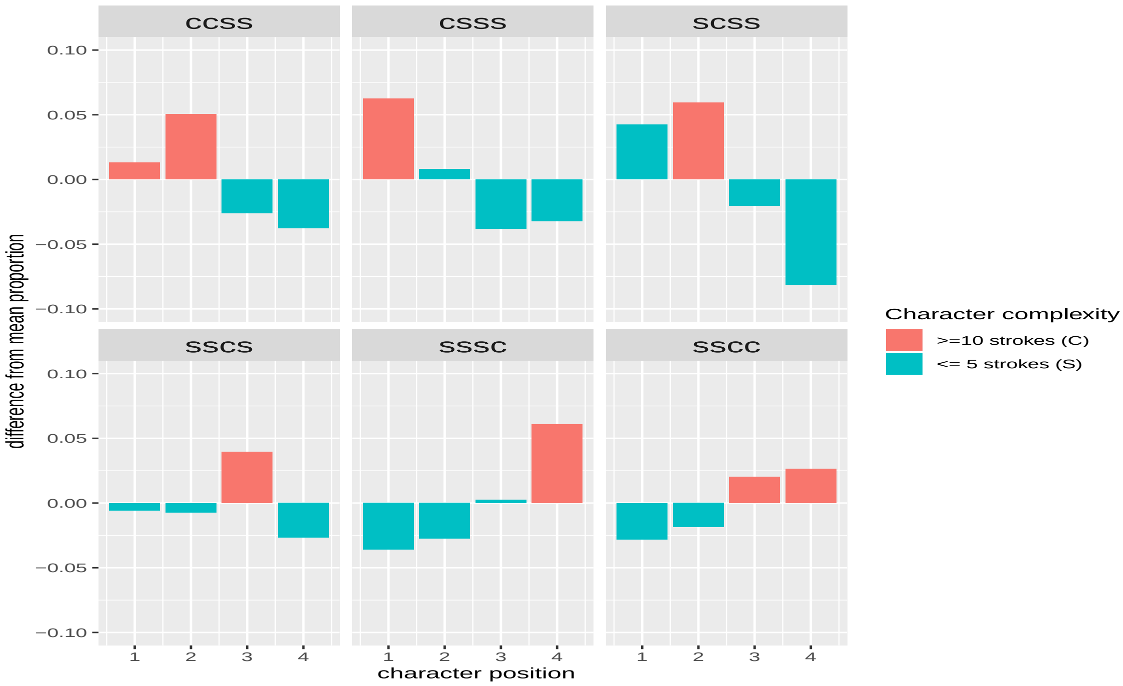
One of the more recent theories of eye movement control in Chinese reading argues for a dynamic account in which the ongoing processing of the text determines saccadic extent (Liu et al., 2015). Liu et al., 2015 provided evidence that saccade length varied as a combined function of foveal load and the availability of parafoveal information. Using a boundary paradigm experiment where foveal load, as measured by the frequency of the fixated word, and the availability of preview information combined independently to affect saccade length (Figure 4.8).

It is possible to create a continuous reading analogue of the boundary paradigm if we assume that the availability of parafoveal information is correlated with the visual complexity of the characters as measured by their stroke count. Since characters with high stroke density are harder to recognise in the parafovea than characters with fewer strokes (Ma & Li, 2015; Wang et al., 2018), we can consider their presence as analogous to an invalid preview, whereas low stroke density is analogous to a valid preview. In Figure 4.9 below, a similar pattern was found as in Liu et al's boundary study: frequency and complexity combine independently to affect saccade length.

This suggests that a dynamic account of saccade targetting is not incompatible with the idea that saccade targetting depends both on the ongoing processing at the launch fixation and the saliency of character targets in the right parafovea.



(a)



(b)

Figure 4.6: Deviation from mean of proportions of fixations on four-character sequences as a function of character complexity patterns: (a) where the sequence was treated as a pseudo-word and only incoming rightward saccades are included, and (b) where intra-sequence saccades are also included

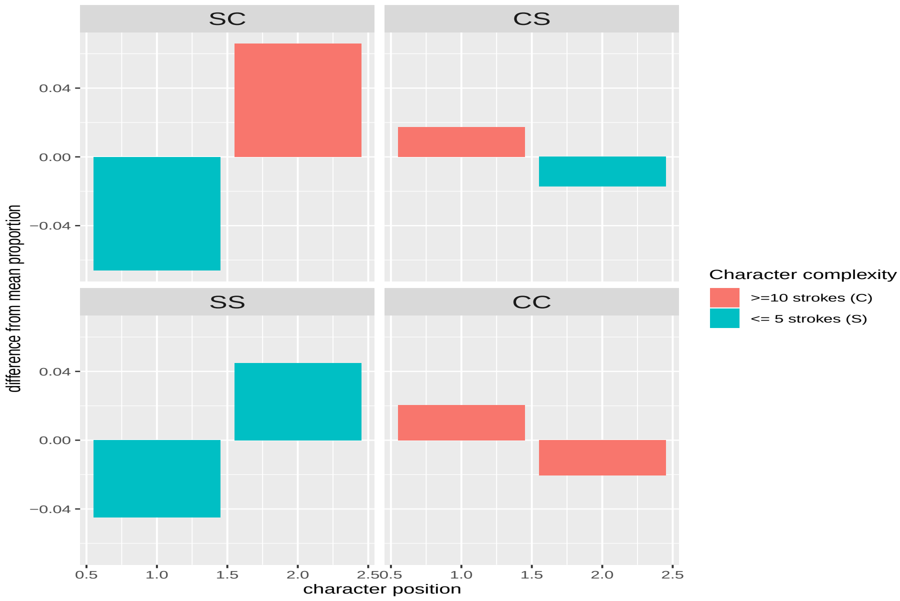


Figure 4.7: Deviation from mean of proportions of first fixations on characters of varying complexity in words of length 2 and incoming rightward saccades of length less than or equal to 7 are included

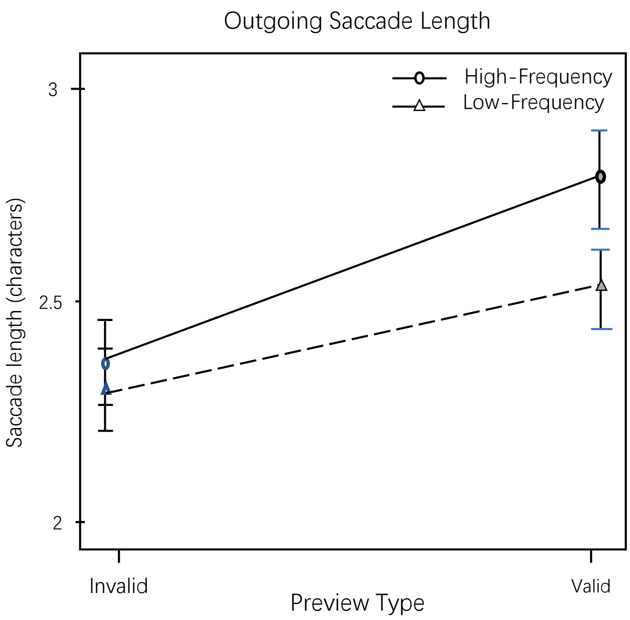


Figure 4.8: Outgoing saccade length as a function of fixated word frequency and preview availability

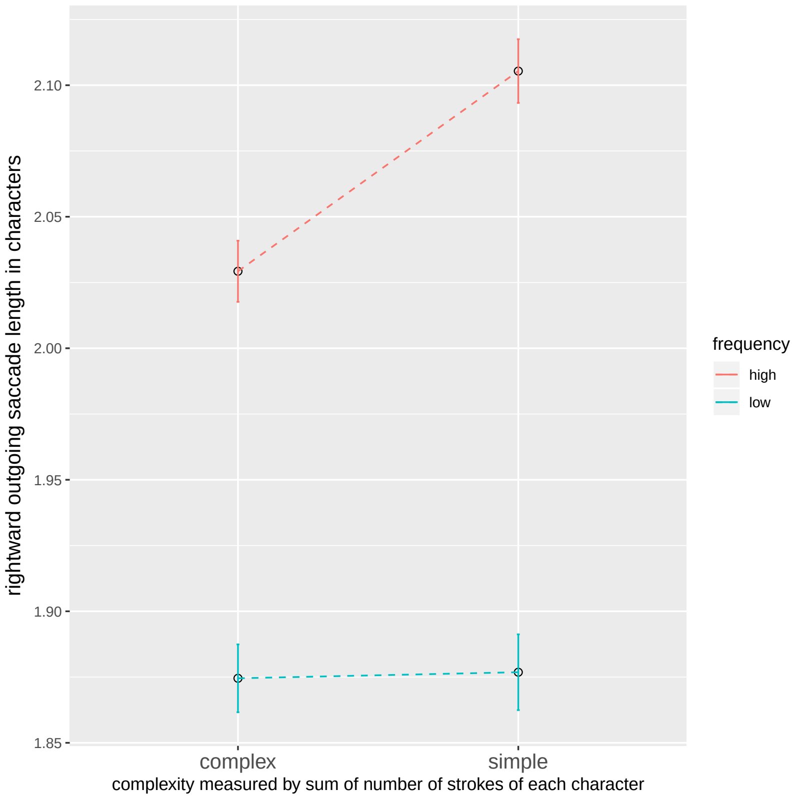


Figure 4.9: Outgoing saccade length as a function of fixated word frequency and preview availability of the following three characters measured by their sum of stroke count

4.3 Conclusion

To summarise, this chapter has explored eye movement control in Chinese reading in a population of young readers. An effort was made to uncover the mechanisms involved in determining where the eye goes next following a fixation. While a number of studies have explored the role of word targetting in Chinese reading, the evidence has always been somewhat equivocal about the pre-eminence of lexical factors. On the other hand, few convincing alternatives have been proposed for a dominant or default eye movement control mechanism. The above analyses show two complementary sets of evidence for saccadic control in Chinese reading: (1) a default, long range mechanism that progresses the eye left-to-right, targetting successive visually complex characters, and (2) a lexically driven short–range mechanism that targets word centres. The challenge is to provide an account of the saccadic process that allows these factors to interact.

Chapter 5

Measuring Chinese Character Discriminability Using a Saccade Latency Metric

5.1 Background

It was proposed in the previous chapter that character complexity played a significant role in eye guidance during reading of Chinese. Therefore, understanding the processes underlying the perception of and discrimination between Chinese characters would provide an important foundation for our understanding of Chinese reading. As mentioned previously, a character, the basic graphemic unit of Chinese, is monosyllabic and represents a basic unit of meaning. There are 3,500 most frequently used characters in mainland China (Ministry of Education of the People's Republic of China, 1988). A character consists of strokes, which in turn can form radicals. Many characters are similar to each other. Given the visual complexity of characters, a discrimination matrix describing the ease, or otherwise, with which a given character can be distinguished from every other of the 3,500 most frequently used characters would be a useful research resource. Moreover, generating a letter discrimination matrix has been the focus of research in alphabetic writing system to facilitate studies on letter recognition, visual research, and reading (Courrieu et al., 2004; Jacobs et al., 1989), but there is nothing simiar for logographic writing system such as Chinese. This is for the fairly obvious reason that gathering data for one-to-all comparisons of Chinese characters would be time consuming and unrealistic.

A limited amount of research has focused on the similarity ratings of the most frequently used Chinese characters. Song et al., 2008 defined the structures of the 6,763 characters in the GB2312 character set according to their graphemic features rather than other factors such as etymology or stroke orders. Based on their description of character structures, the authors proposed an algorithm to calculate a similarity measure between all the 6,763 characters. Ranked by the scores, the top 100 most similar characters were chosen for every one of the 6,763 characters. For example, these characters ‘霭’, ‘蕴’, ‘落’, ‘藩’, ‘莎’ , ‘薄’, ‘藻’, ‘蒲’ are from the list of similar characters to the character ‘蔼’. They are displayed according to their similarity score to ‘蔼’ in descending order. The output of the algorithm was claimed to have accorded with human visual perception. However, this claim is hard to assess since the authors did not provide any empirical evidence. For example, is ‘落’ really more similar to ‘蔼’ than ‘蒲’?

Wang and Xiong (2013) developed an algorithm to calculate character similarities of any two characters. The calculation was based on breaking down characters into different structural levels of and component radicals. To validate the algorithm, the authors chose 100 groups of three similar characters to be use in an experiment. They first let the algorithm pick the two most similar characters in each group. Then they had a group of 20 people subjectively rate the most similar characters within each group. Although the algorithm results performed 98% in accordance with the subjective similarity rating result, it only focused on a small subset of the total number of characters in common usage.

In the present chapter, a new, quasi-empirical discriminability matrix for 3,500 of the most frequently used Chinese characters based on a saccade latency paradigm is introduced (Hill, Radach, & Reilly, 2009). The assumption of the experiment is that the time taken to trigger an eye movement from a cue character to a matching peripheral target character would be negatively correlated with the similarity of the target with a peripheral distractor.

5.2 Data collection

5.2.1 Participants

Twelve adult native speakers of Chinese (7 female, 5 male) participated the experiment. They reported normal or corrected-to-normal vision. Participants were divided into 6 groups. Only participants of the same group completed the same number of experiments. Each group completed 18 * 2, 15 * 2, 10 * 2, 4 * 2, 2 * 2, 1 * 2 experiments respectively.

5.2.2 Materials

Stimuli were the 3,500 most frequently used Chinese characters (Ministry of Education of the People's Republic of China,1988). A character could be either a target or a distractor. All possible combination of these characters would give 3500 x 3499 pairs. Because it was not feasible to ask a single participant to perform the millions of comparisons needed, the 3,500 Chinese characters were randomly divided into 100 groups of 35 characters (see Figure 5.1). Hence, each experiment only consisted of 1,190 trails of the 35 * 34 possible comparisons. Based on the subset of comparisons collected in the experiment a computational model was trained to complete the remaining comparisons for each experiment.

OpenSesame was used to run the experiment (Mathôt et al., 2012). Eye movements were recorded with an EyeLink 1000 system (sampling at 1000 Hz). Characters were presented in the centre of a 19.5-in. Huike E2006 monitor (resolution: 1,440 by 900 pixels; frame rate: 60 Hz). The Song 30px font was used. It took over 90 minutes for a participant to finish one experiment.

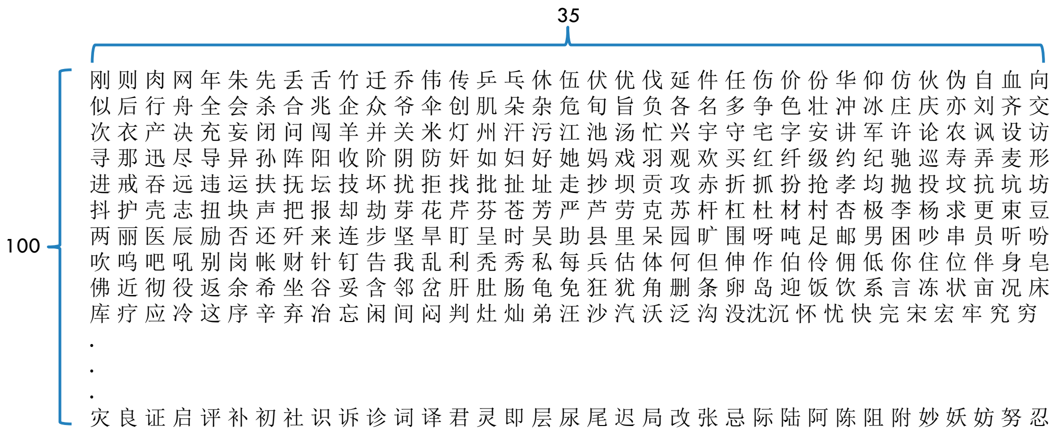


Figure 5.1: Characters used in the experiment

5.2.3 Procedure

Participants were seated 60 cm from the monitor with their head positioned on a chin-rest. Their eyes were in a line with the centre of the screen. Following calibration, participants triggered the first trial by looking at the middle hash of a mask (### # ###). The mask was then followed by a display of three characters in the following arrangement: ### 大 ###. One character was always different to the central character as a distractor on one side, while one was identical to the central character as a target on the other side. For one experiment, half of the targets were on the left-hand side and the other half on the right-hand side. Participants were instructed to gaze at the central character until they made a saccade toward the target. They were asked to do so as rapidly and as accurately as possible. The trial ended once the participant moved their gaze to either side of the central character above a certain tolerance (Figure 5.2). Soon afterwards, the mask was shown for the next trail. There were breaks between every 100 trials when the correct response rates were displayed on the screen. Every experiment started with 100 practice trials with breaks between every 10 trails. Participants were told, ideally, to rest during the breaks, but they were allowed to stop at any time between trials.

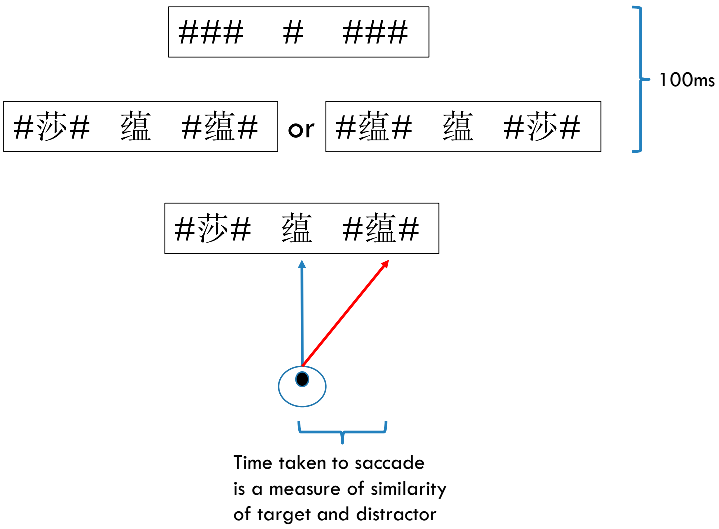


Figure 5.2: Saccade latency paradigm

5.3 Analysis

5.3.1 Date pre-processing

Before analysis, around 0.17 % of data were eliminated due to a bug in Opensesame which caused some characters to be rendered as numbers. Saccade durations of more than 1000ms or less than 140 ms, which represented around 14.8% of the data, was also removed from the analysis.

5.3.2 Result

In order to evaluate what factors affect the overall saccade latency, the saccade latencies for correct responses were analysed using a linear mixed effects model (LMM). Stroke difference, structural difference (e.g., integral vs. compound) and frequency difference between the target and the distractor were treated as fixed effects. Participant and target position were treated as random effects. Target position referred to either the left or right position where the target was presented.

As can be seen from Table 5.1, the results indicate that the greater the stroke difference, the significantly shorter the saccade latency ($t = -8.7$; $p < 0.001$). Figure 5.3 shows saccade latency as a function of stroke difference between the target and the distractor. There is also a slightly weaker effect when the overall character structure of the the target and distractor are the same: similar target and distractor structures increase latency ($t = 2.89$; $p < 0.01$). Interestingly, there was no effect due to any frequency differences between the target and distractor.

Table 5.1: Linear mixed model estimates. Note that strDiff = character structure difference, SDiff = character stroke difference, fDiff = character frequency difference

	Estimate	Std.Error	P
strDiff	0.007	0.003	0.003 **
SDiff	−0.004	0.0004	0.00 ***
fDiff	−0.0006	0.001	0.51
Intercept	5.7	0.3	0.00 ***
Note:	*p≤0.05; **p≤0.01; ***p≤0.001		

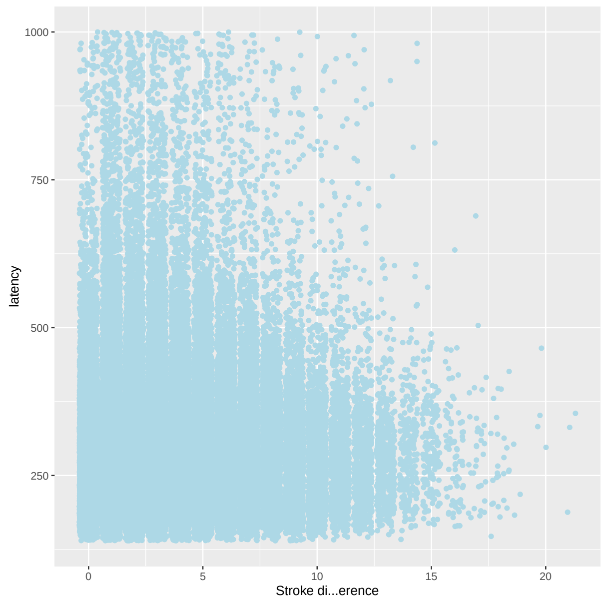


Figure 5.3: Saccade latency as a function of stroke difference between target and distractor

5.4 Model building

Neural networks are one of the main tools used currently in machine learning. As the "neural" part of their name suggests, they are brain-inspired systems which are inspired by the way the human brain learns. A neural network is a system of interconnected artificial neurons that exchange messages between each other. Although neural networks (also called "perceptrons") have been around for decades, they are enjoying something of a renaissance (Haykin, 2010). This is due to the availability of unprecedented amounts of data online, improvements in hardware performance particularly in the area of GPUs, and some algorithmic innovation. Neural networks are very effective tools at finding patterns in data. There are multiple types of neural network, each of which comes with their own specific use cases and levels of complexity. In the area of visual recognition, neural networks have made significant progress the last few years. Some neural networks can achieve reasonable performance on hard visual recognition tasks – matching or exceeding human performance in some domains (Simonyan & Zisserman, 2014).

Convolutional neural networks (CNNs) are widely used in pattern- and image-recognition problems as they have a number of advantages compared to other techniques. There are many CNN models. For example, VGG, GoogLeNet, ResNet are all in widespread use and available in so-called model zoos for use in new image-recognition applications. A CNN has two basic building blocks, i.e., the convolution block and the full connected block. The convolution block which performs feature extraction consists of convolution layers and pooling layers. The fully connected block is a fully connected simple neural network architecture. Based on the features produced by the convolution block, the fully connected layers learn how to use the features to correctly classify images. For instance, given an image, the convolution layer detects features such as two eyes, long ears, four legs, a short tail and so on. The fully connected layers then act as a classifier on top of these features and assign a probability for the input image being, say, a dog. A common technique with large-scale neural network applications is the use of transfer learning to take a network developed for one application (e.g., the ImageNet dataset) and partially retrain it on a related task (Yosinski et al., 2014).

The ImageNet project is a large image database designed for use in visual object recognition research. The ImageNet project runs an annual software contest, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), where software programs compete to correctly classify and detect objects and scenes. ResNet was the best performing network in the 2015 competition which achieved 3.57% error on the ImageNet test set(He et al., 2016). Resnet is a residual learning framework designed to ease the training of networks that are substantially deeper than those used previously. It explicitly reformulates the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions (He et al., 2016). The goal of this chapter is to partially retrain an instance of the ResNet network(He et al., 2016) to generate log latency values as output when presented with target-distractor pairs. The partial training involves adjusting the weights of the final, fully-connected layer of the network and freezing the weights in the preceding layers.

5.4.1 Training & testing

Figure 5.4, below is an example of the training input and output to the modified ResNet network used to generalise the experimental latency data. Note that the network was trained on pairs of distinct characters that were randomly offset on the vertical axis. The goal here is to encourage the network to create position-invariant representations of the characters. Upon presentation of the image, the network is trained to output the log of the saccade latency.

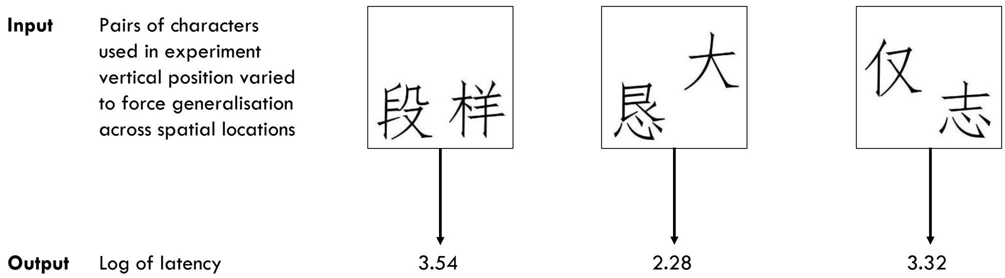


Figure 5.4: Model training

A sample of 60,000 character pairs were selected for training, with another 20,000 held back for testing. Preliminary results indicated that the use of root mean squared error (RMSE) as the loss function may not be the most effective approach since multiple training trials showed some degree of instability with the error values occasionally becoming extremely large. Further work on this model is clearly needed, but the general approach shows merit in allowing us to infer behavioural data for a large-scale experiment from a smaller subset of that experiment.

5.5 Conclusion

The results of this chapter go some way to achieving the aim of providing a useful tool for psycholinguistic research in Chinese reading. Understanding the parameters of the low-level visual processing of Chinese characters is an important component in building a broader understanding of Chinese reading. It is particularly important in gaining an appreciation of potential areas of difficulty for early readers. Nonetheless, there is considerable scope for improving the performance of the neural network. For example, training proved quite unstable using the root mean square error (RMSE) loss function. So other methods for calculating error and representing the desired output will need to explored. In addition, while training is limited to the final layer of weights, a more effective approach may be to fine-tune the network as a whole on the character analysis task. Once a reliable implementation of the neural network is completed, it will also be necessary to validate the generalisation performance of the network by carrying out an additional empirical study of readers.

Ultimately, the goal is to make the model available as an online resource for researchers who can input a set of characters and obtain a matrix of character discriminability measures.

Chapter 6

A Simple Computational Model of Saccadic Control for Chinese

6.1 Models of eye movements control in reading

The development of computational models of eye movement control in reading in recent years has furthered our understanding of the dynamic cognitive process in reading on the basis of empirical findings from eye movement experiments (Engbert et al., 2005; Feng, 2002; Just & Carpenter, 1980; Reichle et al., 2013; Reilly, 1993; Reilly & O'Regan, 1998; Reilly & Radach, 2006; Snell et al., 2018). Based on specific theoretical frameworks, these models produce precise predictions of basic eye movement characteristic such as fixation duration, fixation location and saccade length, which are similar to observed human eye movement patterns. The differences between these models thus can be portrayed according to their theoretical assumption about how perception, cognitive, and motor control processes work together during reading (see Rayner (2009), Reichle (2015) for a more extended discussion). There are three broad categories of models of eye movement control in reading: sequential attention shift (SAS) models, guidance by attentional gradients (GAG) models, and primary oculomotor control (POC) models (Engbert et al., 2002). SAS models assume that attention is shifted sequentially to one word at a time (Engbert et al., 2002; Reichle et al., 2013; Reichle et al., 1998; Reichle et al., 2003). These models were built on Morrison's (1984) theory that reader's attention is shifted word to word and each shift simultaneously initiates a saccade program to the same word. GAG models adopt the assumption that attention is distributed as a gradient and that multiple word are processed in parallel (Engbert et al., 2005; Reilly & Radach, 2006). POC models assume that low-level factors are primarily guiding eye movements during reading (McConkey, 1994; O'Regan, 1990; Reilly & O'Regan, 1998). Given these models cannot account for the influence of lexical-access effects as well as SAS and GAG models, they are not discussed further in the thesis.

6.1.1 E-Z reader

The E-Z Reader is one of the more cited SAS models (Reichle et al., 2013; Reichle et al., 1998; Reichle et al., 2003). E-Z Reader was initially based on Morrison's theory (1984) except that the model assumed separate signals for shifting attention to the next word and for a saccade to be made to that word. It has undergone several changes over the years in which more assumptions were added to make it more psychologically plausible. Nonetheless, the core assumption of sequential attention shift remained the same with one word processed at a time. A schematic diagram of E-Z Reader 7 is shown in Figure 6.1 (Reichle et al., 2003). E-Z Reader 7 comprises a visual system, a word identification system and an oculomotor system. Visual features of words are proceeds at a rate that is modulated by visual acuity limitations at an early stage of visual processing. In order to decouple the trigger for attention shift from that for saccade programming, the word identification system was divided into two processing stages, L1 and L2. The completion of L1 signals the oculomotor systems to begin programming an eye movement to the next word. The completion of L2 causes attention to shift to the next word. The oculomotor system also includes two processing stages, M1 and M2. M1 is the first, labile stage of saccade programming that could be cancelled by the initiation of a later saccade programme. M2 is the second, non-labile stage in which saccades could no longer be cancelled and will be executed when M2 is completed.

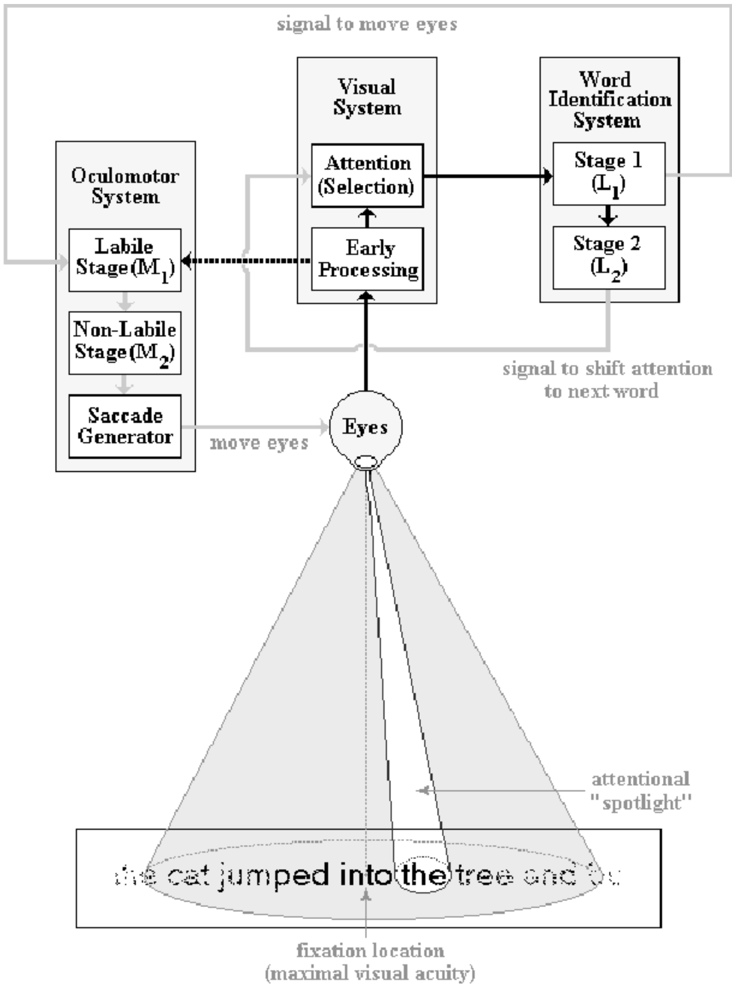


Figure 6.1: A schematic diagram of E-Z Reader 7 (Reichle et al., 2003)

6.1.2 SWIFT

The SWIFT (Saccade-generation With Inhibition by Foveal Targets) model is the most fully developed GAG model (Engbert et al., 2002; Engbert et al., 2005). It was developed based on the framework of saccade generation of Findlay and Walker (1999) and the dynamic field theory of movement preparation of Erlhagen and Schöner

(2002). As seen in the schematic diagram of SWIFT (Figure 6.2), the model comprises two sub-systems: one for saccade programming and one for lexical processing. These two systems are coupled via a foveally-inhibited random timing system and a saccade execution system which moves the eyes during saccades, thus the pathways of when and where the eyes move are separate (Engbert et al., 2002). Similar to E-Z Reader, both saccade programming and lexical processing are completed in two stages. However, SWIFT assumes that lexical information processing is spatially distributed over several words in parallel and modulated by a gradient which is a function of word length and fixation position relative to word center. Also, SWIFT assumes saccade generation is a stochastic process initiated by a random timing inhibited by foveal targets.

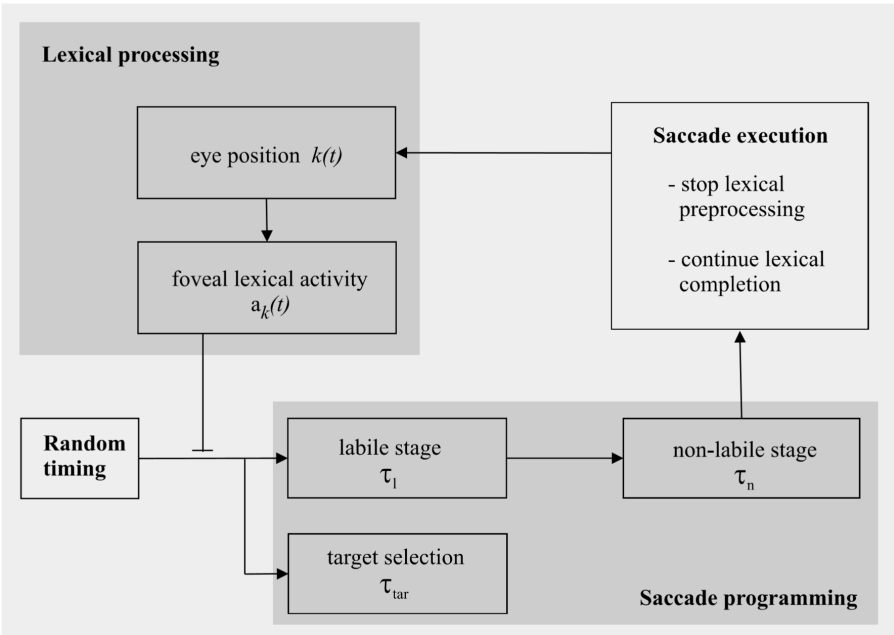


Figure 6.2: A schematic diagram of SWIFT (Engbert et al., 2002)

6.1.3 Glenmore

Glenmore is another GAG model of eye movement control in reading (Reilly & Radach, 2002; Reilly & Radach, 2006). Figure 6.3 is a schematic diagram of Glenmore. The main components of the model are a saliency map which selects the saccade targets, an input module which is a representation of the current perceptual span and codes the visual configuration around the fixation position, an interactive activation word-processing network that identifies words, a fixation centre that controls the decision when to execute a new saccade, and a saccade generator which initiates and executes eye movements. Similar to SWIFT, the model decouples the decision about when to move the eyes from the word recognition process. Glenmore successfully accounts within one mechanism for preview and spillover effects, regressions, progressions, and refixations.

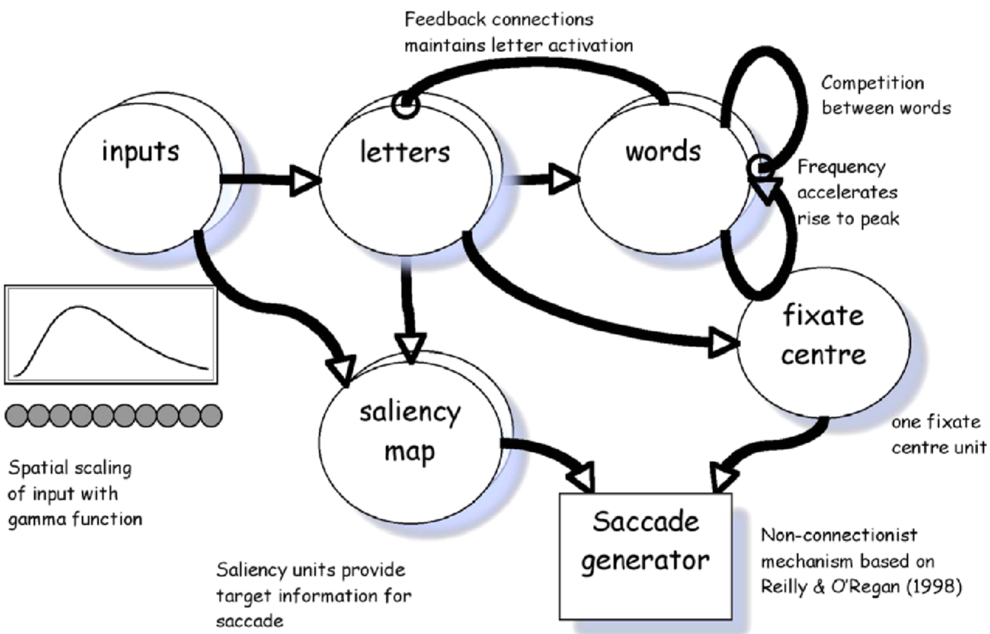


Figure 6.3: A schematic diagram of Glennmore (Reilly & Radach, 2002; Reilly & Radach, 2006)

6.1.4 Models of Chinese reading

There are a number of Chinese models of word recognition that have been developed to explain the coding process of the identifications and positions of characters (Hsiao & Shillcock, 2004; Perfetti & Liu, 2006; Perfetti et al., 2005; Taft & Zhu, 1997; Taft et al., 1999; Xing et al., 2002; Yang et al., 2006). These models have largely advanced our understanding of the orthographic processing of single word in reading Chinese. At the same time there is also a recent modelling effort that aims to deal with the segmentation of Chinese words (Inhoff & Wu, 2005; Li et al., 2011, 2015; Li et al., 2009; Yan et al., 2010). Li et al. (2009) argued that Chinese words segmentation and Chinese word recognition is an interactive process involving top-down and bottom-up factors and proposed a mathematical model to account for it (Figure 6.4). The model is a Chinese version of Interactive Activation (IA) model (McClelland & Rumelhart, 1981) which involves multiple levels: a visual feature level, a character level and a word level. The visual feature level abstracts visual features from the stimulus, the character level receives both perceptual information from the visual feature level and feedback information from the word recognition level. The character level allows up to four characters to be activated in parallel but with only one word being segmented and identified at any given time, which is broadly consistent with serial-attention eye-movement models. The word level receives information from both the character recognizer and the lexicon.

6.1.5 Extension of E-Z reader to Chinese reading

Although there are successful models of eye movements control in reading, most of them have been limited to dealing with the reading of Roman-derived writing systems, with little attention to their generality for non-Roman writing systems like Chinese. So far, only the E-Z Reader model has been extended to Chinese readers to account for aspects of the reading process (Rayner et al., 2007). Rayner et al. (2007) claimed that the Chinese E-Z reader model accounted successfully for both when and where Chinese readers move their eyes. They indicated that the successful simulation suggested that the control of eye movements in reading Chinese is similar to that in an alphabetic language such as English in that it is primarily lexically driven.

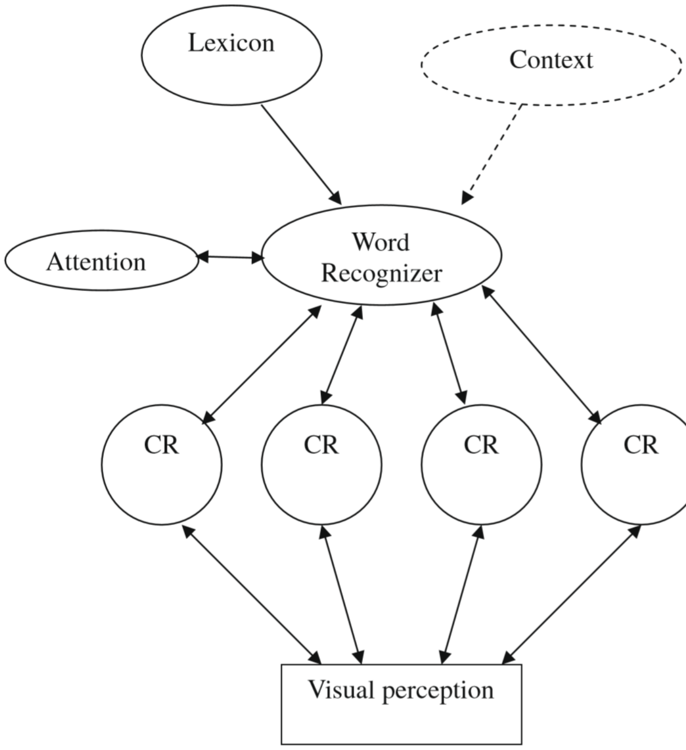


Figure 6.4: The framework of a word segmentation and recognition model (Li et al., 2009)

6.1.6 Machine learning models

There is a class of model that focuses on the use of machine learning approaches to predicting eye movements in perception (Harvey et al., 2019; Kornuta & Rocki, 2016). Notable among these are so-called deep-learning and convolutional neural networks (**CNNs**; Fukushima, 1980; Lecun et al., 2015). However, none has focussed on modelling reading. Moreover, there is some debate about whether this class of model is an appropriate framework for modelling natural perception (Cichy & Kaiser, 2019; Xu & Vaziri-Pashkam, 2020). In Chapter 5, a deep learning network was used to model the confusability for human readers of Chinese characters. However, modelling the reading process involves more than just character and word recognition. We also need to be able to model the temporal dynamics of perception and action, that is the time spent at a particular location in the text and the direction and extent of the subsequent saccade. This is exactly the focus of models such as E-Z Reader, Swift, and Glenmore (Engbert et al., 2002; Engbert et al., 2005; Reichle et al., 1998; Reichle et al., 2003; Reilly & Radach, 2006). While these models tend to assume a character or word perception component, it's not explicitly modelled. On the other hand, the focus of these models is on the interrelationship between perception and saccadic control in order to predict the key reading metrics of viewing time, saccadic direction, and saccadic extent. The resulting models are more comprehensive, but also are abstractions from the low-level details of character processing and saccadic programming. This is the level of focus of the model proposed in the next section.

6.2 Outline of a model

Attempting to provide a theoretical account of even low-level aspects of reading is a challenge. This is primarily because reading comprises a dynamically interacting coalition of processes that influence each other in often unpredictable and non-obvious ways. What Norris (2005), below, has observed about theories of word recognition can just as easily be applied to theories of reading:

> In research on word recognition, [computational] models don't just resolve debates over what theories predict, they are often the only way that even the theorists themselves can be sure what their theories predict. Norris (2005, p.333)

Glenmore is an interactive activation model of eye movement control in reading that can account within one mechanism for preview and spillover effects, regressions, progressions, and refixations (Reilly & Radach, 2006). A typical interactive activation (IA) model comprises a set of interconnected neuron like-units. Activity is transmitted through the network over weighted connections. The units comprising the network execute a function that combines the its inputs and generates a simple numerical output. A network of these units "computes" by circulating activation throughout the network until some stopping criterion has been reached, such as the achievement of a threshold, or the stabilising of overall levels of activity.

The Glenmore model decouples the decision about when to move the eyes from the word recognition process. When the overall level of activity in the recognition system exceeds a specific threshold, a saccade is triggered independently of whether or not a word has been recognised on that fixation. Another feature of the model is the use of a saliency map that acts as an arena for the interplay of bottom-up visual features of the text, and top-down lexical and textual features. These factors combine to create a pattern of spatial activation, the peak of which provides a saccade target.

More specifically, the units of the saliency map receive activation from both visual input and letter/character candidate units. Input from the character units represents cross-talk between the "what" and "where" processing pathways and provides a direct top-down "cognitive" contribution to the evolution of the saliency values for specific regions of the visual field. The spatial location with the highest activity at a certain point in the processing of a fixation acts as the target for the next saccade (Findlay & Walker, 1999).

The outline model in Figure 6.5 illustrates how the Glenmore model framework can be used to integrate the two saccade targeting modes which have been proposed in the previous chapters. Character complexity is a low-level feature of the visual array of characters and it plays a central role in the early stages of saccade targeting. The complexity of characters as measured by their number of strokes makes a direct contribution to the relative saliency of regions in the visual field. In the scenarios described here, just the right parafovea will be focused on, but the design principles of the model can equally account for regressive saccades. As characters and words are recognised, the level of word activation contributes to the saliency of a specific region in the parafovea. The activation of the component characters and the frequency of the word will drive up the the activation of the word. However, word activation ceases to contribute once it crosses a recognition threshold. Thus, activation of a region will persist for lower frequency words, since these reach threshold more slowly. Consequently, a low-frequency word is more likely to be refixated especially if it also comprises high-complexity characters, which themselves contribute to saliency.

A key advantage of IA models is that the common currency of activation allows for the interplay of diverse sources of information. Thus we can have activation from paragraph and lexical levels combining with low-level visual input to affect, say, word recognition. In the case of the account of Chinese reading discussed here, it is necessary to capture the interplay of character complexity and lexical properties on word

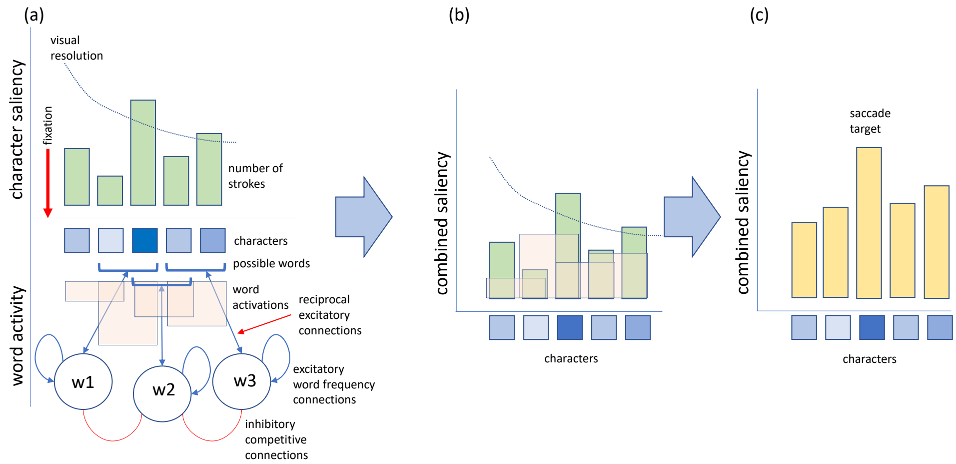


Figure 6.5: A schematic representation of the proposed model for accounting for saccade targeting in reading Chinese. Two interacting mechanisms were envisaged. One which targets characters with relatively larger numbers of strokes in the right parafovea, the other is responsible for within-word saccades (a). The latter consolidates recognition of words following successful segmentation by, where necessary, refixating the segmented word. Various, possibly overlapping, word candidates compete with each other for dominance based on word frequency and contextual factors (not shown here). The dominant word supports its component characters by increasing the level of spatial saliency for those characters (Reilly & Radach, 2006). We can integrate the two sources of influence on saccade targeting by using this saliency concept. So stroke density and word activity combine (b) to raise the level of spatial saliency in the visual field and a saccade is then launched to the location of highest saliency (c).

targetting. The Glenmore model achieves this through the neurologically supported construct of saliency (Findlay & Walker, 1999). In the Chinese instantiation of the Glenmore model, character complexity as measured by number of strokes and lexical frequency can additively combine to strengthen or diminish the saliency of a parafoveal region. Moreover, a region's saliency dynamically alters over time. We can see this illustrated in Figure 6.6.

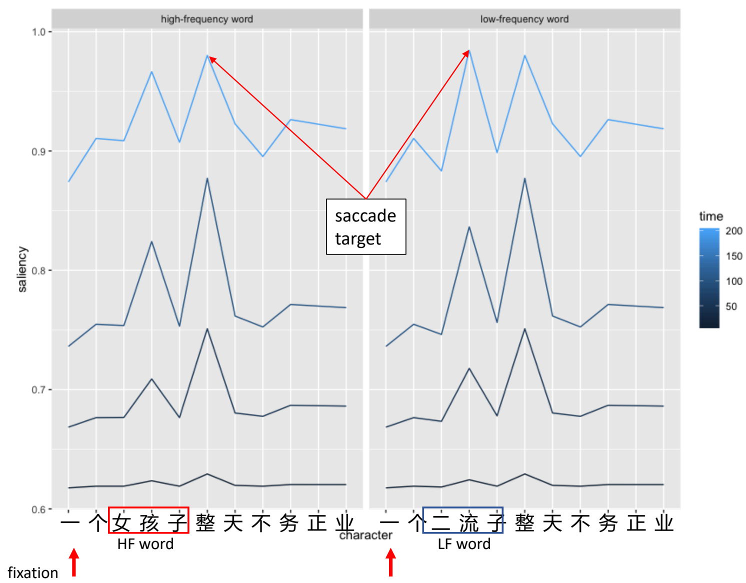


Figure 6.6: A sample run of the Glenmore model adapted for Chinese to illustrate the interplay of character complexity and lexical frequency. The texts in the two runs is almost identical, except for the frequency of the second word (in rectangles). In the left panel, the word is high frequency and low frequency in the right. In both panels we see the evolution over time of saliency values across the right parafovea. Assuming a saccade targets the point of maximum saliency, the eye will fixate the second word if it's low frequency, or the following word if not. The difference in peak location is a function of the dynamic interplay of character complexity and word frequency.

The purpose of the run of the model shown in Figure 6.6 is to illustrate the interplay of character complexity and word frequency in saccadic control. The initial saliency pattern is driven by character complexity and it's only later in the fixation that lexical processing starts to contribute. In the case of the high-frequency word in the right panel, its peak saliency never exceeds that of the complex character to the right. However, the peak in the low frequency word in the right panel gradually exceeds that of the competing complex character after 200 simulation cycles. If a saccade is triggered at 200 ms, the low-frequency word will be targetted. However, in the left panel its high-frequency counterpart is skipped in favour of targetting the complex character. The simulation suggests a testable hypothesis: in the case of the right panel, if a saccade is triggered early, the low-frequency word will be skipped with the eye targetting the saliency peak associated with the most complex character. On the other hand, later saccades will be associated with fixations on the low-frequency word. Consequently, in a situation where you have a high-complexity character following a low-frequency word, early saccades will target the character whereas later ones will target the low-frequency word. This runs counter to what is normally found in reading – word skips normally follow an elevated fixation duration at the launch location (Kliegl & Engbert, 2005).

While validation of this hypothesis would require a controlled experiment, it is possible to search the text corpus of this study for suitable examples. As it turns out, only one sentence example could be found that was close to what was wanted. Nonetheless, the pattern was as predicted (see Figure 6.7): shorter fixation durations preceded the targetting of the complex characters in the right parafovea and the associated skipping of the relatively low frequency word.

6.3 Conclusion

Following the analysis of chapter 4, the analyses show that lexical factors only play a significant role in word targetting when the eye is launched close to the next word. In these cases, the resulting saccades tend to land close to the word centre. Nonetheless, the dominant factor in saccadic control appears to be character complexity. Multiple sources of evidence suggest that readers' eyes preferentially target characters with the

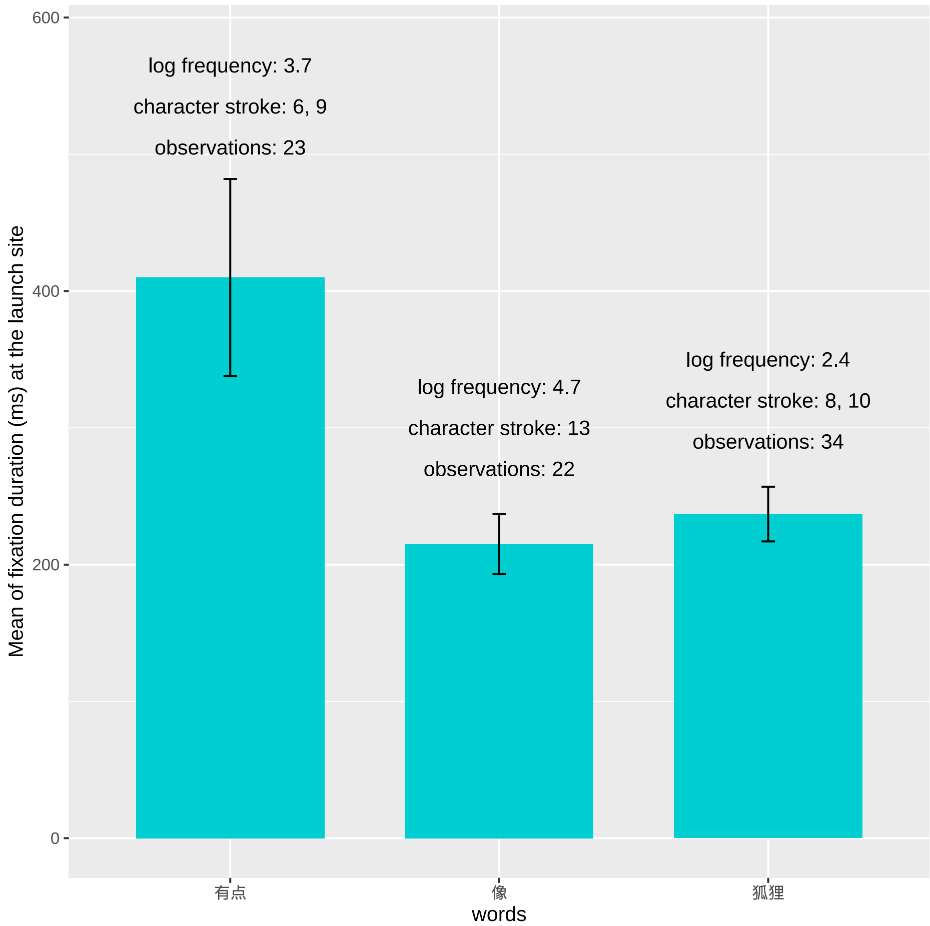


Figure 6.7: Mean fixation duration at the launch location for saccades that land on one of three words in the sentence "浣熊/的/脸/看起来/有点/像/狐狸。", where the launches are rightward from word 4 (看起来) and the landings are on one of the three subsequent words. See text for details.

greatest number of strokes. A key challenge has been to provide a theoretical account that reconciles the lexical and character aspects of the readers' saccadic mechanism in Chinese. The well established Interactive Activation framework (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) provides a possible solution and specifically its instantiation in the Glenmore model of eye-movement control (Reilly & Radach, 2006).

The advantage of the IA framework is that the "activation" in the title can be used to represent influences on the processing of words from a range of sources, such as letter, word frequency, sentential context, and so on. This creates a common currency of information processing which makes it suitable for exploring how lexical and character properties might interact in the processing of text and more specifically in the targetting of eye movements.

The limited model described in this chapter was able to demonstrate the type of dynamic interaction between lexical and character properties that has been proposed in chapter 4. In doing so, the model afforded a number of predictions, one of which appears to be borne out by data in the corpus collected: low frequency words will be skipped if (a) there are complex characters further into right parafovea and (b) a saccade is triggered early in the source fixation.

Obviously, the implications of this “strength-of-character” model need to be teased out further in an experimental rather than corpus-based study. However, the work described in this chapter has created a coherent account of eye movement control in reading Chinese and has provided a platform for exploring these subtle and dynamic processes in more depth.

Chapter 7

Reading Development at the Text Level

7.1 Background

As a reader progresses through a text, depending on their reading goal, they encounter words from which they construct phrases, integrate them into larger sentential and discourse units, and use them to create an isomorphic representation of the writer's conceptual structure. Information is acquired from words or word clusters and then integrated into conceptual units of coarser granularity, such as ideas, events, episodes, narratives. There is considerable evidence that the ongoing cognitive processes involved in reading have a direct impact on the lower-level information processing stages involved in eye movement control (see Radach and Kennedy, 2004, 2013; Rayner, 1998, for overviews). There is also evidence that readers maintain uncertain beliefs about the identities of previously read words, revise these beliefs as a function of grammatical coherence with subsequent linguistic input, and act rapidly on changes in these beliefs through eye movements directed toward the loci of changes in uncertainty. Perceptual input obtained from eye movements is recruited jointly with grammatical knowledge to produce inferences about linguistic form and structure, which in turn play a role in guiding subsequent eye movements (Levy et al., 2009).

7.1.1 Supra-lexical measures

Because of the structural limitations of the eye, the pick-up of word information in reading is a processing bottleneck, but its inherent limitation also presents us with the opportunity to observe the impact of higher level factors relating to sentence structure and meaning on the deployment of this constrained resource. It is possible, therefore, to observe both the impact of processing the current word, but also of prior context through variations in a variety of eye movement parameters. A major challenge in this area of research is to develop quantitative measures of text complexity at a variety of levels that can be used to predict eye movement behaviour. Ideally, one would like to model faithfully the grammatical and conceptual knowledge that the reader is bringing to bear on the task and use it to predict reading behaviour. Levy (2008) proposed the concept of surprisal as a quantitative measure of the cognitive cost required to process a word in a sentence. Surprisal is a text metric that captures how predictable a word is, given the context of preceding words. Substantial progress has been made in developing the measure with the use of conditional probability distributions over interpretations (Boston et al., 2008). A number of easy-to-calculate proxies for surprisal can be derived from various features of large text corpora. If the corpus is large enough, such as is the case with the Google and Microsoft n-gram corpora (Fang et al., 2010; Michel et al., 2011), then we can obtain usable n-gram frequency counts from unigrams (single words) through to 5-grams (e.g., an n-gram is a sequence of N words, a 5-gram is a five-word sequence). The n-gram frequency counts can be used to calculate surprisal values that can act as approximations to both syntactic and to a lesser extent semantic expectations. These in turn can be assessed as predictors of various eye movement parameters.

While surprisal may capture some of the dynamic local constraints on the reader, clearly there are other things going on during reading. For example, if the reader is dealing with an extended text, he or she has the challenge of integrating meaning across sentences. Depending on the coherence of the text, this can prove more or less difficult. Moreover, the current sentence stands in some relation to the text as a whole to the degree that it's more or less central to the theme. A pioneering attempt to quantify global context effects on text reading was reported by Pynte et al. (2008, 2009) using Latent Semantic Analysis (LSA). LSA is a theory and method for analysing documents to find the underlying meaning or concepts inherent in the documents (Landauer et al., 1998). The central idea is that the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other. LSA is one of the most commonly used methods for word meaning representation. However, in recent years neural-network language "embeddings" have received increasing attention (Arora et al., 2016; Kiros et al., 2015; Mikolov et al., 2013a; Mikolov et al., 2013b; Pennington et al., 2014). Furthermore, neural-network derived language embeddings tend to have better performance than LSA on very large training corpora (Altszyler et al., 2016).

Neural-network language embeddings exploit statistical properties of text structure to embed text (words, sentences or paragraphs) into numerical vectors of a fixed number of dimensions (usually between 100 and 500), with more dimensions supporting more nuanced discrimination between meanings. The intuition behind the embedding is that it represents text based on its lexical context accumulated over many millions of instances. Consequently, text appearing in similar contexts will have similar embeddings. Embeddings allow us to easily compute semantic similarity between two texts. A typical way of calculating semantic similarity of language items is to measure the cosine of the angle between the high-dimensional vectors representing the language items; the larger the cosine value, the greater the similarity. A significant recent trend has been the development of so-called universal embeddings (Subramanian et al., 2018; Wieting et al., 2015). Universal embeddings are trained on a variety of data sources and use text classification, semantic similarity, clustering, and other natural language tasks to improve their performance by forcing them to incorporate more general word and sentence features. Google's Universal Sentence Encoder is one example of this approach (Cer et al., 2018; Google research, 2019). Figure 7.1 from Google research (2019) is intended to show that sentence embeddings can be trivially used to compute sentence level semantic similarity scores that achieve excellent performance on the semantic textual similarity (STS) benchmark (Cer et al., 2018). Google's Universal Sentence Encoder is available on TensorFlow Hub (Google research, 2019). The website provides easy-to-use code templates that can be used to encode words, sentences or paragraphs into high dimensional vector embeddings. This then allows the straightforward calculation of similarity among words, sentences and paragraphs.

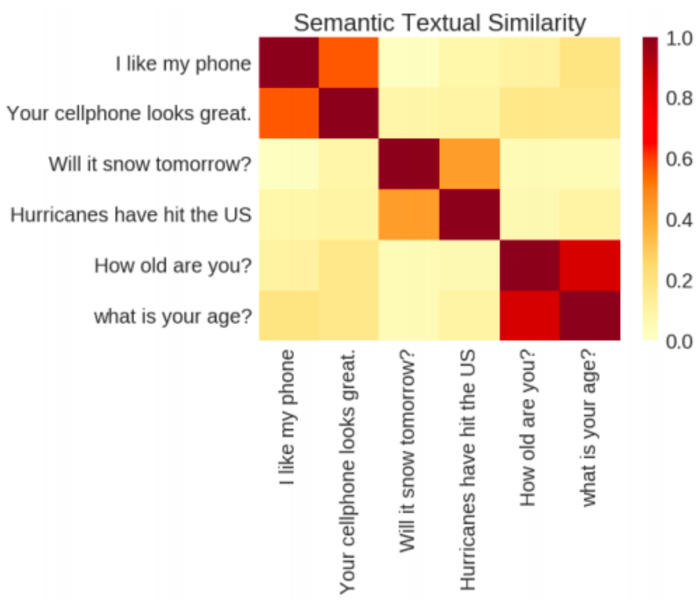


Figure 7.1: Sentence similarity scores using embeddings from the universal sentence encoder

7.1.2 Previous developmental research

In contrast to the substantial literature on adult reading, the number of research publications on eye movements in developing readers is still quite limited (though see Blythe and Joseph, 2011; Radach et al., 2009, for reviews). Much of the groundwork was laid by a set of pioneering studies of text reading (e.g., Buswell, 1922; McConkie et al., 1991; Taylor et al., 1960). These early studies examined elementary school students across various grades, with quantitative analyses mostly restricted to global parameters such as average fixation duration, number of fixations per 100 words, and overall regression frequency. Not surprisingly, this work documented a steady reduction of fixation durations in the number of fixations from grade to grade, whereas the decrease was less pronounced for the proportion of regressive saccades.

McConkie et al. (1991) pioneered the analysis of children's eye movements at the word level, including fine-grained analyses of saccade landing sites within words, and the first quantitative analyses of relations between eye movement parameters and psychometric reading assessments. This was also one of the first studies in which the now common decomposition of viewing times into initial fixation duration, gaze duration and re-reading time was applied to research on developing readers. The idea of this decomposition is to delineate the time course of word processing into time intervals that reflect early orthographic and lexical processing (initial fixations), full lexical access (re-fixation times) and time spent on integration of word meanings into representations on the sentence and text level (rereading times). The analyses of word viewing times in the present chapter will be in line with this logic (see Radach and Kennedy, 2004, for a discussion).

Most current studies on eye movements in developing readers have used experimental designs to address specific research questions, usually comparing children with adults or readers at different stages of their development. This work has usually focused on sub-lexical and lexical components of the reading process, such as letter recognition within the perceptual span (Häikiö et al., 2009; Rayner, 1998) or effects of word length and frequency (Blythe et al., 2011; Blythe et al., 2009; Huestegge et al., 2009; Hyönä & Olson, 1995; Joseph et al., 2009; Joseph et al., 2013). In contrast, higher level, post-lexical processing beyond the word level has received very limited attention. The few exceptions include work on semantic plausibility (Connor et al., 2015; Joseph et al., 2008), syntactic ambiguity (Joseph et al., 2013) and local comprehension monitoring (Vorstius et al., 2013). Van der Schoot et al. (2012) examined global text comprehension using a narrative inconsistency task that allowed individual differences to be assessed in terms of the development and updating of a coherent mental representation of a text passage.

Complementing this previous experimental work, the present chapter will examine supra-lexical effects using surprisal and embedding similarities in a large corpus of eye movement data collected in a longitudinal study of Chinese elementary school students (see below for a detailed description). In the study, the total accumulated time spent on a word during reading was partitioned into three non-overlapping components: first fixation duration (FFD), re-fixation duration (RFD), and the remainder of any viewing time on the word, which will be refer to as re-reading duration (RRD). The central hypothesis of the study is that eye movement metrics are sensitive to surprisal and supra-lexical semantic similarity measures. We should also see developmental changes in sensitivity to these text measures. With respect to local context effects as expressed in surprisal measures, the hypothesis is that surprisal measure will have a stronger effect in the early viewing time phase as it quantifies how predictable a word is, supra-lexical semantic similarity measures will have a stronger influence in the later viewing time phase as they relate to higher-order semantic features of the sentence and paragraph. Furthermore, over the course of development there should be a gradual increase in the sensitivity of readers to such influences. Therefore, it is predicted that there is a greater likelihood of surprisal effects in 5th graders rather than 4th graders, with a gradual increase as they become more skilled readers. Similarly, with respect to global contextual influences it would be expected to see a growing sensitivity to sentence coherence as a child increased their reading proficiency.

7.2 Method

Eye movement data of participants of different reading ability were chosen from the experiment described in chapter 3 for the analysis in this chapter. Equal numbers of readers for the groups “good”, “average” and “poor” were chosen. Due to poor data quality, and participants being transferred to other schools, eye movement data from 19 “good” readers, 20 “average” readers and 20 “poor” readers were used in the analysis of this chapter.

7.3 Analysis

Linear mixed models (Bates et al., 2014) was used in the analysis of the eye movement data from the study. Statistical analysis was conducted using R 3.5.3 (R Development Core Team, 2018) and the lmerTest R package (Kuznetsova et al., 2017). A set of dependent measures were derived from readers' word viewing times. Fixation durations of less than 80 ms and of greater than 800 ms were removed from analyses. Fixations on the first and last words of each line were also excluded. As mentioned previously, the total accumulated time spent on a word during reading was partitioned into three independent components: first fixation duration (FFD), re-fixation duration (RFD), and the remainder of any viewing time on the word, re-reading duration (RRD). These components of a word's viewing time reflect increasingly more advanced stages in the processing of the text. FFD usually reflects the initial processing of a word and tends to reflect local processing constraints. RFD is a measure of the overall difficulty of a word since it reflects the amount of additional fixations the word received before the reader moves on to the next word. RRD measures the total additional time the reader spends on the word following their initial visit and tends to reflect the difficulty a reader has in integrating a word into the an overall understanding of the text. In the study, FFD is expected to reflect the immediate aspects of word processing, RFD somewhat later lexical processing, and RRD higher-level integration processes or more specifically difficulties in performing this integration (see Vorstius et al., 2014 for a recent discussion of these measures).

These three components of viewing time were modelled using the same set of seven fixed independent variables and two random variables as follows.

$$\begin{aligned} DV \sim lg10WF + surp + perp * (word\_sent + sent\_sent \\ + sent\_para + para\_para) + (1 \mid ID) + (1 \mid word) \end{aligned} \tag{7.1}$$

In the analysis described in detail below, the dependent variable DV could be one of the three viewing time decompositions discussed above. Participant ID and word were treated as random effects. Seven additional fixed independent variables are from four categories: word frequency, conditional probability-based surprisal, text similarity, and reading ability (a detailed description of the fixed-effect independent variables is given in Table 7.1).

Surprisal based on n-gram derived conditional probability was calculated for each word in the texts as follows(Levy, 2008):

$$surprisal(w_i) = -\log_2 P(w_i|w_1...w_{i-1}, CONTEXT) \tag{7.2}$$

Where CONTEXT is the extra-sentential context, which will be ignored in the case of this study. $P(w_i|w_1...w_{i-1})$ is the conditional probability of the occurrence of $word_i$ based on its previous words. For example, the conditional probability of $word_i$ based on the last one word can be calculated by the frequency count of $word_{i-1}word_i$ divided

Table 7.1: Description of the fixed-effect independent variable in the mixed-effects models

Variable category	Variable name	Description
word frequency measure	lg10WF	The word frequency measure was taken from a Chinese language subtitle database (Brysbaert and New, 2009). lg10WF is a log frequency measure reflecting the number of times the word appears in the corpus.
surprisal measure	surp	Surprisal values were calculated by the relevant n-grams from Google Chinese Web 5-gram corpora. High values indicate low predictability.
text similarity measures	word_sent, sent_sent, sent_para, para_para	Four sets of values were calculated from embeddings encoded by Google's Universal Sentence Encoder: (1) word_sent measures semantic similarity between fixated word and sentence in which it is contained; (2) sent_sent measures the semantic similarity between the current sentence to the previous sentence; (3) sent_para measures semantic similarity between current sentence and paragraph; (4) para_para measures the semantic similarity between the current paragraph and the previous one.
reading ability	perf	the scores of the literacy test for good, average and poor readers

by the frequency count of $word_{i-1}$. The conditional probability of $word_i$ based on the last two words can be calculated by the frequency count of $word_{i-2}word_{i-1}word_i$ divided by the frequency count of $word_{i-2}word_{i-1}$. Conditional probability based on three or four words can be similarly calculated. As mentioned in the introduction, the conditional probabilities used here were calculated using the Google Chinese Web 5-gram corpus, which consists of Chinese word n-grams and their observed frequency counts generated from over 800 million text tokens. The length of the n-grams in the corpus ranges from unigrams to 5-grams (Fang et al., 2010). In natural language processing practice, it's more common to use trigrams where the probability of $word_i$ is conditional on the probability of its co-occurrence with the previous two words(Jurafsky and Martin, 2014). Also, as n-gram length increases, the problem of data sparsity in the corpus increases. Therefore, surprisal measures are limited to which based on the previous two words in a sentence.

Four measures of text similarity (word-sentence, sentence-sentence, sentence-paragraph, paragraph-paragraph) were calculated using the embeddings derived from Google's Universal Sentence Encoder (Bengio et al., 2003). The Universal Sentence Encoder generates vectors of 128 dimensions for Chinese words, sentences, and paragraphs. The semantic similarity of pairs of language items using their embedding vectors x and y were calculated as follows (also see Figure 7.2):

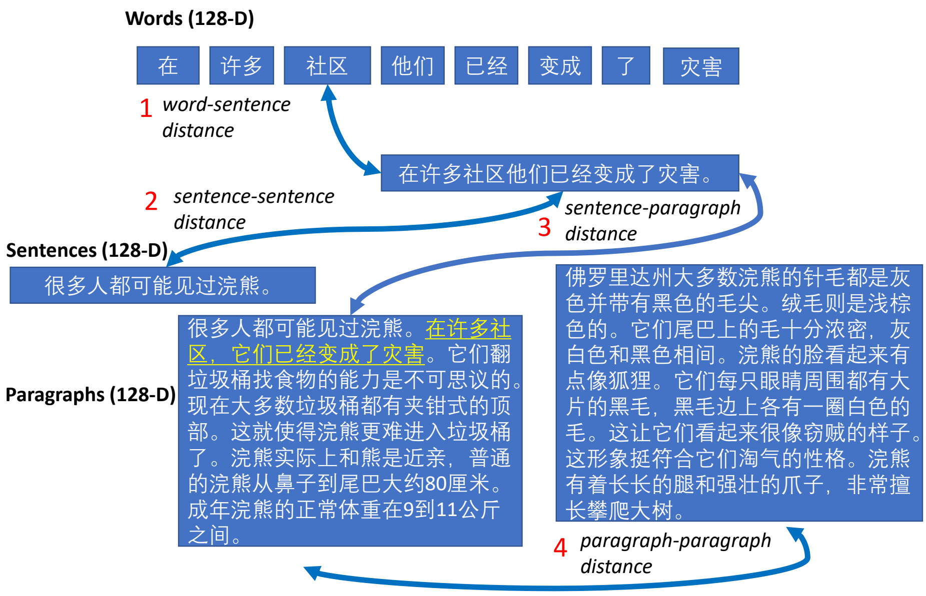


Figure 7.2: Measures of similarity at multiple scales

$$x\_y = 1 - \arccos(\vec{x} \times \vec{y})/\pi \tag{7.3}$$

The variable word_sent measures the semantic similarity between the embeddings representing the fixated word and its containing sentence. This measure would be expected to vary significantly for each word in the sentence. For example, the similarities between function word embeddings and those of the sentence would be quite low compared to those of the sentence's main content words. The variable sent_sent measures the semantic similarity between the sentence in which the fixated word is contained and the preceding sentence. Sentences that are similar by this measure will tend to be dealing with the same topic, whereas a topic-shift between sentences would lead to a decrease in similarity. The variable sent_para measures the semantic similarity between current sentence and its containing paragraph. The measure is intended to quantify the degree to which the current sentence is coherent with its paragraph. It can be viewed as a proxy for the potential impact on the reader of topic shifting or the introduction of new information. The variable para_para measures the semantic similarity between the current paragraph and the previous one. All four measures tap into slightly different aspects of coherence, arguably at different levels of processing. Note that all sentence and paragraph similarity measures are identical for each word in a given sentence.

7.4 Results

The means and standard deviations of the three dependent measures are summarised in Table 7.2. As can be seen from the table, there is a decline in the means and standard deviations of first fixation duration (FFD), re-fixation duration (RFD) and rereading duration (RRD) as readers progress from grade 4 to grade 5. Figure 7.3 below, shows the decomposition of viewing time as a function of reading ability and grade. We can see that the readers classified as poor and average show a slight decline in overall viewing times across grade, while those classified as good show a slight increase in overall viewing times.

Table 7.3 shows the result of the mixed linear models for the three dependent measures. There were overall significant effects of word frequency for all viewing time measures: the higher the frequency, the shorter the viewing time. Reading ability had a statistically significant effect on FFD, with the more able students having shorter

Table 7.2: Numbers of observations, Means and SDs for the three dependent variables (DVs). Note that FFD = first fixation duration; RFD = re-fixation duration; RRD = rereading duration.

	Grade 4			**Grade 5**		
DVs	**N**	**Mean**	**St. Dev.**	**N**	**Mean**	**St. Dev.**
FFD	32,390	248	133	39,129	236	110
RFD	4,417	334	237	4,536	319	232
RRD	10,443	441	351	13,346	433	347

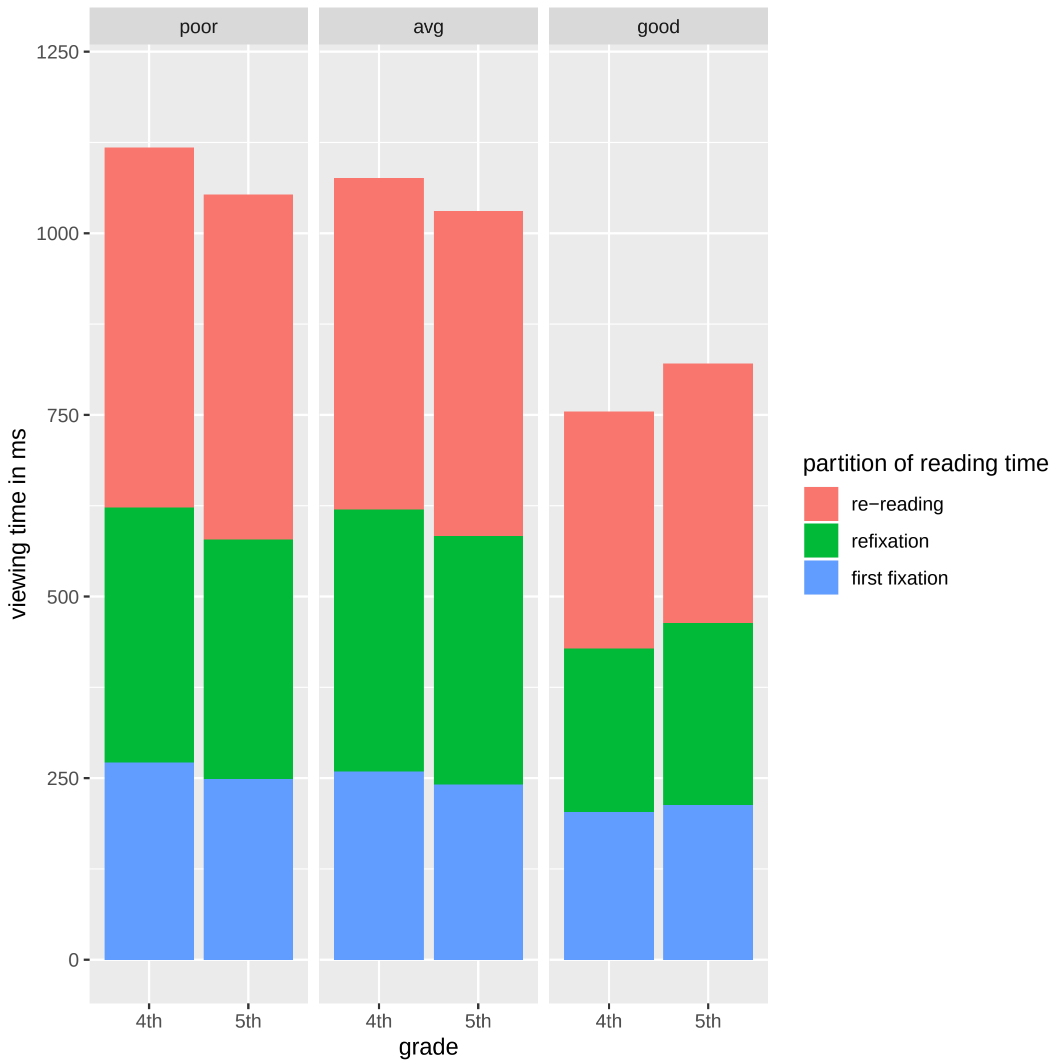


Figure 7.3: Decomposition of viewing times as a function of grade and reading ability

viewing times. Surprisal only had significant effect on FFD: the higher the surprisal, the longer the viewing time. Contrary to what was expected, the text similarity measures only had a statistically significant impact on the early viewing time (FFD): the more similar the current and preceding sentence, the shorter the first fixation duration.

However, the result of the mixed linear models for the three dependent measures done separately for each grade show a slightly different picture of FFD and RRD for fifth graders (Table 7.4).

Table 7.3: Linear mixed model estimates for the three dependent measures

	Dependent variable		
	FFD	**RFD**	**RRD**
lg10WF	**−21.4(1.2)*****	**−19.1 (1.1)*****	**−43.8(2.5)*****
surp	**0.4(0.2)****	**0.3(0.1)***	−0.3(0.2)
perf	**−115.5(61.8)*****	−47.5(43.7)	−115.7(94.8)
word_sent	−182.3 (182.9)	159.8 (134.1)	532.6(291.5)
sent_sent	**−346.7(159)***	−175.1(116.5)	−210.2(253.2)
para_para	101.4(198.7)	−88.7(145.6)	−571.9(316.5)
sent_para	293.7 (160.3)	66(117.5)	277.3(255.4)
perf: word_sent	62.2 (59.7)	−45 (43.8)	−159.5(95.1)
perf:sent_sent	**108.5(52)***	52.2(38.1)	74.5(82.9)
perf:para_para	−23.8 (64.8)	30.9(47.5)	176.9(103.2)
perf:sent_para	−101.7 (52.4)	−20.7(38.5)	−104.5(83.5)
intercept	**592.1(189.4)****	236.5(133.9)	**636.6(290.5)***

Note: $^{*}p\leq0.05$; $^{**}p\leq0.01$; $^{***}p\leq0.001$

Table 7.4: Linear mixed model estimates for the three dependent measures across grade

	Dependent variable					
	FFD$_{G4}$	**FFD**$_{G5}$	**RFD**$_{G4}$	**RFD**$_{G5}$	**RRD**$_{G4}$	**RRD**$_{G5}$
lg10WF	**−21.3 (1.6)*****	**−22.6 (1.3)*****	**−19.6(1.3)*****	**−17.4(1.2)*****	**−39.7(3.2)*****	**−44(2.8)*****
surp	**0.6(0.2)***	0.3(0.2)	0.3(0.2)	0.2(0.1)	−0.4(0.4)	−0.1(0.3)
perf	**−254.7(101.6)***	**−179.8(74.8)***	−113.1(75.5)	−49.6(52.5)	**−315.6(154)***	−53.4(120.6)
word_sent	−49.5 (285.9)	−393.8 (232.6)	201.1(217.2)	93.2(168.2)	**1081.7(443.1)***	38.8(381.6)
sent_sent	−21.3(251.6)	−249.7(204)	−272.4 (191.1)	23.7 (147.4)	−181.4(389.8)	152.9(334.5)
para_para	**−1095.9 (389)****	−385.7 (229.9)	−477.3 (295.4)	−201 (166.5)	**−1630.8(602.7)****	−706(337.1)
sent_para	293.7(246.8)	384.5(207.6)	164(187.4)	−2.7(150.1)	−104.4(382.4)	635.6 (340.5)
perf:word_sent	8.8 (93.3)	141.8 (75.9)	−62.4(70.9)	−19.9(54.9)	**−342.1(144.5)***	2(124.5)
perf:sent_sent	−21.2(82.4)	94.6(66.7)	86.5 (62.6)	−9 (48.2)	51.6(127.6)	−27.1(109.4)
perf:para_para	**332.1 (126.6)****	112.7 (75)	141.3 (96.2)	65.6 (54.2)	**518.2(196.2)****	202.7(123)
perf:sent_para	−96(80.6)	**−135.7(67.9)***	−50.3(61.2)	−0.5(49.1)	28.6(124.8)	−232.9 **(113.3)***
intercept	**1111.9(311.1)*****	**784.8(229.1)*****	**473.9 (231.2)***	233.3(161)	**1261.9(471.9)****	463.8(369.9)
Note:						*p≤0.05; **p≤0.01; ***p≤0.001

7.4.1 First fixation duration

In the overall analysis, reading ability and word frequency were the dominant factors affecting first fixation duration (FFD) and doing so in the obvious direction - higher word frequency and greater reading ability significantly reduced FFD. Increased surprisal also significantly raised FFD - the more unexpected the word in the sentence context, the longer the FFD. One of the similarity measures, sentence-to-sentence, has a marginally significant effect on FFD, suggesting that if the current and preceding sentence are similar enough, it will shorten FFD for the words in the current sentence. However, this effect is more pronounced for poorer readers, as evidenced by the significant ability-by-similarity interaction. Figure 7.4 shows FFD as a function of current and preceding sentence similarity and reading ability after removal of between-subject and between-word variance of the dependent variables (Hohenstein & Kliegl, 2014). In this figure, we can see that the source of the significant interaction is a slight floor effect, whereby there is scope for shortening the FFD as result of sentence similarity in the case of less proficient readers, but that there's less room for improvement for the shorter FFDs of more able readers.

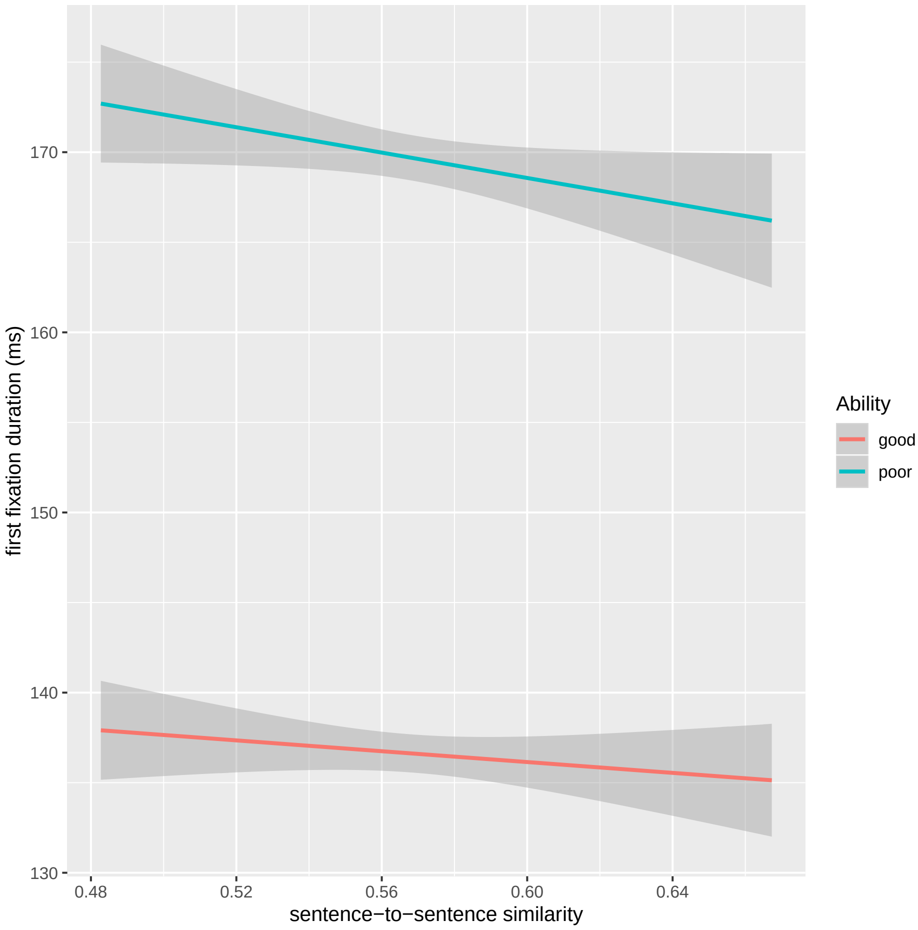


Figure 7.4: FFD as a function of current and preceding sentence similarity and reading ability: random effects removed

In Table 7.4, grade 4 values are compared to grade 5 and apart from consistent frequency and ability effects across grades, it appears that greater paragraph similarity benefits grade 4 students more than grade 5. Moreover, reading ability also plays a role in this effect, with the less able readers benefiting more from paragraph similarity as indicated by the significant interaction with ability. This interaction is visualised in Figure 7.5, which shows FFD for 4th and 5th grade as a function of current and preceding paragraph similarity and reading ability after removal of between-subject and between-word variance of the dependent variables. The greater benefit for the less able readers is another instance of the floor effect mentioned above, where there is an uncompressable lower bound on FFD in grade 4, which limits the amount of improvement that can be gained from exploiting paragraph similarity. The pattern of data for grade 5 in the same figure shows an overall decrease in FFD, with the pattern of interaction reversed – poor readers showing no benefit from paragraph similarity, but more able readers benefiting. Note, however that this apparent interaction is not statistically significant.

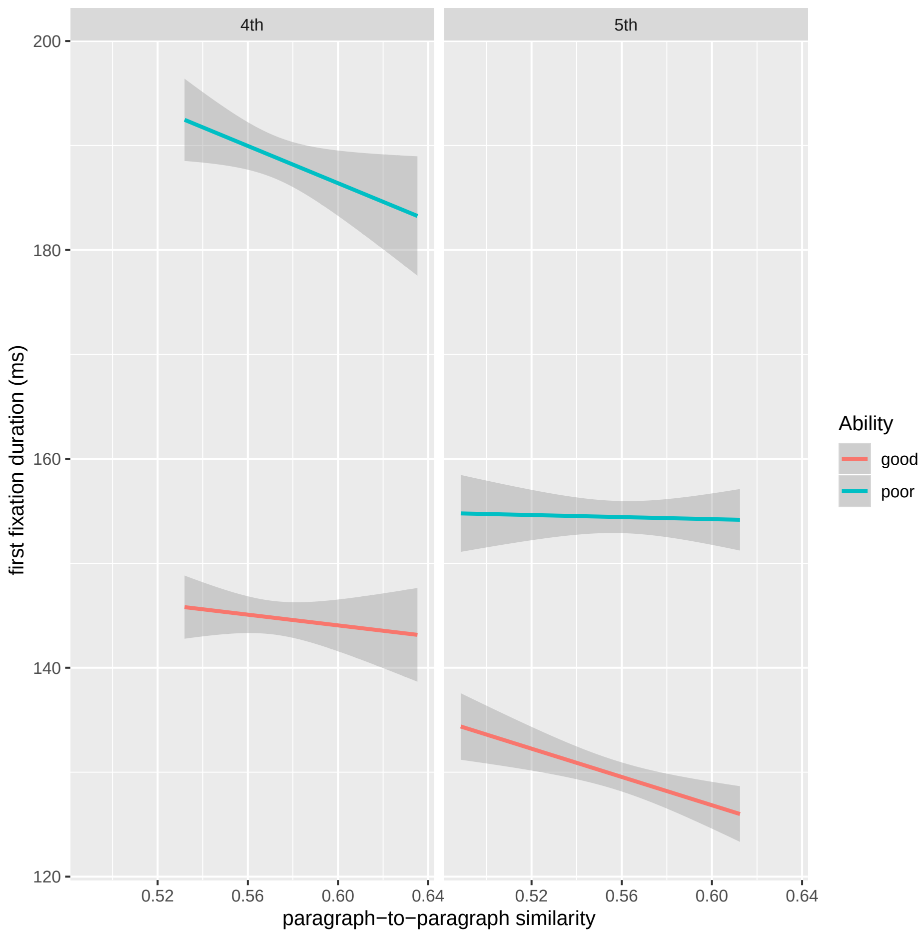


Figure 7.5: FFD for 4th and 5th grade as a function of current and preceding paragraph similarity and reading ability: random effects removed

7.4.2 Refixation duration

In the global analysis of refixation duration (RFD), while there is a significant effect of word frequency and a marginal effect of surprisal, no other independent measures reach statistical significance. When we look at the analysis by grade, there is only a significant effect of word frequency. The absence of significant effects may be a consequence of the relatively small number of word refixations in the data. This probably arises from the unique characteristics of Chinese text. In the texts used for the study, 54% of the words consisted of just one character, while 42% comprised two characters. More generally, 70% of Chinese words contain two characters, 20% contain one character, and 10% contain three or more characters (Modern Chinese Frequency Dictionary, 1986). The shorter average word length in the texts reflects the fact that they were designed for reading by children. Furthermore, given the spatial compactness of the Chinese writing system, one fixation will tend to suffice in most cases for the satisfactory identification of words. Overall, therefore, the RFD measure may not be as reliable a metric of word processing in the case of Chinese reading as it is for alphabetic writing systems with more heterogeneous word lengths. In effect, the RRD measure for Chinese reading might tend to incorporate the RFD metric that normally would be distinct in alphabetic reading.

7.4.3 Re-reading duration

In the global analysis, only word frequency had a significant effect on re-reading duration (RRD) and in the expected direction. No other independent measures reach overall statistical significance.

However, when RRD is analysed separately by grade, there is a significant sensitivity to various similarity measures among grade 4 readers. The strongest effects are for paragraph similarity, with high similarity reducing RRD. Figure 7.6 shows RRD for 4th and 5th grade as a function of current and preceding paragraph similarity and reading ability after removal of between-subject and between-word variance in the dependent variables. The source of the interaction between ability and paragraph similarity in 4th grade readers appears this time to be the greater benefit good readers derive from paragraph similarity compared to less able readers. A similar pattern is seen for 5th grade readers, though not a statistically significant one.

Figure 7.7 shows the marginally significant interaction between reading ability and word-sentence similarity for grade 4 readers. Te source of the interaction is not entirely clear and, moreover, it is not clear why greater word-sentence similarity should lead to an increase in RRD for both ability levels. One possibility is that it's due to the reader's integration efforts. Recall that RRD measures repeated visits to a word, so perhaps greater similarity gives rise to an increase in confirmatory fixations. The elevated trend in RRD for good readers is also apparent in grade 5, though without reaching significance.

7.4.4 Path analysis of semantic similarity measures

Given the, albeit relatively small, colinearity of the family of semantic similarity measures used in the preceding analysis and the potential complexity of interpretation of their influence as discussed above, it was decided to unpack the pattern of interdependence among the measures in more detail using path analysis (Rosseel, 2012). Specifically, the degree to which these measures directly and indirectly influenced two of the three dependent measures: FFD and RRD was explored. RFD was omitted from these analyses because of the relatively low numbers of observations involved after partitioning the data and also the problematic nature of RFD given the homogeneous word length of Chinese, particularly Chinese in children's texts. It would also be informative to see if the patterns of influence varied across grade and/or categories of reading ability. Therefore, it was hypothesised that the early viewing time measures, such as FFD, would show less direct and indirect influence from the sentence

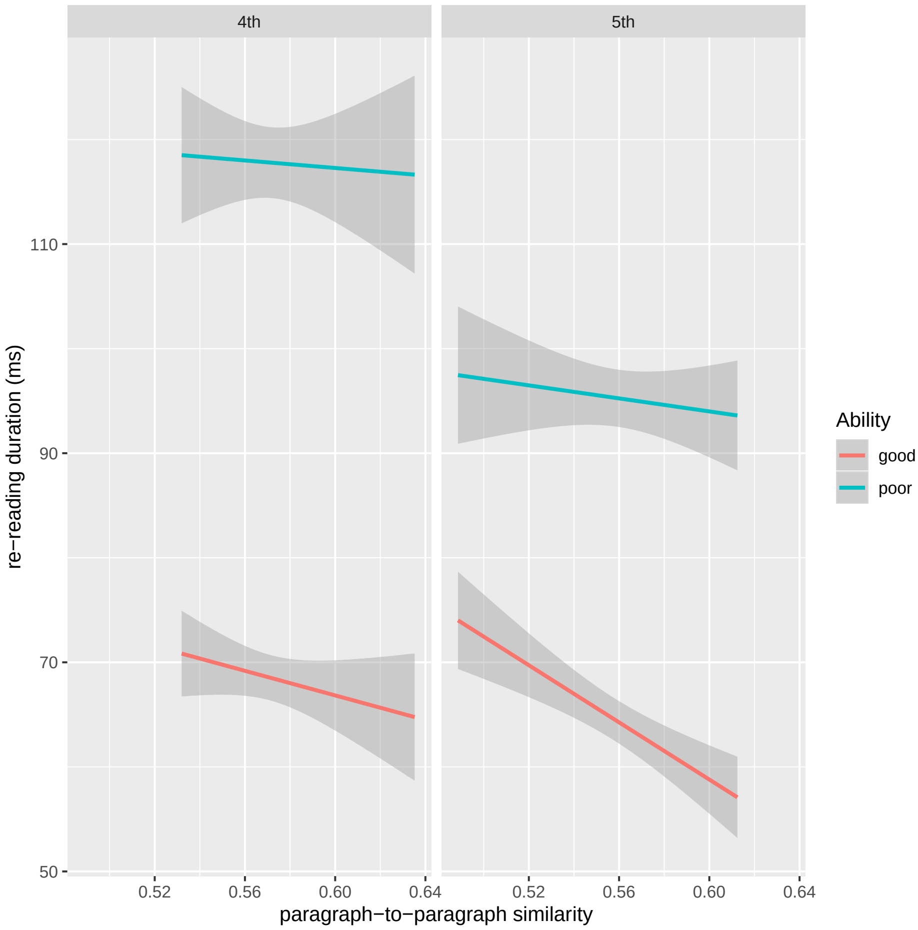


Figure 7.6: RRD for 4th and 5th grade as a function of current and preceding paragraph similarity and reading ability: random effects removed

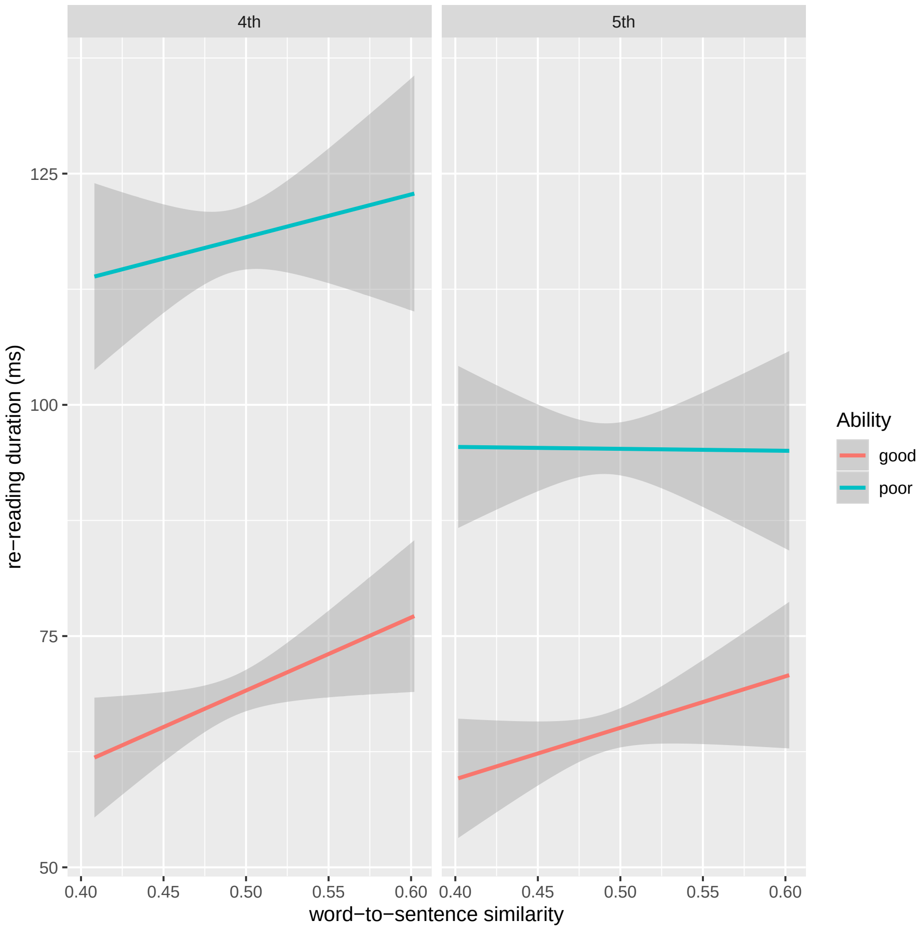


Figure 7.7: RRD for 4th and 5th grade as a function of word and sentence similarity and reading ability: random effects removed

and paragraph similarity measures than the later RRD. Furthermore, earlier grades and poorer readers should show less direct and indirect influence of extra-sentential similarities than either later grades or better readers on the assumption that less skilled readers are more narrowly focused on the local lexical context and less on the sentential and paragraph level.

Table 7.5 is the output from an overall path analysis examining the pattern of influences (direct and indirect) on FFD and RRD. The first sub-figure of Figure 7.8 represents all possible paths of influence. Using a criterion of $p < 0.05$ to assess each path's significance, it is apparent that RRD shares the same pattern of influences as FFD but with additional pathways from sentence-paragraph similarity. This difference can be viewed as a progression from sentence level to paragraph level influences to which FFD and RRD are expected to be differentially sensitive. Note that in Figure 7.8 the dashed paths represent negative estimates, which indicate a reduction in viewing time as a function of increased similarity, positive estimates indicate an increase in viewing time. Note also that word frequency and surprisal are included in the path analysis but are not shown in the path diagrams. The inclusion of surprisal may explain the surprising absence word-sentence similarity effects in this and all subsequent path analyses.

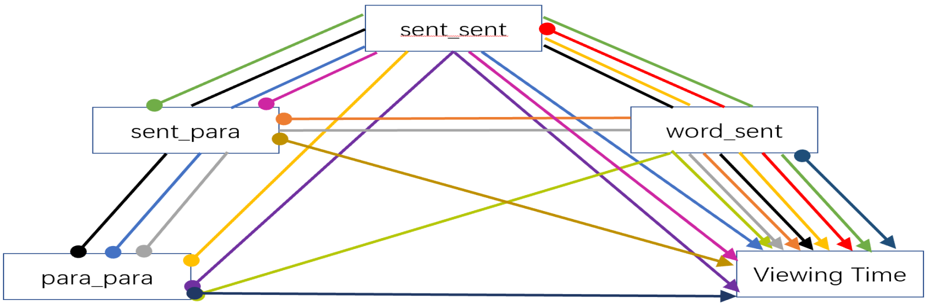
The strength of various paths' influence on reading time measures can be considered to reflect efforts by the reader to integrate what they are reading into their developing understanding of the text. Therefore, an analysis of the strength of paths for readers in different grades and of different reading ability should tell us something about readers' growing sensitivity to higher-order properties of the text. The main area where we would expect to see differences is in the FFD and RRD viewing time measures, since the preceding LMM analyses have shown that semantic similarity measures have a significant effect on these two variables.

Table 7.6 shows comparative path analyses for the FFD measure between grades 4

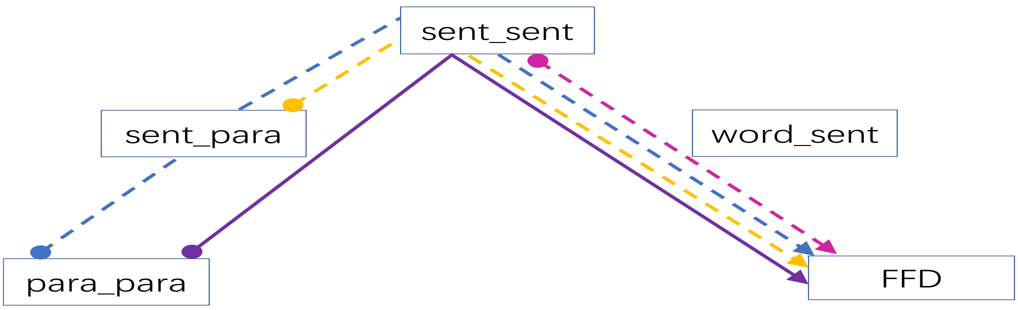
Table 7.5: Estimates (and their standard errors) for the path analyses of the impact of the similarity measures on the three dependent measures. Note that pp = paragraph_paragraph similarity; sp = sentence_paragraph similarity; ss = sentence_sentence similarity; ws = word_sentence similarity.

	FFD	**RRD**
pp→	0.206 (0.115)	−0.176 (0.105)
ss→	**−0.736 (0.097)*****	**−0.191 (0.088)***
sp→	0.045 (0.097)	**−0.188(0.089)***
ws→	−0.145 (0.111)	−0.000 (0.101)
pp→sp→ss→ws→	0.001 (0.000)	0.000(0.000)
pp→sp→ws→	−0.001 (0.001)	−0.000(0.000)
pp→sp→ss→	**−0.07 (0.009)*****	**−0.018 (0.008)***
pp→ss→ws→	−0.001 (0.001)	−0.000 (0.001)
pp→ss→	**0.139 (0.018)*****	**0.036(0.017)***
pp→ws→	−0.002 (0.001)	−0.000 (0.001)
pp→sp→	0.015 (0.032)	**−0.062 (0.029)***
sp→ss→ws→	0.002 (0.001)	0.000 (0.001)
sp→ws→	−0.002(0.002)	−0.000 (0.001)
sp→ss→	**−0.212 (0.028)*****	**−0.055 (0.025)***
ss→ws→	0.006 (0.005)	0.000 (0.004)

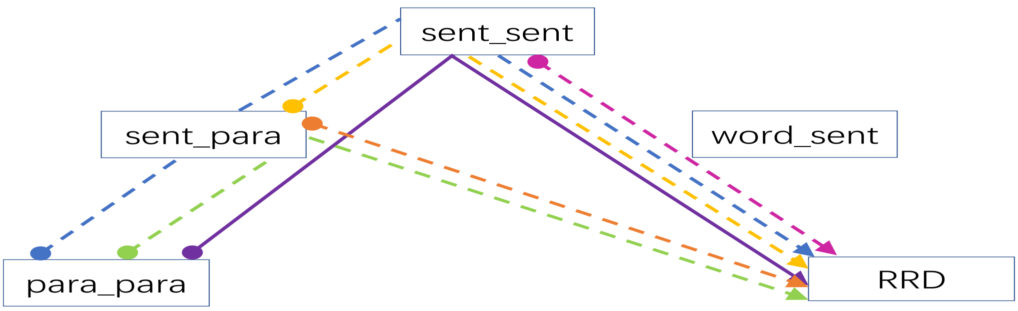
Note: *p≤0.05; **p≤0.01; ***p≤0.001



(a) All



(b) FFD



(c) RRD

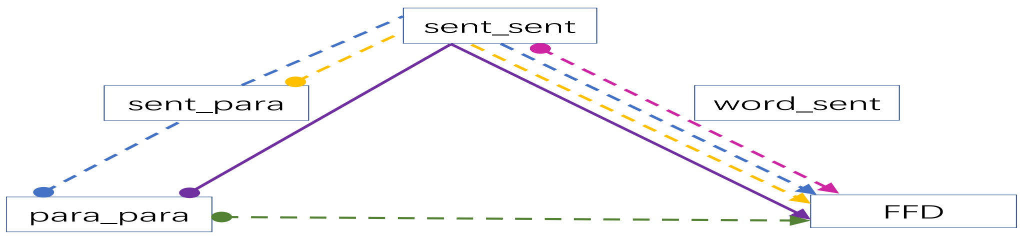
Figure 7.8: Path analysis showing the patterns of influence of the four text similarity measures used in the analysis of viewing time data: (a) all possible paths; (b) significant paths influencing first fixation duration (FFD); and (c) significant paths influencing re-reading duration (RRD). Distinct paths are illustrated with different colours, where the start of the path is represented by a filled circle and its end by an arrowhead. In (b) and (c) only paths with p<0.05 are graphed, dashed paths indicate a negative estimate, and solid lines a positive one.

and 5, and between good and poor readers. Figure 7.9 shows a difference in pathways of FFD for different classes of readers. The pattern of pathways for FFD for 4th graders and poor readers are very similar, with the exception of a different sign for the direct path from paragraph-paragraph similarity. However, grade 5 and good readers have identical sets of significant paths. This pattern of influences suggests that early readers and poor readers are affected by the effort for story-level integration even to the extent of it affecting early lexical processing.

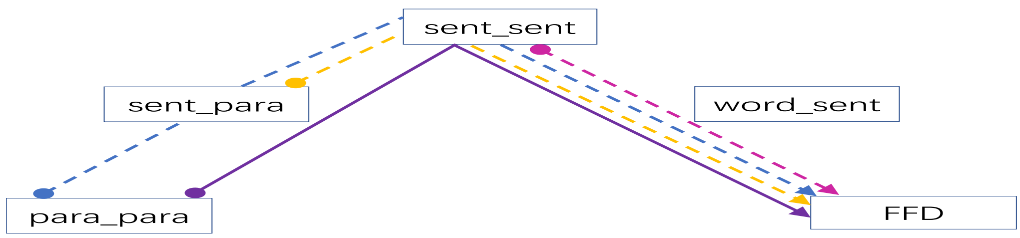
Table 7.6: Estimates (and their standard errors) for the path analyses of the impact of the similarity measures on the first fixation time as a function of grade and reading ability. Note that pp = paragraph_paragraph similarity; sp = sentence_paragraph similarity; ss = sentence_sentence similarity; ws = word_sentence similarity.

	$\mathbf{FFD}_{G4}$	$\mathbf{FFD}_{G5}$	$\mathbf{FFD}_{poor}$	$\mathbf{FFD}_{good}$
pp→	**−0.517 (0.221)***	−0.058 (0.139)	**0.484 (0.198)***	−0.24 (0.198)
ss→	**−0.749 (0.146)*****	**−0.55 (0.132)*****	**−0.831 (0.165)*****	**−0.744 (0.170)*****
sp→	0.269 (0.143)	0.073 (0.134)	0.251 (0.166)	0.085 (0.169)
ws→	−0.249 (0.165)	−0.117 (0.151)	−0.272 (0.19)	−0.106 (0.193)
pp→sp→ss→ws→	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)
pp→sp→ws→	−0.005 (0.004)	0.001 (0.001)	−0.001(0.001)	−0.000(0.001)
pp→sp→ss→	**−0.095 (0.019)*****	**−0.049 (0.012)*****	**−0.078 (0.016)*****	**−0.068 (0.016)*****
pp→ss→ws→	−0.001 (0.001)	−0.000 (0.001)	−0.002(0.002)	−0.001(0.002)
pp→ss→	**0.317 (0.062)*****	**0.037 (0.009)*****	**0.158 (0.032)*****	**0.154 (0.035)*****
pp→ws→	−0.02 (0.013)	0.002 (0.002)	−0.003 (0.003)	−0.001 (0.002)
pp→sp→	0.119 (0.063)	0.021 (0.04)	0.083 (0.055)	−0.027 (0.053)
sp→ss→ws→	0.000 (0.000)	0.002 (0.003)	0.003 (0.002)	0.001 (0.002)
sp→ws→	−0.012 (0.008)	0.003 (0.004)	−0.003(0.003)	−0.002(0.003)
sp→ss→	**−0.214 (0.042)*****	**−0.164 (0.039)*****	**−0.237 (0.047)*****	**−0.215 (0.049)*****
ss→ws→	0.001 (0.001)	0.007 (0.009)	0.012 (0.008)	0.004 (0.008)

Note: $^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$

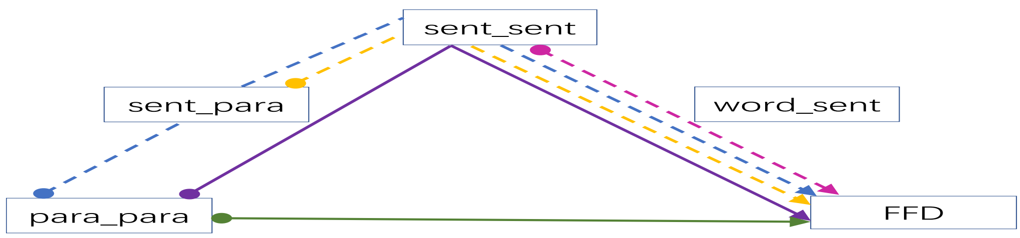


(a) Grade 4



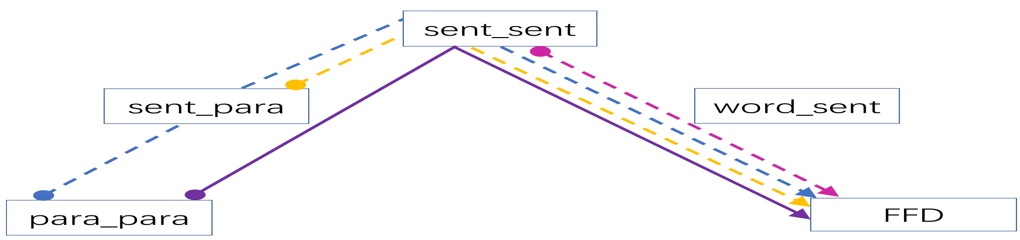
(b) Grade 5

−0.1in



(c) poor readers

−0.1in



(d) good readers

−0.1in

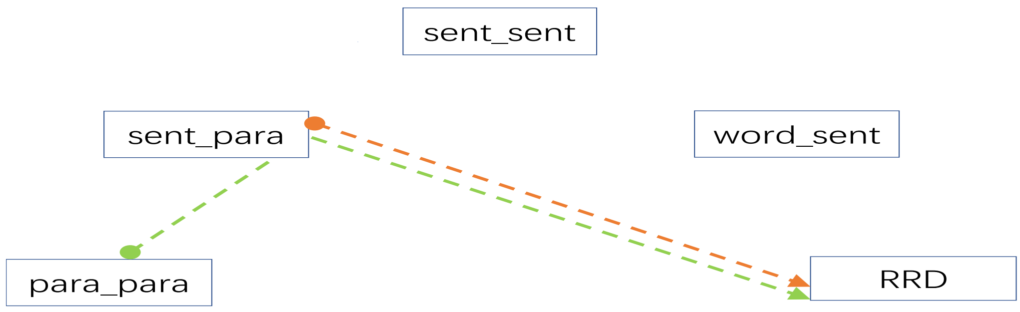
Figure 7.9: Results of a path analysis showing the pattern of influences of the four similarity measures used in the analysis of viewing time data. The viewing time components in question are (a) first fixation duration (FFD) for 4th graders, (b) FFD for 5th graders, (c) FFD for readers classified as "good" and (d) FFD for readers classified as "poor". Distinct paths are illustrated with different colours, where the start of the path is represented by a filled circle and its end by an arrowhead. Only paths with $p<0.05$ are graphed. Dashed paths indicate a negative estimate, solids lines a positive one.

Finally, Table 7.7 shows the results of comparative path analyses for the two levels of reading ability and two grades for the RRD viewing time measure. Interestingly, none of the paths influencing RRD were significant in the case of either grade 4 or poor readers. There were, however, significant sentence and paragraph-level effects for good readers and grade 5 readers. It appears that in the case of less skilled readers, sentence and paragraph-level information is not available to these readers or, if it is, the readers' limited resources are focused on the more immediate task of word processing (cf. Figure 7.10).

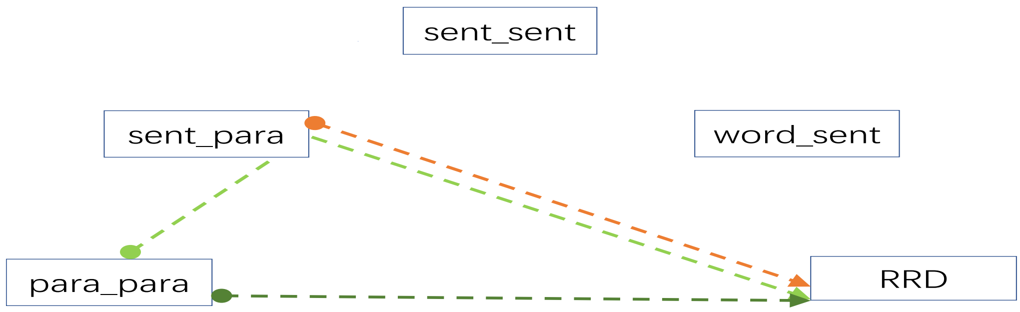
Table 7.7: Estimates (and their standard errors) for the path analyses of the impact of the similarity measures on the re-reading time as a function of grade and reading ability. Note that pp = paragraph_paragraph similarity; sp = sentence_paragraph similarity; ss = sentence_sentence similarity; ws = word_sentence similarity.

	RRD$_{G4}$	**RRD**$_{G5}$	**RRD**$_{poor}$	**RRD**$_{good}$
pp→	−0.33 (0.206)	−0.186 (0.125)	−0.136 (0.197)	**−0.443(0.158)****
ss→	−0.256 (0.137)	−0.139 (0.119)	−0.058 (0.164)	−0.143(0.135)
sp→	−0.103 (0.133)	**−0.256 (0.12)***	−0.105(0.165)	**−0.271 (0.134)***
ws→	0.18 (0.154)	−0.143 (0.135)	−0.088 (0.189)	0.152(0.154)
pp→sp→ss→ws→	−0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	−0.001 (0.001)
pp→sp→ws→	0.004 (0.003)	0.001(0.001)	−0.000(0.001)	0.001(0.001)
pp→sp→ss→	−0.032 (0.017)	−0.012 (0.01)	−0.006 (0.015)	−0.013 (0.012)
pp→ss→ws→	0.000 (0.000)	−0.000 (0.001)	−0.001 (0.002)	0.001 (0.001)
pp→ss→	0.108 (0.058)	0.009 (0.008)	0.011 (0.031)	0.029(0.028)
pp→ws→	0.014 (0.012)	0.002 (0.002)	−0.001 (0.002)	0.002(0.002)
pp→sp→	−0.045 (0.059)	**−0.078 (0.036)***	−0.035(0.055)	**−0.086 (0.043)***
sp→ss→ws→	−0.000 (0.000)	0.002 (0.002)	0.001(0.002)	−0.002(0.002)
sp→ws→	0.009 (0.007)	0.003 (0.003)	−0.001(0.002)	0.002(0.002)
sp→ss→	−0.073 (0.039)	−0.041 (0.035)	−0.017(0.047)	−0.041(0.039)
ss→ws→	−0.001 (0.001)	0.008 (0.008)	0.004 (0.008)	−0.006(0.005)

Note: *p≤0.05; **p≤0.01; ***p≤0.001



(a) Grade 5



(b) Grood readers

Figure 7.10: Results of a path analysis showing the pattern of influences of the four similarity measures used in the analysis of viewing time data. The viewing time components in question are (a) re-reading duration (RRD) for 5th graders, and (b) RRD for good readers. Distinct paths are illustrated with different colours, where the start of the path is represented by a filled circle and its end by an arrowhead. Only paths with $p<0.05$ are graphed. Dashed paths indicate a negative estimate, solids lines a positive one.

7.5 Conclusion

The results of this chapter demonstrate that text similarity measures have a significant impact on moment-to-moment processing of words in reading. Previous research has demonstrated that n-gram and LSA contextual measures have an impact on word viewing times in adult readers (Pynte et al., 2008, 2009). However, this is the first attempt to track the developmental trajectory of these influences in Chinese early readers as well as readers with differing reading abilities.

While the underlying factors driving the developmental change in response to supra-lexical properties of the text clearly need to be explored further, on the basis of what has been found one should be cautious about attributing the main changes in the developing reader merely to changes in the efficiency of lexical processing (cf. Engbert et al., 2005; McDonald et al., 2005; Reichle et al., 2003; Reilly and Radach, 2006 for arguments along these lines). The study finds that there are clear and robust contextual effects impacting on the local processing of words during a fixation and the nature of these effects changes as the reader becomes more proficient.

Another contribution of the findings in this chapter is to present an easy-to-use set of tools for linking the low-level aspects of fixation durations to a hierarchy of sentence-level and paragraph-level features that can be computed automatically. The use of the decomposition of word viewing times into immediate and later components combined with measures of sentence and paragraph coherence illuminates the time-course of the reader's processing of a text. Similar to the study by Radach et al. (2008), broader contextual constraints impact on low-level aspects of the reading process have been demonstrated. However, the techniques described here allow these affects to be measured post-hoc rather than requiring them to be incorporated into an experimental design, as was the case with Radach et al's (2008) paper.

Finally, the similarity-based measures could be used to assess text for their suitability for readers of different levels of ability. While there are text complexity measures such as “Lexile” (Lexile, 2019) available for English texts, nothing comparable exists for Chinese. The measures described here could be applied to any language for which there is a large text corpus. Moreover, the Lexile measure is primarily calculated as a function of word frequency and sentence length. The text coherence measures described here could usefully augment a lexile-like measure to provide quantitative measures of the semantic characteristics of the sentences and texts in addition to word frequency and sentence length.

Chapter 8

General Discussion

The final chapter comprises two sections. The first summarises the findings of the thesis and their contribution to the literature on Chinese reading. The second outlines some suggested directions for future research based on the these finding.

8.1 Summary of findings

At the beginning of the thesis it was reported that several large-scale national interventions in reading instruction in the US had failed to deliver any significant improvements in student performance. Evaluators of the interventions concluded that this was due to a failure to take sufficient account of the specific characteristics of individual readers and tailoring the interventions accordingly. Motivated, in part, by these results the thesis focussed on attempting to gain an insight into the underlying processes of readers of Chinese from the early perceptual level through to the formation of meaning representations in the hope that a focus on individual readers or specific sub-groups of them may lead to deeper understanding and ultimately more effective approaches to the teaching of reading. While the overall goals were ambitious, the results of the research described here makes some progress towards these goals in both devising new techniques for studying readers and new results that deepen our understanding of the reading process both in Chinese and potentially in other writing systems. The key questions set out at the beginning of the thesis and the relevant findings addressing them are summarised below:

- *Characteristics of Chinese readers at different points in the reading ability spectrum from poor to good and their progression from 4th to 5th grade*

  The thesis is one of the few large scale studies of young Chinese readers and the resulting corpus of data is a valuable research resource in itself. As well as providing a substantial corpus of eye movement data, the study also amassed a range of complementary IQ and language ability measures to provide context for the analyses of the eye movement data.

  The overall results of the data corpus yielded few surprises, but confirmed the alignment of the pattern of readers' learning and developmental progression with those of readers of other writing systems. This, in turn, confirmed the validity of using the key viewing time metrics of first fixation duration (FFD), refixation duration (RFD), and re-reading duration (RRD). However, the RFD measure proved of limited utility because of the unique features of Chinese, specifically the relatively short word-lengths involved in the age-appropriate texts used.

- *Factors affecting eye movement control in reading Chinese*

  There is an ongoing debate about what factors determine where the eye goes when reading Chinese script. Accounts of eye movement control derived from spaced alphabetic writing systems tend to be word-focussed, with the main driver of the proposed mechanism being the targetting of word centres. Words present a relatively easy target in spaced writing, but in Chinese the issue is complicated by the lack of inter-word spaces as well as ambiguity about word segmentation. Therefore, lexically driven saccadic control models are faced with significant challenges when faced with Chinese.

The thesis presents results that suggest the control of eye movements in Chinese is a function of both character and lexical factors. Evidence is presented that complex characters involving large numbers of strokes are preferentially targetted by the reader. There is also evidence that eye movement control comes under the influence of lexical factors such as word frequency when the eye is near the beginning of a word. Clearly these two drivers of eye movement control interact to give the overall pattern of eye movement behaviour. However, to successfully understand how this happens requires a suitable modelling framework and its computational instantiation.

- *Computational models as tools to explore Chinese reading*

  This thesis has made use of three computational tools to help further our understanding of the reading data: a machine learning model trained on human data to develop a system for predicting character confusion; a model of word, sentence, and paragraph relatedness based on vector-based representations of text referred to as embeddings; and a computational model of eye movement control that extends an earlier model designed for an alphabetic writing system. While the character confusion model has still some way to go before it can be successfully employed, the text similarity methodology and computational model both represent significant additions to the reading researchers toolkit for studying Chinese reading.

- *Supra-lexical influences on viewing times*

  Probably the most immediately useful finding of the thesis is the establishment of relationships between eye movement data and the embedding-based text similarity measures developed. More significantly, these relationships also correlated with increases in reading ability. The similarity measures can potentially be used as diagnostics of readers' sensitivity to sentential and paragraph level factors and also be used as metrics of text coherence and age appropriateness. They also offer the possibility of incorporating the text level factors into an enhanced version of the proposed computational model.

8.2 Future work

One way to manage the complexity of the reading process as well as accommodating individual differences in reading ability is to use individualised computational models. Therefore, future work building on the research described here will be aimed at (1) demonstrating the utility of an extended version of Glenmore in understanding the reading process at an individual level; (2) accounting for reading performance data across the ability range, from poor to more able readers; and (3) integrating both decoding and comprehension research efforts at the empirical and modelling levels. The work described in this thesis lays the foundation for developing individualised models, which in turn will offer the potential for targetting teaching interventions more effectively.

References

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & cognition*, *24*(4), 477–492.

Altszyler, E., Sigman, M., Ribeiro, S., & Slezak, D. F. (2016). Comparative study of lsa vs word2vec embeddings in small corpora: A case study in dreams database. *arXiv preprint arXiv:1610.01520*.

Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. in 1. t. guthrie (ed.), comprehension and teaching: Research reviews (pp. 77-117). newark. *DE: International Reading Association*.

Arora, S., Liang, Y., & Ma, T. (2016). A simple but tough-to-beat baseline for sentence embeddings.

Aslin, R. N., & Shea, S. L. (1987). The amplitude and angle of saccades to double-step target displacements. *Vision research*, *27*(11), 1925–1942.

Bai, X., Li, X., & Yan, G. (2015). Eye movement control in chinese reading: A summary over the past 20 years of research. *Psychological Development and Education*, *31*(1), 85–91.

Bai, X., Meng, H., Wang, J., Tian, J., Zang, C., & Yan, G. (2011). The landing positions of dyslexic, age-matched and ability-matched children during reading spaced text. *Acta Psychologica Sinica*, *43*(08), 851–862.

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1277.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision research*, *19*(9), 967–983.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, *3*(Feb), 1137–1155.

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading, In *Proceedings of the 48th annual meeting of the association for computational linguistics*. Association for Computational Linguistics.

Blythe, H. I., Häikiö, T., Bertam, R., Liversedge, S. P., & Hyönä, J. (2011). Reading disappearing text: Why do children refixate words? *Vision research*, *51*(1), 84–92.

Blythe, H. I., & Joseph, H. S. (2011). Children's eye movements during reading.

Blythe, H. I., Liversedge, S. P., Joseph, H. S., White, S. J., Findlay, J. M., & Rayner, K. (2006). The binocular coordination of eye movements during reading in children and adults. *Vision research*, *46*(22), 3898–3908.

Blythe, H. I., Liversedge, S. P., Joseph, H. S., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision research*, *49*(12), 1583–1591.

Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, *45*(3), 391–411.

Boston, M., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *The Mind Research Repository (beta)*, (1).

Brysbaert, M., & New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior research methods*, *41*(4), 977–990.

Buswell, G. T. (1922). *Fundametal reading habits, a study of their development.* Chicago, Ill. : The University of Chicago.

Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Et al. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175.*

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: Effects of reading skill. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(3), 209.

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 225.

Chen, H.-C. (1992). Reading comprehension in chinese: Implications from character reading times, In *Advances in psychology.* Elsevier.

Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. *Trends in Cognitive Sciences*, *23*(4), 305–317. https://doi.org/https://doi.org/10.1016/j.tics.2019.01.009

Clifton, C., Staub, A., Rayner, K., van Gompel, R., Fischer, M., Murray, W., & Hill, R. (2007). Eye movements: A window on mind and brain. Amsterdam, Netherlands: Elsevier.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., Underwood, P., & Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child development*, *80*(1), 77.

Connor, C. M., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading*, *19*(2), 114–134.

Courrieu, P., Farioli, F., & Grainger, J. (2004). Inverse discrimination time as a perceptual distance for alphabetic characters. *Visual Cognition*, *11*(7), 901–919.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, *42*(5), 621–636.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). Swift: A dynamical model of saccade generation during reading. *Psychological review*, *112*(4), 777.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological review*, *109*(3), 545.

Fang, L., Yang, M., & Lin, D. (2010). *Chinese web 5-gram version 1.* Retrieved March 4, 2019, from https://catalog.ldc.upenn.edu/LDC2010T06

Feng, G. (2002). Share: A stochastic, hierarchical architecture for reading eye-movement.

Feng, G., Miller, K., Shu, H., & Zhang, H. (2009). Orthography and the development of reading processes: An eye-movement study of chinese and english. *Child development*, *80*(3), 720–735.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–674.

Fisher, D., & Shebilske, W. (1985). There is more that meets the eye than the eyemind assumption. *Eye movements and human information processing*, 149–158.

Florida Department of Education, a. (2009). *Florida assessments for instruction in reading administration manual: Grades 3-12* (tech. rep.). State of Florida, Department of Education.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178–210.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 862.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193–202.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). Reading first impact study. final report. ncee 2009-4038. *National Center for Education Evaluation and Regional Assistance*.

Gilbert, L. C. (1959a). Saccadic movements as a factor in visual perception in reading. *Journal of Educational Psychology*, *50*(1), 15.

Gilbert, L. C. (1959b). Speed of processing visual stimuli and its relation to reading. *Journal of Educational Psychology*, *50*(1), 8.

Google research. (2019). Universal sentence encoder.

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of experimental child psychology*, *102*(2), 167–181.

Harvey, W., Teng, M., & Wood, F. (2019). Near-optimal glimpse sequences for improved hard attention neural network training.

Haykin, S. (2010). *Neural networks and learning machines, 3/e*. Pearson Education India.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition, In *Proceedings of the ieee conference on computer vision and pattern recognition*.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 166.

Hongxia, M., Bai, X., Yan, G., & Yao, H. (2014). The number of strokes influences initial landing positions during chinese reading. *Psychological Science*, *37*(4), 809–815.

Hsiao, J. H.-w., & Shillcock, R. (2004). Connectionist modeling of chinese character pronunciation based on foveal splitting, In *Proceedings of the annual meeting of the cognitive science society*.

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, *49*(24), 2948–2959.

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1430.

Hyönä, J., & Pollatsek, A. (1998). Reading finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1612.

Hyönä, J., & Vainio, S. (2001). Reading morphologically complex clause structures in finnish. *European Journal of Cognitive Psychology*, *13*(4), 451–474.

Inhoff, A. W., & Wu, C. (2005). Eye movements and the identification of spatially ambiguous words during chinese sentence reading. *Memory & cognition*, *33*(8), 1345–1356.

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of verbal learning and verbal behavior*, *23*(5), 612–624.

Inhoff, A. W., Lima, S. D., & Carroll, P. J. (1984). Contextual effects on metaphor comprehension in reading. *Memory & Cognition*, *12*(6), 558–567.

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 20.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes, In *Eye guidance in reading and scene perception*. Elsevier.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & psychophysics*, *40*(6), 431–439.

Inhoff, A. W., & Topolski, R. (1994). Use of phonological codes during eye fixations in reading and in on-line and delayed naming tasks. *Journal of Memory and Language*, *33*(5), 689–713.

Jacobs, A. M., Nazir, T. A., & Heller, O. (1989). Perception of lowercase letters in peripheral vision: A discrimination matrix based on saccade latencies. *Perception & Psychophysics*, *46*(1), 95–102.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., Douglas, A., Gersten, R., Newman-Gonchar, R., Dimino, J., Et al. (2009). Effectiveness of selected supplemental reading comprehension interventions: Impacts on a first cohort of fifth-grade students. ncee 2009-4032. *National Center for Education Evaluation and Regional Assistance*.

Joseph, H. S., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, *61*(5), 708–723.

Joseph, H. S., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, *49*(16), 2078–2086.

Joseph, H. S., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, *42*(2), 207–223.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1312.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3). Pearson London.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kajii, N., Nazir, T. A., & Osaka, N. (2001). Eye movement control in reading unspaced text: The case of the japanese script. *Vision Research*, *41*(19), 2503–2510.

Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 68.

Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Skip-thought vectors, In *Advances in neural information processing systems*.

Kliegl, R. (1981). Automated and interactive analysis of eye fixation data in reading. *Behavior Research Methods & Instrumentation*, *13*(2), 115–120.

Kliegl, R., & Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychonomic Bulletin & Review*, *12*(1), 132–138.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European journal of cognitive psychology*, *16*(1-2), 262–284.

Kornuta, T., & Rocki, K. (2016). Utilization of deep reinforcement learning for saccadic-based object visual search.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, *25*(2-3), 259–284.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature Cell Biology*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), https://www.pnas.org/content/106/50/2 21086–21090. https://doi.org/10.1073/pnas.0907664106

Lexile. (2019). The lexile framework for reading.

Li, X., Liu, J., Yu, H., Jiang, X., Liang, F., Wang, Y., & Bai, X. (2019). The decision of saccadic target selection in chinese reading: The interactive way of visual and lexical information. *Studies of Psychology and Behavioral*, *17*(4), 496–503.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in chinese reading. *Journal of Experimental Psychology: General*, *143*(2), 895.

Li, X., Liu, P., & Rayner, K. (2011). Eye movement guidance in chinese reading: Is there a preferred viewing location? *Vision Research*, *51*(10), 1146–1156.

Li, X., Liu, P., & Rayner, K. (2015). Saccade target selection in chinese reading. *Psychonomic Bulletin & Review*, *22*(2), 524–530.

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of chinese words during reading. *Cognitive Psychology*, *58*(4), 525–552.

Lima, S. D., & Inhoff, A. W. (1985). Lexical access during eye fixations in reading: Effects of word-initial letter sequence. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(3), 272.

Liu, Y., Huang, R., Li, Y., & Gao, D. (2017). The word frequency effect on saccade targeting during chinese reading: Evidence from a survival analysis of saccade length. *Frontiers in psychology*, *8*, 116.

Liu, Y., & Reichle, E. D. (n.d.). Using eye movements to understand the leakage of information during chinese reading. *Psychonomic bulletin & review*, 1–7.

Liu, Y., Reichle, E. D., & Li, X. (2015). Parafoveal processing affects outgoing saccade length during the reading of chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1229.

Liu, Y., Reichle, E. D., & Li, X. (2016). The effect of word frequency and parafoveal preview on saccade length during the reading of chinese. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 1008.

Liu, Y., Yu, L., & Reichle, E. D. (2019). The dynamic adjustment of saccades during chinese reading: Evidence from eye movements and simulations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(3), 535.

Liu, Y., Yu, L., Fu, L., Li, W., Duan, Z., & Reichle, E. D. (n.d.). The effects of parafoveal word frequency and segmentation on saccade targeting during chinese reading. *Psychonomic bulletin & review*, 1–10.

Lonigan, C. J., & Phillips, B. M. (2009). Reducing children's risk for later reading disabilities: The role of tier 1 and tier 2 instruction in preschool. *Perspectives on language and literacy*, *35*(5), 21.

Ma, G., & Li, X. (2015). How character complexity modulates eye movement control in chinese reading. *Reading and Writing*, *28*(6), 747–761.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314–324.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological review*, *88*(5), 375.

McConkey, G. (1994). What are" normal" eye movements during reading: Toward a mathematical description. *Eye Movement in Reading*, 315–327.

McConkie, G. W., Kerr, P. W., & Dyre, B. P. (1994). Eye movements in reading. Elsevier Science Inc Tarrytown, NY.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. the location of initial eye fixations on words. *Vision research*, *28*(10), 1107–1118.

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: Ii. frequency of refixating a word. *Perception & Psychophysics*, *46*(3), 245–253.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586.

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, N. R., & Wolff, P. M. (1991). Children's eye movements during reading. *Vision and visual dyslexia*, *13*, 251–262.

McDonald, S. A., Carpenter, R., & Shillcock, R. C. (2005). An anatomically constrained, stochastic model of eye movement control in reading. *Psychological review*, *112*(4), 814.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Et al. (2011). Quantitative analysis of culture using millions of digitized books. *science*, *331*(6014), 176–182.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality, In *Advances in neural information processing systems*.

Norris, D. (2005). How do computational models help us develop better theories. *Twenty-first century psycholinguistics: Four cornerstones*, 331–346.

O'Regan, J. K. (1980). The control of saccade size and fixation duration in reading: The limits of linguistic control. *Perception & Psychophysics*, *28*(2), 112–117.

O'Regan, J. K. (1990). Eye movements and reading. *Reviews of oculomotor research*, *4*, 395–453.

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading, In *Eye movements and visual cognition*. Springer.

O'Regan, J. K. (1992). Solving the" real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *46*(3), 461.

O'Regan, J. K., & Lévy-Schoen. (1987). Eye movement strategy and tactics in word recognition and reading. *The Psychology of Reading*.

O'Regan, J., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. É. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(2), 250.

O'Regan, K. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, *25*(6), 501–509.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation, In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)*.

Perfetti, C. A., & Liu, Y. (2006). *Reading chinese characters: Orthography, phonology, meaning, and the lexical constituency model.* na.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on chinese for general theories of reading. *Psychological review*, *112*(1), 43.

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 148.

Pollatsek, A., & Rayner, K. (1982). Eye movement control in reading: The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(6), 817.

Pynte, J., New, B., & Kennedy, A. (2008). On-line contextual influences during reading normal text: A multiple-regression analysis. *Vision research*, *48*(21), 2172–2183.

Pynte, J., New, B., & Kennedy, A. (2009). On-line contextual influences during reading normal text: The role of nouns, verbs and adjectives. *Vision research*, *49*(5), 544–552.

R Development Core Team. (2018). *R: A language and environment for statistical computing* [ISBN 3-900051-07-0]. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological research*, *72*(6), 675–688.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. *European Journal of Cognitive Psychology*, *16*(1-2), 3–26.

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology*, *66*(3), 429–452.

Radach, R., Schmitten, C., Glover, L., & Huestegge, L. (2009). How children read for comprehension: Eye movements in developing readers. *Beyond decoding: The biological and behavioral foundations of reading comprehension*, 75–106.

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *49*(2), 151.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological bulletin*, *85*(3), 618.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*(1), 21–30.

Rayner, K. (1985). The role of eye movements in learning to read and reading disability. *Remedial and Special Education*, *6*(6), 53–60.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of experimental child psychology*, *41*(2), 211–236.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of eye movement research*, *2*(5), 1.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition*, *14*(3), 191–201.

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, *38*(8), 1129–1144.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological science in the public interest*, *2*(2), 31–74.

Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability on the eye movements of chinese readers. *Psychonomic Bulletin & Review*, *12*(6), 1089–1093.

Rayner, K., Li, X., & Pollatsek, A. (2007). Extending the e-z reader model of eye movement control to chinese readers. *Cognitive science*, *31*(6), 1021–1033.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision research*, *16*(8), 829–837.

Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *The Quarterly journal of experimental psychology*, *33*(4), 351–373.

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 912.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509.

Reichle, E. D. (2015). Computational models of reading: A primer. *Language and Linguistics Compass*, *9*(7), 271–284.

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S., White, S. J., & Rayner, K. (2013). Using ez reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*(2), 110–149.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, *105*(1), 125.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The ez reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, *26*(4), 445–476.

Reilly, R. (1993). A connectionist framework for modeling eye-movement control in reading.

Reilly, R. G., & O'Regan, J. K. (1998). Eye movement control during reading: A simulation of some word-targeting strategies. *Vision Research*, *38*(2), 303–317.

Reilly, R. G., & Radach, R. (2002). Glenmore: An interactive activation model of eye movement control in reading, In *Proceedings of the 9th international conference on neural information processing, 2002. iconip'02.* IEEE.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*(1), 34–55.

Res, V. (1996). Reading unspaced text is not easy: Comments on the implications of epelboim et al.'s (1994) study for models of eye movement control in reading. *Vision Res*, *36*(3), 461–470.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://www.jstatsoft.org/v48/i02/

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Ii. the contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60.

Salthouse, T. A., Ellis, C. L., Diener, D. C., & Somberg, B. L. (1981). Stimulus processing during eye fixations. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(3), 611.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once) comprehension is supported by regressions during reading. *Psychological science*, *25*(6), 1218–1226.

Shen, D.-L., Bai, X.-J., Zang, C.-L., Yan, G.-L., Feng, B.-C., & Fan, X.-H. (2010). Effect of word segmentation on beginners' reading: Evidence from eye movements. *Acta Psychologica Sinica*, *42*(02), 159–172.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). Ob1-reader: A model of word recognition and eye movements in text reading. *Psychological review*, *125*(6), 969.

Song, R., Lin, M., & Ge, S.-l. (2008). Similarity calculation of chinese character glyph and its application in computer aided proofreading system. *Journal of Chinese Computer Systems*, *129*(10).

Spragins, A. B., Lefton, L. A., & Fisher, D. F. (1976). Eye movements while reading and searching spatially transformed text: A developmental examination. *Memory & Cognition*, *4*(1), 36–42.

Stenner, A., Burdick, H., Sanford, E., & Burdick, D. (2007). The lexile framework for reading technical report. *Durham, NC: Metametrics*, 6.

Subramanian, S., Trischler, A., Bengio, Y., & Pal, C. J. (2018). Learning general purpose distributed sentence representations via large scale multi-task learning. *arXiv preprint arXiv:1804.00079*.

Sun, J. (2013). Jieba chinese text segmentation. GitHub, Inc.

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 761.

Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in chinese character recognition. *Journal of memory and Language*, *40*(4), 498–519.

Taylor, S. E., Frackenpohl, H., & Pettee, J. L. (1960). *Grade level norms for the components of the fundamental reading skill*. Educational Developmental Laboratories.

Tsai, J.-L., & McConkie, G. W. (2003). Where do chinese readers send their eyes?, In *The mind's eye*. Elsevier.

Van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. (2012). How do children deal with inconsistencies in text? an eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*(7), 1665–1690.

Vauras, M., Hyönä, J., & Niemi, P. (1992). Comprehending coherent and incoherent texts: Evidence from eye movement patterns and recall performance. *Journal of Research in Reading*, *15*(1), 39–54.

Vitu, F. (1991). The existence of a center of gravity effect during reading. *Vision Research*, *31*(7-8), 1289–1313.

Vitu, F., McConkie, G. W., & Zola, D. (1998). About regressive saccades in reading and their relation to word identification, In *Eye guidance in reading and scene perception*. Elsevier.

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception & Psychophysics*, *57*(3), 352–364.

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & psychophysics*, *47*(6), 583–600.

Vorstius, C., Radach, R., & Lonigan, C. J. (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition*, *22*(3-4), 458–485.

Vorstius, C., Radach, R., Mayer, M. B., & Lonigan, C. J. (2013). Monitoring local comprehension monitoring in sentence reading. *School Psychology Review*, *42*(2), 191.

Wang, D., & Xiong, S. (2013). New algorithm for similarity calculation of chinese character glyph. *Appl. Res. Comput*, *30*(8), 2395–2397.

Wang, X., & Tao, B. (1993). *Primary school literacy test and evaluation scale*. Shanghai education press.

Wang, Y., Zhao, B., Chen, M., Li, X., Yan, G., & Bai, X. (2018). Influence of the frequency of fixated words and the number of strokes of parafoveal words on saccadic target selection in chinese reading. *Acta Psychologica Sinica*, *50*(12), 1336–1345.

Wei, W., Li, X., & Pollatsek, A. (2013). Word properties of a fixated region affect outgoing saccade length in chinese reading. *Vision Research*, *80*, 1–6.

White, S. J., Lantz, L. M., & Paterson, K. B. (2017). Spontaneous rereading within sentences: Eye movement control and visual sampling. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 395.

Wieting, J., Bansal, M., Gimpel, K., & Livescu, K. (2015). Towards universal paraphrastic sentence embeddings. *arXiv preprint arXiv:1511.08198*.

Xing, H., Shu, H., & Li, P. (2002). Self-organizing connectionist model of character acquisition in chinese, In *Proceedings of the annual meeting of the cognitive science society*.

Xu, Y., & Vaziri-Pashkam, M. (2020). Limited correspondence in visual representation between the human brain and convolutional neural networks. *bioRxiv*, https://www.biorxiv.org/content/early/2020/03/14/2020.03.12.989376.full.pdf. https://doi.org/10.1101/2020.03.12.989376

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of chinese readers. *British Journal of Psychology*, *97*(2), 259–268.

Yan, M., Kliegl, R., Richter, E. M., Nuthmann, A., & Shu, H. (2010). Flexible saccade-target selection in chinese reading. *The Quarterly Journal of Experimental Psychology*, *63*(4), 705–725.

Yang, H.-M., & McConkie, G. W. (1999). Reading chinese: Some basic eye-movement characteristics (J. Wang, H.-C. Chen, R. Radach, & A. Inhoff, Eds.). In J. Wang, H.-C. Chen, R. Radach, & A. Inhoff (Eds.), *Reading chinese script*. Psychology Press.

Yang, J., Zevin, J. D., Shu, H., McCandliss, B. D., & Li, P. (2006). A “triangle model” of chinese reading, In *28th annual conference of the cognitive science society, mahwah, nj*. Citeseer.

Yen, M.-H., Tsai, J.-L., Tzeng, O. J., & Hung, D. L. (2008). Eye movements and parafoveal word processing in reading chinese. *Memory & Cognition*, *36*(5), 1033–1045.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?

Zang, C., Zhang, M., Guo, X., Liu, J., Yan, G., & Bai, X. (2012). Several effects on chinese lexical processing: Evidence from eye movements. *Advances in Psychological Science*, *20*(9), 1382–1392.

Zang, C., Liang, F., Bai, X., Yan, G., & Liversedge, S. P. (2013). Interword spacing and landing position effects during chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 720.

Zhang, H., & Wang, X. (1989). Standardization research on raven's standard progressive matrices in china. *Acta Psychologica Sinica*, *21*(02), 3. http://journal.psych.ac.cn/xlxb/CN/abstract/article_758.shtml

Zhang, X., & Yan, G. (2005). Age of acquisition, familiarity, concreteness and frequency effects of words on college student's eye movements in reading. *Studies of Psychology and Behavior*, *3*.

Appendix A

Code of a small-scale extension to the Glenmore model of eye movement control in Chinese reading

```
#Mini-implementation of Glenmore to model the interaction between
#character and word-driven saliency in Chinese reading

# install.packages('gifski')
# devtools::install_github('thomasp85/gganimate')
# devtools::install_github("kassambara/ggpubr")
#install.packages("DEoptim")
#library(ggpubr)
library(tidyverse)
library(ggplot2)
library(gganimate)
library(gifski)
library(DEoptim)
renderer=gifski_renderer()

rm (list = ls()) # reset workspace
gaussian <- function(x, mu, sig, height=1.0) {
  if (height==0.0) {
    height = 1.0/(sig * sqrt(2.0 * pi))
  }
  diff = (((x-mu)*(x-mu))/(2 * sig*sig))
  return (height * exp(-diff))
}
# allows a vector of distribution heights
gaussian2 <- function(x, mu, sig, height) {
  diff = (((x-mu)*(x-mu))/(2 * sig*sig))
  return (height * exp(-diff))
```

```
}
sigmoidPlus <- function (x, offset=0, temp=1){
  return (1./(1. + exp(-(x-offset)/temp)))
}
# Sentence 1
# 一个/女孩子/整天/不务正业
# HF word
sent_1 <- list(
  t(matrix(c(                                # char to word matrix
    c(1,1,0,0,0,0,0,0,0,0,0),
    c(0,0,1,1,1,0,0,0,0,0,0),
    c(0,0,0,0,0,1,1,0,0,0,0),
    c(0,0,0,0,0,0,0,1,1,1,1)),
    nrow=11, ncol=4)),
  c(1,3,3,9,3,16,4,3,5,5,5),                 # strokes
  c(5.1,2.8,3,1),                            # word freq
  c(4, 11),                                  # n_word n_chars
   # parafoveal scaling
  c(1, 0.992217938, 0.969233234, 0.932102492, 0.882496903, 0.822577562,
    0.754839602, 0.681940751, 0.60653066, 0.531095991, 0.457833362),
  # chars
  c('一','个','女','孩','子','整','天','不','务','正','业'),
  # words
  c('一个','女孩子','整天','不务正业'),
  # desired saliency
  c(0.80, 0.83, 0.84, 0.90, 0.83, 0.96, 0.84, 0.83, 0.85, 0.85, 0.85))

# Sentence 2
# 一个/二流子/整天/不务正业
# LF word
sent_2 <- list(
  t(matrix(c(
    c(1,1,0,0,0,0,0,0,0,0,0),
    c(0,0,1,1,1,0,0,0,0,0,0),
    c(0,0,0,0,0,1,1,0,0,0,0),
    c(0,0,0,0,0,0,0,1,1,1,1)),
    nrow=11, ncol=4)),
  c(1,3,2,10,3,16,4,3,5,5,5),
  c(5.1,0.8,3,1),
  c(4, 11),
  c(1, 0.992217938, 0.969233234, 0.932102492, 0.882496903, 0.822577562,
    0.754839602, 0.681940751, 0.60653066, 0.531095991, 0.457833362),
  c('一','个','二','流','子','整','天','不','务','正','业'),
  c('一个','二流子','整天','不务正业'),
  c(0.80, 0.83, 0.84, 0.96, 0.83, 0.90, 0.84, 0.83, 0.85, 0.85, 0.85))
```

```
params <- c(0.1077204,  0.1740187,  0.2333088, -2.7671077,
            0.5941790,  0.1555598 )
param_names <- c('R_2_C', 'C_2_W', 'C_2_S', 'W_2_C', 'W_2_W', 'S_2_S')
names(params) <- param_names

# generalised run_sim function
run_sim <- function(iter, params, sent) {
  # initialise various variables
  #
  char_log <-     sal_log <- wrd_log <- tap_log <-  NULL
  max_strokes <-  23
  n_word <-       sent[[4]][1]
  n_char <-       sent[[4]][2]
  chars_accum <-  rep(0, n_char)
  chars_out <-    rep(0, n_char)
  sal_out <-      rep(0, n_char)
  sal_accum <-    rep(0, n_char)
  w_inp <-        rep(0, n_word)
  w_out <-        rep(0, n_word)
  # turn on or off feedback to chars from words (1=on; 0=off)
  w_tap <-        rep(1, n_word)

  # Extract params
  R_2_C <- params['R_2_C']
  C_2_W <- params['C_2_W']
  C_2_S <- params['C_2_S']
  W_2_C <- params['W_2_C']
  W_2_W <- params['W_2_W']
  S_2_S <- params['S_2_S']

  # Word identification threshold and steepness
  W_thresh <- 0.9
  W_temp <- 5.0

  for (i in 1:iter) {
    #input to chars from retina
    retina <- sent[[2]] * sent[[5]]
    chars_accum <- chars_accum + retina * R_2_C

    # input to chars from words
    chars_accum <- chars_accum + (w_out * W_2_C * w_tap) %*% sent[[1]]

    # input to saliency from chars and self
    sal_accum <- sal_accum + chars_out * C_2_S
```

```
    sal_accum <- sal_accum + sal_out * S_2_S

    # input to words from chars
    w_inp <- w_inp + w_tap * ((chars_out * C_2_W) %*% t(sent[[1]]))

    # input to words from word freq
    w_inp <- w_inp + sent[[3]] * W_2_W

    # output
    chars_out <- gaussian2(chars_accum, 50, 50, height=sent[[2]]/max_strokes)
    sal_out <- gaussian (sal_accum, 50, 50)
    w_out <- sigmoidPlus(w_inp, offset=50, temp=W_temp)

    for(j in 1:n_word) {
      # Lock word if over threshold
      if (w_tap[j]==1 & w_out[j]>W_thresh) w_tap[j] <- 0
    }

    # Build data logs
    char_log <- rbind(char_log, chars_out)
    sal_log <-  rbind(sal_log, sal_out)
    wrd_log <-  rbind(wrd_log, w_out)
    tap_log <-  rbind(tap_log, w_tap)
  }
  # Return combined dataframe
  return(as.data.frame(cbind(sal_log, tap_log, wrd_log, char_log),
                       row.names=1:iter))
}



# run_sim for use with DEoptim, so it returns RMSE between
#saliency and target vectors

run_sim_RMSE <- function(iter, params, sent) {
  # name params
  param_names <- c("R_2_C", "C_2_W", "C_2_S", "W_2_C", "W_2_W", "S_2_S")
  names(params) <- param_names

  # initialise various variables
  max_strokes <-  23
  n_word <-       sent[[4]][1]
  n_char <-       sent[[4]][2]
  chars_accum <-  rep(0, n_char)
  chars_out <-    rep(0, n_char)
```

```
sal_out <-      rep(0, n_char)
sal_accum <-    rep(0, n_char)
w_inp <-        rep(0, n_word)
w_out <-        rep(0, n_word)
# turn on or off feedback to chars from words (1=on; 0=off)
w_tap <-        rep(1, n_word)

# Extract params
R_2_C <- params['R_2_C']
C_2_W <- params['C_2_W']
C_2_S <- params['C_2_S']
W_2_C <- params['W_2_C']
W_2_W <- params['W_2_W']
S_2_S <- params['S_2_S']

# Word identification threshold and steepness
W_thresh <- 0.9
W_temp <- 5.0

for (i in 1:iter) {
  #input to chars from retina
  retina <- sent[[2]] * sent[[5]]
  chars_accum <- chars_accum + retina * R_2_C

  # input to chars from words
  chars_accum <- chars_accum + (w_out * W_2_C * w_tap) %*% sent[[1]]

  # input to saliency from chars and self
  sal_accum <- sal_accum + chars_out * C_2_S
  sal_accum <- sal_accum + sal_out * S_2_S

  # input to words from chars
  w_inp <- w_inp + w_tap * ((chars_out * C_2_W) %*% t(sent[[1]]))

  # input to words from word freq
  w_inp <- w_inp + sent[[3]] * W_2_W

  # output
  chars_out <- gaussian2(chars_accum, 50, 50, height=sent[[2]]/max_strokes)
  sal_out <- gaussian (sal_accum, 50, 50)
  w_out <- sigmoidPlus(w_inp, offset=50, temp=W_temp)

  for(j in 1:n_word) {
    # Lock word if over threshold
    if (w_tap[j]==1 & w_out[j]>W_thresh) w_tap[j] <- 0
```

```
    }
  }
  # Return RMSE between target and sal output
  return(sqrt(sum((sal_out-sent[[8]])^2)))
}

# Run multiple sentences through the optimiser
run_sent_batch_DE <- function (params, sents) {
  acc_rmse <- 0
  for (sent in sents) {
    acc_rmse <- acc_rmse + run_sim_RMSE(150, params, sent)
  }
  return(acc_rmse)
}

set.seed(55912072) # lowest 0.08230607
lower <- c(0,0,0,-3,0,0)
upper <- c(3,3,3,0,3,3)
sents<- list(sent_1, sent_2)

DEoptim (run_sent_batch_DE, lower=lower, upper=upper,
         control=DEoptim.control(itermax=5000), sents)

# First sentence
s_s1.df <- run_sim (220, params, sent_1)
W_thresh <- 0.9

# character animation
char_s1.df <- s_s1.df %>%
  rownames_to_column (var="time") %>%
  select(time, V20:V30) %>%
  pivot_longer (-time, names_to="char",
                values_to="activation", names_prefix="V") %>%
  mutate(time = as.integer(time), char = as.integer(char)-19)

char_s1.p <- ggplot(char_s1.df, aes(x= char, y=activation)) +
  geom_line() +
  scale_x_discrete(limits=1:sent_1[[4]][2],
                   labels = sent_1[[6]]) +
  ylab("activation") +
  xlab("character")

char_s1.p <- char_s1.p + transition_time(time) +
  ggtitle("Character activation (time: {frame_time})")
```

```
# Saliency animation
sal_s1.df <- s_s1.df %>%
  rownames_to_column (var="time") %>%
  select(time, V1:V11) %>%
  pivot_longer (-time, names_to="char",
                values_to="sal", names_prefix="V") %>%
  mutate(time = as.integer(time), char = as.integer(char))

sal_s1.p <- ggplot(sal_s1.df, aes(x=char, y=sal)) +
  geom_line() +
  scale_x_discrete(limits=1:sent_1[[4]][2],
                   labels = sent_1[[6]]) +
  ylab("saliency") +
  xlab("character")

sal_s1.p <- sal_s1.p + transition_time(time) +
  ggtitle("Saliency (time: {frame_time})")

# Word tap
word_tap_s1.df <- s_s1.df %>%
  rownames_to_column(var="time") %>%
  select(time, V12:V15) %>%
  pivot_longer (-time, names_to="word",
                values_to="tap", names_prefix="V") %>%
  mutate(time = as.integer(time),
   word = as.factor(as.integer(word)-11),
         tap = as.factor(ifelse(tap==1, "on", "off")))

# Word activation animation
word_s1.df <- s_s1.df %>%
  rownames_to_column(var="time") %>%
  select(time, V16:V19) %>%
  pivot_longer (-time, names_to="word",
                values_to="activation", names_prefix="V") %>%
  mutate(time = as.integer(time), word = as.factor(as.integer(word)-15)) %>%
  left_join (word_tap_s1.df, by = c("time", "word"))

word_s1.p <- ggplot(word_s1.df, aes(x=word, y=activation)) +
  geom_col(aes(fill=tap)) +
  geom_hline(yintercept=W_thresh, colour="blue") +
  annotate("text", x = 5, y = W_thresh+0.05,
         label = "identification\nthreshold") +
  scale_x_discrete(name="word", limits=1:(sent_1[[4]][1]+1),
                   labels = c(sent_1[[7]], ' ')) +
  scale_y_continuous(name="activation", limits = c(0,1))
```

```
word_s1.p <- word_s1.p + transition_time(time) +
  ggtitle("Word activation (time: {frame_time})")


# Save the graphs
animate(char_s1.p, fps=5, nframes=220)
anim_save("char_s1.gif", device = cairo_pdf)
animate(sal_s1.p, fps=5, nframes=220)
anim_save("sal_s1.gif", device = cairo_pdf)
animate(word_s1.p, fps=5, nframes=220)
anim_save("word_s1.gif", device = cairo_pdf)



### Second sentence
s_s2.df <- run_sim (220, params, sent_2)
W_thresh <- 0.9

# character animation
char_s2.df <- s_s2.df %>%
  rownames_to_column (var="time") %>%
  select(time, V20:V30) %>%
  pivot_longer (-time, names_to="char",
                values_to="activation", names_prefix="V") %>%
  mutate(time = as.integer(time), char = as.integer(char)-19)

char_s2.p <- ggplot(char_s2.df, aes(x= char, y=activation)) +
  geom_line() +
  scale_x_discrete(limits=1:sent_1[[4]][2],
                   labels = sent_1[[6]]) +
  ylab("activation") +
  xlab("character")

char_s2.p <- char_s2.p + transition_time(time) +
  ggtitle("Character activation (time: {frame_time})")

# Saliency animation
sal_s2.df <- s_s2.df %>%
  rownames_to_column (var="time") %>%
  select(time, V1:V11) %>%
  pivot_longer (-time, names_to="char",
                values_to="sal", names_prefix="V") %>%
  mutate(time = as.integer(time), char = as.integer(char))
```

```
sal_s2.p <- ggplot(sal_s2.df, aes(x=char, y=sal)) +
  geom_line() +
  scale_x_discrete(limits=1:sent_1[[4]][2],
                   labels = sent_1[[6]]) +
  ylab("saliency") +
  xlab("character")

sal_s2.p <- sal_s2.p + transition_time(time) +
  ggtitle("Saliency (time: {frame_time})")

# Word tap
word_tap_s2.df <- s_s2.df %>%
  rownames_to_column(var="time") %>%
  select(time, V12:V15) %>%
  pivot_longer (-time, names_to="word",
                values_to="tap", names_prefix="V") %>%
  mutate(time = as.integer(time), word = as.factor(as.integer(word)-11),
         tap = as.factor(ifelse(tap==1, "on", "off")))

# Word activation animation
word_s2.df <- s_s2.df %>%
  rownames_to_column(var="time") %>%
  select(time, V16:V19) %>%
  pivot_longer (-time, names_to="word",
                values_to="activation", names_prefix="V") %>%
  mutate(time = as.integer(time),
    word = as.factor(as.integer(word)-15)) %>%
  left_join (word_tap_s2.df, by = c("time", "word"))

word_s2.p <- ggplot(word_s2.df, aes(x=word, y=activation)) +
  geom_col(aes(fill=tap)) +
  geom_hline(yintercept=W_thresh, colour="blue") +
  annotate("text", x = 5, y = W_thresh+0.05,
    label = "identification\nthreshold") +
  scale_x_discrete(name="word", limits=1:(sent_1[[4]][1]+1),
                   labels = sent_1[[7]]) +
  scale_y_continuous(name="activation", limits = c(0,1))

word_s2.p <- word_s2.p + transition_time(time) +
  ggtitle("Word activation (time: {frame_time})")

#Animate it
animate(char_s2.p, fps=5, nframes=220)
anim_save("char_s2.gif")
animate(sal_s2.p, fps=5, nframes=220)
```

```
anim_save("sal_s2.gif")
animate(word_s2.p, fps=5, nframes=220)
anim_save("word_s2.gif")

sal_s1_samp.p <- sal_s1.df %>%
  filter(time %in% c(10,50,100,200)) %>%
  mutate(sent=1) %>%
  ggplot(aes(x=char, y=sal, group=time, colour=time)) + geom_line()

print(sal_s1_samp.p)

sal_s2_samp.p <- sal_s2.df %>%
  filter(time %in% c(10,50,100,200)) %>%
  ggplot(aes(x=char, y=sal, group=time, colour=time)) + geom_line()

print(sal_s2_samp.p)


sal_samp.df <-
  rbind (
    sal_s1.df %>%
      filter(time %in% c(10,50,100,200)) %>%
      mutate(sent=1),
    sal_s2.df %>%
      filter(time %in% c(10,50,100,200)) %>%
      mutate(sent=2)
  )

sal_samp.p <-
  sal_samp.df %>%
  ggplot(aes(x=char, y=sal, group=time, colour=time)) +
  geom_line() +
  facet_wrap (~sent)
```