FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

An Eye-tracking Corpus study of Chinese-English bilingual reading

Longjiao Sui

Supervisor: Prof. Dr. Nicolas Dirix
Co-supervisor: Prof. Dr. Evy Woumans

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Doctor of Psychology

Academic year 2022–2023

GHENT
UNIVERSITY

CONTENTS

ACKNOWLEDGEMENTS

If a Ph.D. program is a novel, mine could be *The Mysterious Affair at Styles*, as it was full of unexpected twists and turns. I started my Ph.D. in 2019 and then experienced the COVID pandemic. Due to the pandemic, I faced challenges in collecting data from interpreters, which led me to change my Ph.D. project from language production to word recognition. In addition, the lockdown further complicated matters by restricting access to the laboratory and posing challenges in recruiting participants for the new project. I also experienced changing principal supervisor because he took up a position elsewhere.

Despite these challenges, my promoter, Prof. Nicolas Dirix, provided me with steadfast support, enabling me to overcome challenges and complete my studies. I would like to express my sincere gratitude to him for his invaluable guidance throughout my research process and his understanding of the difficulties encountered by an international student (e.g., language). Without his continuous support and encouragement, I would not have been able to overcome the challenges and achieve my academic goals.

I am also grateful for Prof. Wouter Duyck, my former supervisor, who has provided invaluable support to my research even after taking up the position at NVAO. In addition, I would like to express my gratitude to my doctoral guidance committee members, Prof. Evy Woumans, Prof. Bart Defrancq, and Prof. Arnaud Szmalec. I also thank Prof. Marc Brysbaert and Rob Hartsuiker and all the professors who have helped me. Their academic support and guidance have greatly benefited me. Furthermore, I am very grateful to the participants who took part in my experiments and supported me during the pandemic.

CHAPTER 1. INTRODUCTION

Reading is often considered one of the most important and indispensable ways to access information. People can read to gain knowledge from books or websites; they can get timely information from schedules at airports or train stations; and they can use it as a primary way to communicate with others, such as by reading emails and text messages. Visual word processing is undoubtedly vital, and its processes have been extensively studied in alphabetic languages accordingly. Multiple methodologies have been used in word recognition research, including lexical decision tasks and eye-tracking measures as well as event-related potentials (ERPs, e.g., Sereno & Rayner, 2003) and neuroimaging (e.g., Fiez & Petersen, 1998).

The lexical decision task is a well-established, widely used approach for studying word recognition, with approximately 3,800 publications on Web of Science[1]. In the lexical decision task, participants were asked to decide whether the string of letters presented on the screen was an existing word by pressing a corresponding button. The reaction times for the subject to respond to it were measured. This paradigm laid the foundation for understanding word recognition (e.g., Duyck et al., 2008; Brysbaert et al., 2016). It has advantages as it is comparatively easy to collect data due to not having excessive equipment requirements. However, a growing number of researchers have questioned whether the responses in determining the authenticity of a word can reflect the natural reading performance (e.g., Kuperman et al., 2013). Indeed, research has shown that controlled experiments cannot fully reflect word recognition processing in natural reading (see Dirix et al., 2019; Kuperman et al., 2013). Consequently, eye-tracking, the latter research approach, has received increasing attention in recent years.

In eye-tracking, participants are asked to silently read the presented text while recording highly accurate temporal and spatial information about their eye movements through the equipment. The presented text could either be an isolated sentence or a paragraph. This type of reading is referred to as “natural reading”, as the primary task of the participant is to read at their own pace. In addition, reading a chunk of words or paragraphs is much closer to how people read in daily life than reading individual words. Since reading in context involves numerous factors that may influence the reading of subsequent words, such as syntax, only the results observed in sentence reading can reflect word recognition in daily life,

[1] measured 30 April 2023

not in isolated word conditions. Hence, eye-tracking research allows an unbiased investigation of the word recognition process in natural reading. This approach has other advantages, such as allowing the study of multiple measures (i.e., fixations, saccades, skips, regressions; Rayner, 2009) as well as the stages in word recognition (i.e., early- and late-stage; Boston et al., 2008).

There are a few types of eye movements, including fixations, saccades, skips, and regressions. Fixations refer to the duration of fixating on a word; saccades refer to the movements from one fixation point to another; skips refer to skipping a word as the eyes move and fixating on subsequent ones; and regressions refer to moving back to where it has already read. The different fixation measures are believed to reflect different stages of word recognition (e.g., Clifton et al., 2007). The first fixation duration (FFD, the duration to fixate a word for the first time in the first-pass reading) and gaze duration (GD, all fixation durations on a word before fixating the next (right-side) one in the first-pass reading) are referred to as “early” measures and are considered to reflect the “early” stages of word recognition. Total reading time (TRT, all fixation durations on a word) is a "late" measure and is assumed to reflect the "late" stage of word recognition. Based on the application of this method, some eye-movement control models of reading, such as the E-Z reader model (e.g., Reichle et al., 1998, 2003), have been proposed.

In general, the methodologies described above (e.g., lexical decision tasks and eye-tracking studies) establish the existing knowledge of word recognition. Notably, however, most existing well-known phenomena (e.g., frequency and word length effects) are based on observations in alphabetic languages, where words are composed of letters and are separated by spaces in sentences (e.g., Rayner et al., 1996, 2011; Slattery et al., 2007). Hence, relevant models of visual word recognition are based on findings in alphabetic languages (e.g., the learning hypothesis, Morton, 1970; the lexical entrenchment hypothesis, Diependaele et al., 2013 for frequency effect; DRC model, Coltheart et al., 2001; E-Z Reader model, Reichle et al., 2003 for eye movement control in word recognition). However, an important question is whether these phenomena and models hold for structurally different languages, such as Chinese.

Chinese writing system

The Chinese writing system differs markedly from that of alphabetic languages in terms of orthography, morphology, sentence structure, and grammar. Chinese characters are equally spaced box-shaped strings composed of strokes. The more strokes a character has, the more visual complexity the character becomes. The degree of visual complexity seems to impact word recognition, as characters with high complexity take longer to read (e.g., Liversedge et al., 2014). In addition, each Chinese character has its pronunciation and carries semantic information. The visually distinct characters are the constituents of Chinese words and can appear in various positions in different words. A limited number of characters can comprise over 56,000 words (Li & Su, 2022). In Chinese, about 95% of words are one- and two-character words (Li & Pollatsek, 2020), so their word length is much shorter on average than in English.

What further differentiates Chinese from alphabetic languages is that there is no visual demarcation from each other in a sentence. In this case, a character in a sentence may be a one-character word or part of a multi-character word with the preceding or following characters (also see Chapter 3). Furthermore, the concept of words was not introduced until the twentieth century, and characters remain the primary units in Chinese dictionaries until now. For these reasons, many researchers contend that Chinese characters, rather than words as in alphabetic languages, serve as the fundamental processing unit in Chinese reading (e.g., Hoosain, 1992). In fact, some Chinese word recognition models suggest that characters are the basic reading unit (e.g., Hsiao & Shillcock, 2004, 2005). However, a substantial body of research has shown that words are the primary processing unit in Chinese sentence reading (Li et al., 2013, 2014; Bai et al., 2008). Nevertheless, character frequency still appears to influence word recognition (e.g., Li et al., 2014; Yan et al., 2006; Yu et al., 2021). In addition, since words are a primary unit in Chinese reading (e.g., Li et al., 2014) and lack visual separation in sentences, a language-specific process of segmenting words is necessary for reading Chinese sentences. As a result, some models indicate that words are an important processing unit for Chinese reading (e.g., Rayner et al., 2007). Some of these models also suggest that character properties may still play a role, despite not being the fundamental reading unit (e.g., Li & Pollatsek, 2020).

In recent decades, there has been an increasing number of studies investigating Chinese reading performance, yet the quantity and research scope are still limited. Existing research on Chinese reading has shown that phenomena such as frequency effects observed in alphabetic languages were also found in Chinese reading (e.g., Ma et al., 2015; Li et al., 2014; Yan et al., 2006; Zhou et al., 2018 for frequency effect, Zang et al., 2018 for word length effect; Rayner et al., 2005 for predictability, etc.). What’s more, due to the particular features of Chinese orthography, there are language-specific factors, such as complexity, involved in Chinese reading. Of course, whether language-specific factors affect the observed phenomena, resulting in different performance compared to that found in alphabetic languages, like frequency effect, remains to be confirmed. Nevertheless, there could be numerous disparities between native language reading in alphabetic languages and Chinese. Hence, investigating Chinese reading performance is of theoretical importance, as it is crucial for identifying the reading process specific to Chinese or universal across languages and for understanding the cognitive and linguistic processes it involves. With the increasing number of bilinguals worldwide, another question is how bilinguals with different writing systems in the two languages, i.e., Chinese-English bilinguals, read. To investigate this issue, we will delve into bilingualism in the next section.

Bilingualism

Over the past few decades, numerous studies have investigated the effects of bi- and multilingualism on cognition and language processing (e.g., Assche et al., 2012; Dijkstra, 2005; Dijkstra & Van Heuven, 2002 in word recognition; Griffin & Bock, 1998; Costa et al., 1999 in language production), with the steadily increasing number of populations speaking more than one language. A common consensus is that all known languages are activated, supported by substantial evidence (e.g., Dijkstraet al., 1999; Duyck, 2005; Jared & Kroll, 2001; Van Hell & Dijkstra, 2002). In addition, the co-activation of languages affects each other (e.g., Brysbaert & Duyck, 2010), with an example of the cognate effect (e.g., Van Assche et al., 2009; for an alternative account that effects like cognates are due to the increased language exposure, see Brysbaert et al., 2017). Nevertheless, the language processing of bilinguals is undoubtedly somewhat different from that of monolinguals. Therefore, it is crucial to understand the language processing of bilinguals.

Indeed, such differences occur for example in word recognition. When monolinguals read a word, the target word is activated with a group of orthographic neighbors (words that are orthographically similar to the target word, differing only by one letter, e.g., Coltheart et al., 1977, cited from Dirix et al., 2017). As for bilinguals, in addition to the within-language neighbors, the cross-language neighbors from the non-target language are also activated, e.g., Dirix et al., 2017; Whitford & Titone, 2019). As a result, for bilinguals the target word is selected from a group of co-activated candidates, within and across languages, which in comparison to monolinguals causes additional interference or facilitation (dependent on the language) in language processing (Dirix et al., 2017, van Heuven et al., 1998). However, Chinese and English are different writing systems whose orthographies are dissimilar. Hence, it raises an interesting question of whether the word recognition for Chinese-English bilinguals differs from that of alphabet-language bilinguals, as there are no cross-linguistic orthographic neighbors between Chinese and English. Can the recognition of words when reading in a language be affected by a distinct non-target language?

Exploring the reading performance of Chinese-English bilinguals is theoretically essential. It contributes to verifying and refining existing theories and models that aim to explain universal cognitive processes and behaviors of reading (e.g., (bilingual) reading models or frequency effect theories) but are based on data from same-alphabet language bilinguals. Such research can provide valuable insights into whether reading behavior may differ across languages or language pairs. Moreover, it is also empirically important, given the considerable number of Chinese-English bilinguals worldwide. Understanding how these bilinguals process the two entirely different languages during reading and whether there are any potential adverse effects is crucial for enhancing their reading efficiency, particularly in the second language.

This dissertation

Apparently, eye-movement studies are an effective method for investigating natural reading performance and are a closer approach to reading in everyday life than experimental studies (i.e., lexical decision tasks). However, exploring the eye-movement performance of bilinguals in both first- and second-language reading is quite a challenge. Firstly, collecting high-quality data with a substantial size is a challenge. Data with high statistical power can detect reliable phenomena, even the minimal effects (also see Brysbaert, 2019). Yet, the collection of large amounts of eye-movement data is not only effortful, time-consuming, and costly but also requires specialized equipment.

Secondly, preparing the natural reading material requires considerable effort. Currently, most research on Chinese reading uses reading material especially constructed for research purposes, that for example only includes words with widely accepted boundaries to avoid ambiguity. Indeed, such manipulation can avoid the difficulties and efforts involved in segmenting words when preparing word-level interest areas for exporting eye movement measures. However, the use of controlled materials may result in deviations from the natural reading performance (also see Dirix et al., 2019). It is especially true for Chinese reading, where Chinese words are the primary units in sentence reading, and there are no visual delimitations between them in sentences. Therefore, word segmentation is necessary and may affect word recognition (e.g., Li & Pollatsek, 2020). In addition, such materials lack naturalism, which limits the exploration of processing ambiguous words (for research on ambiguous Chinese words, see Ma et al., 2014).

What's more, many existing experiments and corpora only explore target words presented in isolated, unrelated low-constraint sentences, and some even use the same sentence frame where only the target word differs (e.g., Cui et al., 2021; Ma et al., 2015). While it is true that some of these ways may require significantly less effort in word segmentation, it is important to note that contextual information at the text level can impact the reading of subsequent words. The performance of target words in isolated, controlled sentences may not fully reflect the performance of word recognition in natural reading.

Finally, reading materials of different content lack comparability. Disparate reading materials may affect word recognition differently, such as the predictability of sentences. As a result, the comparisons of the first and second language reading performance of Chinese-English bilinguals, as well as their comparison with bilinguals of different language pairs, will be severely interfered. Obviously, preparing natural reading materials that are comparable to those in other languages is extremely time-consuming and laborious, especially when it comes to Chinese materials that require word segmentation.

In sum, the availability of high-quality natural reading data is crucial for studying Chinese reading performance. Given this, the authors set up the very first natural reading eye-tracking corpus (Chinese Ghent Eye-Tracking Corpus, GECO-CN) of Chinese-English bilinguals, which will be introduced in Chapter 2. In addition, it is the first eye-tracking corpus for the natural reading of Chinese a novel. This large-scale database is freely accessible online and can be utilized by researchers for research purposes. In this database, thirty native Chinese adults with English as their second language read the entire novel. Half of them read the first half of the novel in their first language (Simplified Chinese) and the second half in their second language, while the other half read in the opposite language order. In addition, these participants completed a series of language proficiency tests (in first and second languages), as they are known to affect word recognition (e.g., Diependaele et al., 2013).

Notably, this database and the Ghent Eye-Tracking Corpus (GECO, Cop et al., 2017a) used the same reading material in different language versions and shared the identical procedure. GECO investigated the eye movements of English monolinguals and Dutch-English bilinguals reading an entire novel (Cop et al., 2017a). Such a database often has high statistical power and incorporates a variety of stimuli. It enables the examination of the effects, hypotheses, and models in reading without the necessity of data collection. Indeed, GECO (Cop et al., 2017a) has been widely used to explore effects related to bilingualism (e.g., Cop et al., 2017b; Dirix et al., 2017) and the similarities and differences between bilinguals and monolinguals (e.g., Cop et al., 2015). Given this, the corpus we propose is feasible for investigating the within-group performance of Chinese-English bilinguals and for comparing the performance of bilinguals with different writing systems of the native language in conjunction with GECO.

In Chapter 3, we explore the frequency effects of Chinese words and characters during natural reading using data from the GECO-CN (Sui et al., 2022). The word frequency effect is known to be one of the robust effects based on the findings observed in alphabetic languages. However, Chinese words are composed of a limited number of characters, as mentioned. It means that the frequency of a character is often higher than a word. Given Chinese character properties are likely to affect word recognition due to the visual features (lacking the boundaries to delineate words), historical reasons (the relatively recent introduction of the concept of words; see Sui et al., 2022), and empirical evidence (some studies have observed facilitative, inhibitory, or absent effects; see Yan et al., 2006; Cui et al., 2021; Li et al., 2014), it may also affect word frequency.

If so, the high-frequency character may facilitate the recognition of low-frequency words, resulting in a limited difference in reading times between high- and low-frequency words. In addition, if the language-specific factor (e.g., character) affects the word frequency effect, then the underlying process of the word frequency effect should at least be somewhat different from alphabetical languages. Hence, our inquiry allows us to learn whether the Chinese word frequency effect is influenced by language-specific factors as well as to verify whether character frequency effects affect Chinese word processing. This study is the first to question that Chinese word frequency effects may be affected by language-specific factors. It is also the first to explore the frequency effects of two-character words and both their characters in Chinese text reading as continuous variables instead of categorizing them into dichotomous frequency categories.

In Chapter 4, we compare the word frequency effects in the first and second languages for bilinguals with different first-language writing systems but with the same second language as English. Currently, there are several existing theories of the word frequency effect, which can explain the findings of alphabetic language monolinguals and bilinguals quite well (i.e., the learning hypothesis, Morton, 1970; the lexical entrenchment hypothesis, Diependaele et al., 2013; the rank hypothesis, Murray & Forster, 2004; for discussion, see Duyck et al., 2008). Yet whether these theories can support the findings of Chinese bilinguals needs to be investigated.

Interestingly, the existing theories make different presumptions about word frequency effects in the first and second languages of bilinguals with various language pairs (i.e., Chinese-English and Dutch-English). Most of these models predict that if the language exposures, which affect word frequency effects, are comparable, word frequency effects should be similar across bilinguals, regardless of whether it is their first or second language or whether their languages share the same language system. That is, the word frequency of Chinese- and Dutch-English bilinguals should be comparable in size in the first or second language when there is no significant difference between their language proficiencies. In contrast, according to an extension of the rank hypothesis model (Duyck et al., 2008), bilinguals with alphabetic language pairs (e.g., Dutch-English) may exhibit comparable or larger word frequency effects in their first language than those with non- and alphabetic language pairs (e.g., Chinese-English) but have larger word frequency effects in their second language. However, the comparison of frequency effects between Chinese- and Dutch-English bilinguals in their native languages and second languages remains unclear and will, therefore, be explored in Chapter 4.

References

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), 1277–1287.

Boston, Marisa & Hale, John & Kliegl, Reinhold & Patil, Umesh & Vasishth, Shravan. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*. 2(1): 1-12.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), Article 16. https://doi.org/10.5334/joc.72

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?. *Bilingualism: Language and cognition*, 13(3), 359-371.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 441–458. https://doi.org/10.1037/xhp0000159

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Elsevier. https://doi.org/10.1016/B978-008044980-7/50017-3

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017a). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*, 49(2), 602–615.

Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (2017b). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, 20(4), 747-769.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic bulletin & review*, 22, 1216-1234.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, 41(3), 365–397. https://doi.org/10.1006/jmla.1999.2651

Cui, L., Wang, J., Zhang, Y., Cong, F., Zhang, W., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology*, 74(4), 610-633.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dijkstra, T. (2005). Bilingual Visual Word Recognition and Lexical Access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–201). Oxford University Press.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, 5(3), 175-197.

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41, 496–518.

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior Research Methods, 51*(6), 2800–2816.

Dirix, N., Cop, U., Drieghe, D., & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(6), 887-915. https://doi.org/10.1037/xlm0000352

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 663–679. https://doi.org/10.1037/0278-7393.33.4.663

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. (2008). The frequency effect in second-language visual word recognition: evidence against a language-selective rank hypothesis. *Psychonomic Bulletin and Review*, 15(4), 850-855.

Fiez, J. A., & Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences*, 95(3), 914-921.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, 38 (3), 313–338.

Hoosain, R. (1992). Psychological Reality of the Word in Chinese. In H. C. Chen, & O. J. L. Tzeng (Eds.), *Language Processing in Chinese* (pp. 111-130). Amsterdam, NY: North-Holland. http://dx.doi.org/10.1016/S0166-4115(08)61889-0

Hsiao, J. H., & Shillcock, R. (2004). Connectionist modelling of Chinese character pronunciation based on foveal splitting. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society* (pp. 601–606). Erlbaum.

Hsiao, J. H., & Shillcock, R. (2005). Differences of split and non-split architectures emerged from modelling Chinese character pronunciation. In B. G. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty Seventh Annual Conference of the Cognitive Science Society* (pp. 989–994). Erlbaum.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words?. *Journal of memory and language*, 44(1), 2-31.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology, 66*(3), 563–580.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, 127(6), 1139–1162.

Li, X., & Su, X. (2022). *Lexicon of common words in contemporary Chinese Second Edition*. The Commercial Press.

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), 895–913.

Li, X., Gu, J., Liu, P., & Rayner, K. (2013). The advantage of word-based processing in Chinese reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 879–889.

Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition, 22*(3-4), 441–457.

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1046–1059. https://doi.org/10.1037/a0035389.

Ma, G., Li, X., & Rayner, K. (2015). Readers extract character frequency information from nonfixated-target word at long pre-target fixations during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 41(5), 1409-1419.

Morton, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Murray, W. S., & Forster, K. I. (2004). Serial Mechanisms in Lexical Access: The Rank Hypothesis. *Psychological Review*, 111(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.

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., 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., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1188–1200. https://doi.org/10.1037/0096-1523.22.5.1188

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. Jo*urnal of Experimental Psychology: Human Perception and Performance*, 37(2), 514–528. https://doi.org/10.1037/a0020990

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157. https://doi.org/10.1037/0033-295X.105.1.125

Reichle, E., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*(4), 445-476. doi:10.1017/S0140525X03000104

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in cognitive sciences*, 7(11), 489-493.

Slattery, T.J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition*, 35, 1283-1292.

Sui, L., Woumans, E., Duyck, W., Dirix, N., (submitted) Word Frequency Effects Differ Across Writing Systems: The Word and Character Frequency Effect in Chinese Reading. *Language Learning*.

Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in Psychology*, 3, 174. doi:10.3389/fpsyg.2012.00174.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological science*, 20(8), 923-927.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, 9(4), 780–789.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3), 458–483. https://doi.org/10.1006/jmla.1998.2584

Whitford, V., & Titone, D. (2019). Lexical entrenchment and cross-language activation: Two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Language and Cognition*, 22(1), 58-77.

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.

Yu, L., Liu, Y., & Reichle, E. D. (2021). A corpus-based versus experimental examination of word- and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of Experimental Psychology: General*, 150(8), 1612–1641. https://doi.org/10.1037/xge0001014

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2018). Investigating word length effects in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 44(12), 1831–1841. https://doi.org/10.1037/xhp0000589

Zhou, J., Ma, G., Li, X., & Taft, M. (2018). The time course of incremental word processing during Chinese reading. *Reading and Writing*, 31(3), 607-625.

CHAPTER 2. GECO-CN: Ghent Eye-Tracking COrpus of Sentence Reading for Chinese-English Bilinguals[2]

Abstract

The current work presents the very first eye-tracking corpus of natural reading by Chinese-English bilinguals, whose two languages entail different writing systems and orthographies. Participants read an entire novel in these two languages, presented in paragraphs on screen. Half of the participants first read half of the novel in their native language (Simplified Chinese) and then the rest of the novel in their second language (English), while the other half read in the reverse language order. This article presents some important basic descriptive statistics of reading times and compares the difference between reading in the two languages. However, this unique eye-tracking corpus also allows the exploration of theories of language processing and bilingualism. Importantly, it provides a solid and reliable ground for studying the difference between Eastern and Western languages, understanding the impact and consequences of having a completely different first language on bilingual processing. The materials are freely available to be used by researchers interested in (bilingual) reading.

Keywords: Bilingualism, Reading, Eyetracking, Corpus

[2] Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. Behavior Research Methods, 1-21.

Reading is an essential skill for distilling content from written text. Psycholinguistic researchers have strived to understand the nature of the cognitive processes underlying reading, hoping to improve reading efficiency.

Over the past few decades, computational models of reading have been proposed to explain reading processes and a wide range of phenomena that are observed in English, the dominant language investigated in academic literature. Although these models seem to account for processing of alphabetic writing in general (Coltheart et al., 2001; McClelland & Rumelhart, 1981; Kintsch, 1988; Reichle et al., 2003), they are not always useful to explain processing a structurally different language, such as Chinese. Additionally, reading literature in general, and eye-tracking research in particular, has studied Chinese to some extent, but not sufficiently relative to its widespread use and great demographic relevance: fewer than 11% of all eye-tracking investigations use Chinese (Siegelman et al., 2022).

Although efforts have been made in recent decades to account for Chinese reading (Taft & Zhu, 1997; Perfetti et al., 2005; Rayner, Li & Pollatsek, 2007; Li & Pollatsek, 2020), models are still limited and need to be further validated. This is not surprising, since Chinese and English are dramatically different in orthography, sentence structure, and grammar. They do, however, share some similar theoretical hallmark phenomena that affect reading, such as frequency and prediction effects (Rayner & Duffy, 1986; Yan et al., 2006; Liversedge et al., 2014; Zola, 1984; Balota et al., 1985; Rayner & Well, 1996; Rayner et al., 2006; Rayner et al., 2005). Still, the peculiar properties of Chinese orthography (see Reichle & Yu, 2018 for a discussion) raise numerous questions that do not apply to the alphabetic writing system. One such peculiarity is that Chinese does not contain spaces between words, which will be discussed in length below (Li & Pollatsek, 2020). These language-specific characteristics may have a unique influence on natural reading. Investigating Chinese reading is theoretically relevant and necessary for cross-lingual comparisons that aim to understand the universal cognitive processes underlying reading.

Besides the differences between native language reading in English and Chinese, questions also arise for people that master both these writing systems, as is the case in Chinese-English bilinguals. How do people read when their two languages are completely different? Do their two languages interfere with each other while reading? There are a large number of Chinese-English bilinguals around the world, including many international students studying in a second language, yet it is still unclear how they could manage these two completely different languages and if there are any potential adverse impacts.

Understanding these questions is urgent and essential for theoretical reasons, such as shaping existing and future models of (bilingual) reading.

The prerequisite for the theoretical understanding of Chinese reading is the availability of high-quality natural reading data. This study therefore presents the first-ever corpus of Chinese-English bilingual natural reading employing eye-tracking to investigate the online reading process. Our participants read half of a complete novel in Chinese and the other half in English. Before reporting on the study, this paper first introduces some main differences between Chinese and English, summarises the important findings in experimental eye-tracking research of Chinese reading, and briefly reviews the existing eye-tracking corpus work. Next, we introduce the experimental procedure of and results from the Chinese Ghent Eye-tracking Corpus (GECO).

Chinese Writing System

The Chinese writing system is remarkably different from the Indo-European writing system in many ways. First, the morphology is different. English and other alphabetic writing systems are composed of letters, and the length of a word may vary depending on the number of letters. Chinese character, however, is a type of string formed by a number of strokes called characters. Each character is the same square size and equally spaced. A character is composed of radicals that are combined by strokes in a certain manner. Radicals denote phonological or semantic information, and their position can vary within different characters. Changing the number of strokes alters the visual complexity of the character, but not its length. For example, there are two strokes in the word 二 (TWO) and thirteen strokes in word 数 (NUMBER), while the character length of these two words is one. Visual complexity influences word identification to some extent, with longer fixation durations for more highly complex characters (Yang & McConkie, 1999; Su & Samuels, 2010; Liversedge et al., 2014).

Chinese words are comprised of characters in a flexible way, with some default rules. This unique feature allows a limited number of characters (approximately 5,000 unique characters commonly used) to compose an astounding number of words (about 56,000 commonly used words, based on Li & Su, 2022) and allows the creation of new words that can be widely accepted and understood. A character can be part of different words, sometimes with different pronunciations. For example, the pronunciation of the character 扇 is different when it is part of the words 扇子 (shàn·zi, FAN in English) and 扇动 (shān, FLAP in English).

The word length increases with the number of characters, as in Western scripts. The most common word types are two-character words, which account for about 70% of the most used word types (Li & Su, 2022). However, the most frequently encountered tokens are actually one-character words (see Li et al., 2015; Li & Pollatsek, 2020, for details).

In addition, Chinese writing is different from the alphabetical writing system because words are not separated by spaces. Chinese characters are presented one after the other without any visible boundaries. Importantly, in alphabetic writing, interspaces have a strong influence on identifying words and eye-movement control (Rayner et al., 1998; Perea, & Acha, 2009; but see Epelboim et al., 1994). This lack of clear delimitations can make reading difficult, resulting in longer fixations and different eye movement patterns compared to reading with apparent word boundaries. These models based on alphabetic languages also emphasise the importance of spaces that influence the landing position of the eyes (e.g., E-Z Reader model, Reichle et al., 2003).

The Chinese writing system, without explicit word segmentation, therefore seems fundamentally different from the word-based writing system. The unspaced writing style troubles the concept of word boundaries (Liu et al., 2013). As such, different readers may have different opinions as to whether the text (e.g., 躺在, LIE DOWN in English; see Liu et al., 2013) consists of two one-character words or one two-character word. This segmentation issue raises several questions: Are word units as important in Chinese, in which boundaries are ambiguous, as they are in English? Are distinctive and meaningful characters the basic units in Chinese reading? These questions are not without ground. Until twentieth centuries, there was only the concept of character. Although Chinese sentences are now read horizontally from left to right, they used to be written vertically and read character by character from top to bottom. And even now, characters are still the basic unit in the Chinese dictionaries. In clarifying these issues, a series of studies investigated the importance of words during Chinese reading.

Native Chinese readers have been shown to spend more time reading sentences when there are spaces between each character or within words (two-character word) than under traditional unspaced or spaces between words conditions (Bai et al., 2008). In addition, when reading a sentence using the moving window paradigm in which a sentence is completely masked apart from the fixated point, readers spend more time reading the sentence if the composed characters of a word are not presented simultaneously (Li et al., 2013). These results thus showed that readers' performance was influenced by visibility of the entire word, arguing against the assumption that the character is the basic representational unit of reading. Other arguments may be obtained from recognition task data. If the character is the basic unit of the reading process, the properties of the word should not affect word recognition. In a study by Li et al. (2009), Chinese readers were shown four characters very briefly (80 ms). These four characters constituted either a four-character word or two two-character words. The participants were able to report the four-character word, but usually reported only the first word when two two-character words were presented. This result suggests that text recognition is driven more by word representations than by characters, consistent with the word superiority effect found in English (Reicher, 1969; McClelland & Rumelhart, 1981). When having to provide a button-response to the location (top or bottom) of the character embedded in the noise but not the character itself, participants are faster if the target character can compose a word with the adjacent legible character than when it cannot (Li & Pollatsek, 2011). Additionally, the variance in eye movement measures such as fixation durations and fixation probabilities can be better explained more by words than by character-driven parameters in mixed-effect regression models (Li et al., 2014).

Based on the above evidence, it is reasonable to believe that word representations are salient to Chinese readers during reading, even if their boundaries are visually ambiguous. Yet, it is still not entirely clear how readers segment words while reading (for research on word segmentation on ambiguous word, see e.g., Ma et al., 2014; Huang & Li, 2020; for a Chinese reading model explaining word segmentation, see Li & Pollatsek, 2020) and why word segmentation is inconsistent among different readers (Liu et al., 2013). One of the main questions is how the processes underlying Chinese reading differ from that of alphabetic languages. Answering these questions is vital in order to verify and complement existing computational models for reading.

Eye Movements in Chinese Reading

Over the years, efforts have been made to explore Chinese online reading processes, with eye-tracking as an effective method. This technique enables the detection of participants' saccades and fixations with high spatial and temporal accuracy. Saccades are the rapid eye movements from one point to another, while fixations are the pauses made to extract information from the text. This method allows participants to read at their own pace without any time pressure (in contrast to studies on lexical decisions, which require speeded responses, and which contains a decision component that may be dependent on strategic factors), with words embedded in meaningful sentences, and is as such more naturalistic in comparison with experimental tasks used in the field.

Despite the very different orthographies of Chinese and English, there are some similarities between reading in these two languages, according to studies exploring target word recognition in isolated sentences. The effects of frequency, predictability, and word length found in Chinese reading are consistent with previous evidence in alphabetic languages such as English. Chinese readers spend less time fixating on high-frequency words than on low-frequency words (Wei et al., 2013; Yu et al., 2020). Furthermore, fixations on highly predictable words in a sentence are shorter and skipping rates are higher (Rayner et al., 2005). Similar patterns are found for short in comparison with longer words, with longer fixations and less skips for longer words (Zang et al., 2018). Additionally, Rayner and colleagues (2007) claims that eye-movement control in reading appears to be fundamentally similar between Chinese and English readers (also see Li et al., 2014; for a review, see Li et al., 2022). They tested native Chinese readers with isolated sentences. They demonstrated that by using parameters derived from English reading (e.g., frequency, predictability, distance to the fixations point, and fixed parameters), the E-Z Reader Model made fairly good predictions of eye movements, close to the actual performance of participants reading in native Chinese.

There is no doubt however that the marked differences between Chinese and alphabetic writing systems (e.g., in spelling and grammar) also leads to differences in visual word processing. Character complexity, for instance, influences recognition of words embedded within sentences; as the number of strokes increases, fixation durations increase as well, whereas skipping probability decreases (Liversedge et al., 2014). Character frequency may further affect eye-movement behaviour, with fixations on target words in isolated sentences being longer when the initial character frequency of the two-character word is low (Yan et al., 2006). Still, it must be noted that these effects were not present in the study by Li et al. (2014), although the lack of reliable results in this work could be due to the substantial correlation between word and character properties (e.g., frequency). Later, Yu et al. (2020; but see also Yan et al., 2006) found evidence for an inhibitory character frequency effect, observing that word identification is slowed when the initial character is highly frequent. Yu et al. suggest that factors such as low sentence constraint and large orthographic neighbourhood might account for the discrepancy with previous studies.

In terms of the general reading patterns, fixation durations appear to be somewhat longer in native Chinese adults than their English or other European languages counterparts

(Liversedge et al., 2016; Feng et al., 2009; Rayner et al., 2005). Like English readers, Chinese readers also occasionally make regressions to previous content while reading but they do this more frequently than English readers (Feng et al., 2009; Rayner et al., 2005). The occurrence of word skipping, however, seems somewhat more inconsistent across different studies. When reading isolated sentences, native readers appear to skip about 42% of Simplified Chinese characters (Chen et al., 2003), 3-25% of Simplified Chinese words (Rayner et al., 2005; Yan et al., 2006), and 10% of Traditional Chinese words (Tsai et al., 2004). Word skipping probabilities between Chinese and English readers using comparable materials written in Simplified Chinese and English do not differ significantly (Rayner et al., 2005). However, when reading expository texts written in Simplified Chinese, word skipping rates reach 47%, somewhat higher than, for instance, English and Finnish readers with the same material in different language versions (Liversedge et al., 2016).

Furthermore, the perceptual span of Chinese readers seems to be much smaller than that of English readers, regardless of whether the test material is in Simplified Chinese (Inhoff & Liu, 1998) or Traditional Chinese (Chen & Tang, 1998). When employing the moving window paradigm (McConkie & Rayner, 1975) to manipulate the perceptual span size around the fixation point by masking the rest of the words, Chinese span is shown to be one character to the left of the fixation and 2-3 characters to the right (approximately 0.9-1.2 degrees of visual angle per character, Inhoff & Liu, 1998; Chen & Tang, 1998). In comparison, the perceptual span in English is 3 – 4 letters to the left and 14 – 15 letters to the right of the fixation point (approximately 1 degree of visual angle per three characters; Rayner et al., 1980). This difference might be due to the lack of inter-word space, resulting in greater information density of Chinese (Yan et al., 2006). A similar pattern was found in the saccadic amplitude, as the size of forwarding (rightward) saccades is 2-2.5 characters in Chinese reading (Inhoff & Liu, 1998; Rayner, Li, Williams, et al., 2007) and 7.5 letters in English reading (Rayner, Li, Williams, et al., 2007).

To sum up this paragraph, although there has been some progress in studying the Chinese reading process, there is still much to be unravelled. Crucial to note here as well is that most of the studies reported above only explore the target word recognition in isolated sentences, and do not consider sentence-level processes. Reading, in fact, goes beyond the level of words and cannot be fully grasped with small-scaled studies on limited stimuli sets. To illustrate, the percentage of skipping rates is much higher when reading a paragraph than when reading an isolated sentence (Yan et al., 2006; Liversedge et al., 2016). Similarly, sentence- and paragraph-level processes may influence also other characteristics of reading that may only be discovered in natural reading of longer, running text. It is also worth noting that the limited amount of collected data sometimes leads to a lack of power to detect reliable phenomena, such as the character frequency effect that failed to be found in the work of Li and colleagues (2014). Below, we introduce a methodology without these limitations for exploring online reading processes.

Eye-Tracking Corpora

Eye-tracking corpus research is an approach in which researchers collect a large amount of data that allows for in-depth analyses with high statistical power and the ability to detect even minimal effects. In contrast to small-scaled experimental studies that generally use a limited number of stimuli or sentences to investigate reading behaviour, this type of research by nature includes a wide range of stimuli. It entails the possibility of taking language variations into account to provide a comprehensive picture of written language processing. This is especially important because research has shown that there is an alarmingly low degree (less than 17%) of shared variance between the widely used paradigms in visual word recognition (e.g., lexical decision) and eye-tracking data (e.g., gaze durations and first fixations durations; See Kuperman & Van Dyke, 2013; Dirix et al., 2019), implying that natural reading processes may not be completely, and even not considerably, captured using only tasks like lexical decision. In addition, such large databases allow examining existing hypotheses and models, investigating multiple main effects and interactions of factors involved in reading, and evaluating the replicability of effects obtained in studies without conducting new experiments (Demberg & Keller, 2008; Whitney, 2011; Kuperman & Van Dyke, 2013; Chuang et al., 2021).

As such, this method of exploring eye movement performance with sizable data has many advantages. And although in general the number of eye-tracking corpora is very limited, a few studies have adopted this approach in recent years. The Dundee corpus (Kennedy & Pynte, 2005) was probably the first eye-tracking corpus of natural reading. In this study, ten native French and ten native English participants read 20 newspaper texts written in their native language. The English texts consisted of 56,212 tokens in total, and the French texts contained a total of 52,173 tokens. Another example is a corpus gathered by Frank and colleagues (2013), who had 43 participants read 205 British English sentences with a total of 1,931 tokens. In addition, the Zurich Cognitive Language Processing Corpus (ZuCo, Hollenstein et al., 2018) provided EEG and eye-tracking data from 12 native English speakers reading sentences extracted from movie reviews and the Wikipedia relation extraction dataset, with a total of 21,629 tokens. Later, Hollenstein and colleagues (2019) presented the Zurich Cognitive Language Processing Corpus 2.0 (ZuCo 2.0), an EEG and eye-tracking corpus of 18 native English speakers reading 739 sentences. Furthermore, the Provo Corpus (Luke & Christianson, 2018) presented eye movement data from 470 native speakers of American English reading 55 short passages with 2,689 tokens and 1,197 unique word types in total.

Corpora not including English include the German-language Potsdam Sentence Corpus (Kliegl et al., 2006) with data of 1,138 tokens read by 222 readers, the Dutch Eye-Movements Online Internet Corpus (DEMONIC; Kuperman et al., 2010) containing data of 55 participants reading 1,746 tokens, the Russian Sentence Corpus (RSC, Laurinavichyute et al., 2019) with 96 Russian monolinguals and reading a total of 1,362 tokens, and the Beijing Sentence Corpus (Pan et al., 2021) with eye-tracking data of 60 Chinese native participants reading sentences from newspapers, totalling in 936 types and 1,685 tokens.

Important to mention is that all previously mentioned corpora are limited to monolingual data. However, the majority of the population nowadays speaks more than one language, and this number is steadily increasing. To address research questions on bilingualism, there has been a growing body of studies in related fields such as linguistics, education, and psychology in recent decades (Kuperman et al., 2022). Surprisingly, to our knowledge there are currently only three eye-tracking corpora which include data of second language (L2) reading. The first work to introduce a corpus of bilingual reading is the Ghent Eye-Tracking Corpus (GECO, Cop, Dirix, Drieghe, & Duyck, 2017), which was used to answer several questions about bilingual reading. Participants read a novel with 59,716 tokens in the Dutch version and 54,364 tokens in the English version. Paragraphs were displayed on the screen, simulating the process of reading a book. The study recorded the eye movements of 19 unbalanced Dutch-English bilinguals who read half of the novel in the first language (L1) and the other half in L2 (the order was counterbalanced between subjects), and also included data of a set of 14 English monolingual participants (who read the book entirely in their native language). In another recent study, the Bilingual Russian Sentence Corpus (BiRSC, Parshina, 2020) recruited 50 English-Russian Heritage speakers and 27 L2 learners, classified them as beginners and advanced speakers. The study asked beginners to read 30 sentences and advanced speakers to read 72 sentences. However, this corpus only had participants reading isolated sentences in their L2 (Russian). The most recent Multilingual Eye Movement Corpus (MECO, Siegelman et al., 2022, Kuperman et al., 2022), was a large-scale multi-lab study, collecting data on bilinguals (12 groups with different native languages) and English monolinguals. The bilingual participants read 12 short texts in L1 with 1487~ 2412 tokens (depending on the language) and 1653 tokens in L2. The majority of the bilingual groups had a European language as L1, such as Dutch, German, Italian, and Spanish, and all had English as their L2.

Although these corpora are generally larger than small-scaled experiments in terms of collected data, some still have a rather limited number of stimuli. The small amount of testing materials, however, may result in variable fixation times of words across various language contexts, in this case the specific texts, as Dirix and colleagues (2019) have shown when comparing the two databases (GECO, Cop, Dirix, Drieghe, & Duyck, 2017; Dundee corpus, Kennedy & Pynte, 2005). However, averaging repeated presentations of the target word can decrease noise and provide a more stable eye movement estimate, which only large databases can achieve.

Another drawback is that the materials, instructions, and participants in available corpora are predetermined. To illustrate, a corpus may not include the essential stimuli or age group needed to address a specific research question. Furthermore, many corpora display unrelated sentences in an isolated way, with the exception of the Dundee corpus, GECO, and MECO. The lack of a naturalistic language context and diverse stimuli could limit the exploration of natural reading. Yet, even when texts are coherent, they still represent only part of all genres. In the case of GECO, the text is a murder mystery, and it is uncertain whether results from this specific text can be fully generalised to other types of fiction and to non-fiction. Indeed, Brysbaert (2019) already demonstrated that reading rates (expressed in words per minute – wpm) are faster for fiction (260 wpm) than for non-fiction (238 wpm).

The practical and theoretical importance of these corpora have been demonstrated by their extensive use in empirical studies. Based on these corpora, many experiments have (re-)evaluated theoretical frameworks, such as syntactic processing complexity (Demberg & Keller, 2008), hierarchical structure in sentence processing (Frank & Bod, 2011), predictability of computational models (Mitchell et al., 2010; Hollenstein et al., 2021), and test factors that impact reading behaviour, for example word characteristics such as frequency and predictability (Kennedy et al., 2013), adjacent words (Pynte & Kennedy, 2006), and age of acquisition (Dirix & Duyck, 2017). As such, corpora are an interesting and productive breeding ground for new empirical research.

Regarding the bilingual corpora, GECO (Cop, Dirix, Drieghe, & Duyck, 2017) has been applied in several studies that explored differences between L1 and L2 written language processing. A first general comparison between reading in the L1 and L2 on a sentence level showed that L2 reading is more time-consuming (205 ms longer, with 13% more fixations) and is 5% less prone to skipping than L1 reading (Cop, Drieghe & Duyck, 2015). Furthermore, the frequency effect is larger when reading in L2, and it is negatively correlated with proficiency in L1 regardless of the text language in which a bilingual is reading (Cop, Keuleers, et al., 2015). Also, the study on the effects of age of acquisition in bilinguals (Dirix & Duyck, 2017) has shown that words learned early are recognised more quickly in both L1 and L2. Importantly, L2 reading performance also appears affected by the age at which the translation equivalent in L1 was learned. Moreover, there is some evidence of parallel bilingual language activation, even when reading in a single language. Reading times in the L2 seem to benefit from the density of the cross-language neighbourhood (Dirix, Cop, et al., 2017), with reading times in L2 being shorter for words with a denser orthographic neighbourhood in L1. Other support of parallel activation of languages comes from the cognate facilitation effect when reading the narrative text in the L1 and L2 (Cop, Dirix, Van Assche, et al., 2017). All these studies illustrate how bilingual eye-tracking corpora may be an important data source for empirical studies assessing a wide range of research hypotheses.

Surprisingly, since the presentation of the first bilingual eye-tracking corpus (Cop, Dirix, Drieghe, & Duyck, 2017), no other corpora of the same size have appeared. From the perspective of bilingualism research, Dutch-English is the only language pair for which such dataset is available. It is the current study's aim to fill this void in the literature by presenting the first Chinese-English corpus of bilingual reading.

GECO-CN

Composing a Chinese reading corpus can be challenging. One of the problems, for instance, already lies in defining Chinese words boundary due to the lack of visual clue between words. Hence, researchers are cautious, selecting only words with generally accepted boundaries to avoid confusion and disagreement (Yan et al., 2010; Pan et al., 2021). This also applies to the recently published Beijing Sentence Corpus (Pan et al., 2021). Modifications were made to the sentences in order to avoid ambiguous word boundaries, resulting in written language that was different from what a reader may naturally expect. Furthermore, although single-character words are the most frequently encountered Chinese tokens, the corpus holds only 348 single-character words out of 1685 tokens, or about 20%.

The problem of unclear word boundaries in Chinese is indeed a big challenge. However, one must face it because it is an essential part of the actual performance of Chinese readers, which is the first reason why it is important to have an eye-tracking corpus that simulates real-world reading, also for non-Western languages. Controlled stimuli will bias reading performance away from natural variability, affecting core characteristics like fixation durations and skipping probabilities. As noted above, artificial materials can result in a much higher percentage of two-character words, making the average word length of the materials longer than in natural texts. Since word length influences eye movement behaviour (Zang et al., 2018), participants consequently show longer fixation durations and lower skipping probabilities than reading in real life. In addition, natural reading materials are essential for understanding how readers effectively segment words without negatively affecting reading performance and whether the diversity of word segmentation among participants impacts reading and fixation landing points.

Another problem with reading studies in general is that most of the existing eye-tracking experiments employ isolated words and sentences. Nevertheless, in daily life, text is mostly read in long paragraphs and semantically connects to each other. Reading a coherent and meaningful text of comparable length involves rich linguistic processes (e.g., syntactic parsing, for a review, see Rayner & Reichle, 2010) and cognition (e.g., working memory, Miller et al., 2006; Peng et al., 2018). That is, in addition to integrating the meaning of words, parsing syntactic information, and identifying ambiguous sentences, the preceding context (Rayner & Well, 1996; Rayner et al., 2005), prior knowledge (Woolley, 2011), and many other processes also play a role in natural reading. Investigations with artificially designed experiments may only partially tap into these components, profoundly affecting the reading processes. The results obtained in experimental tasks, such as lexical decision and naming tasks, are insufficient and fail to predict the performance in natural reading, as shown by Kuperman, Drieghe, et al. (2013) and Dirix et al. (2019).

Different from the widely studied and established alphabetic language reading models, the theoretical models on Chinese reading are still developing. Currently, almost all the existing Chinese reading models remain at the word level and are not fully empirically validated (see Reichle & Yu, 2018, for review). However, an ambitious reading model should explain more than a word-level process and take other coordination processes into account (Kuperman, Drieghe, et al., 2013). More importantly, with the globalisation of our society and the growing number of bilinguals, a reliable reading model should also consider the process of reading in languages other than the mother tongue, especially when the L2 may be qualitatively different from the native one, for example by having different orthographies and writing systems.

This study aims to contribute to answering the questions above. Here, we present the very first Chinese-English eye-tracking corpus for bilinguals reading an entire novel. It is also the first eye-tracking corpus of Chinese reading of paragraphs. Native Chinese speakers read half of the novel in Chinese and the other half in English (the order of which part was read in L1/L2 was counterbalanced between subjects). In total, each participant read about 5,000 sentences. This methodological paper summarises their eye movement data (including the distributions and reliability coefficients of several eye-tracking measures), basic descriptive statistics of the Chinese and English reading materials, and background characteristics of the participants.

This database can be employed to address previous limitations discussed in the introduction, since it allows investigating and comparing diverse aspects involved in reading, examining the validity and generalisability of existing experimental research or models based on limited test materials. For example, future research can validate assumptions of the E-Z Reader Model (Rayner, Li & Pollatsek, 2007) and Chinese reading model (CRM), Li & Pollatsek, 2020) in predicting eye movements when reading Chinese in paragraphs using data of this eye-tracking corpus. In addition, this study uses different language versions of the same reading material to explore Chinese and English, languages which have apparent discrepancies in spelling and syntax. This creates the possibility of investigating the specificity of potential or well-known factors involved in reading, such as the homophone effect (Chen et al., 2009). Clarifying these issues is helpful to the construction and development of both Chinese and universal reading models.

Furthermore, this eye-tracking corpus allows examining the interaction between two very different linguistic systems. Languages that are very dissimilar in orthography may not influence word recognition for each other in the same way as those originated from a linguistic family (e.g., cross-language neighbourhood effect, Dirix, Cop, et al., 2017; cognates effect, Van Assche et al., 2009), although they are activated even when reading unilingual text (Van Heuven et al., 1998; Dijkstra, & Van Heuven, 2002).

Finally, our corpus allows comparisons between readers with a variety of language pairs. Under the concern of geographical difficulties, material incomparability, and the limited number of data, the comparisons between bilinguals with different L1 are somewhat difficult. However, this study shares almost identical reading materials with the Dutch-English GECO (Cop, Dirix, Drieghe, & Duyck, 2017), which facilitates comparison of two completely different L1s: Dutch and Chinese.

Method

Participants

Thirty-two native speakers of Chinese, born in mainland China and studying in Belgium, with English as L2, participated in the study with remuneration for their time. Two participants were excluded from the analysis: one due to excessive head movements, the other due to the possibility of non-attentive reading[3]. The remaining 30 participants (8 males) with an average age of 25.3 years (range: 20-29; SD = 2.60) were Master or Ph.D. students at Ghent University or Leuven University. The average age of acquisition for English was around eight years old (range: 3-18; SD = 3.23). No participants reported language and/or reading deficits and all had a normal or corrected-to-normal vision.

In addition to the eye-tracking experiment, participants completed the LEAP-Q questionnaire (Language Experience and Proficiency Questionnaire, Marian et al., 2007) to investigate their language background; the HSK test (Chinese Proficiency Test, level 6) to explore Chinese proficiency; and three tasks to objectively assess English proficiency (in accordance with the first GECO, Cop, Dirix, Drieghe, & Duyck, 2017, and to facilitate cross-corpora comparisons between Dutch and Chinese native speakers). The three tasks were the LexTALE (Lexical Test for Advanced Learners of English; Lemhöfer & Broersma, 2012), an unspeeded lexical decision task; WRAT 4 (Wilkinson & Robertson, 2006), a spelling task; and a classic speeded lexical decision task (See details in Table 1).

Based on the classification of LexTALE, two participants were in the lower intermediate group (below 59%), 16 were in the upper intermediate group (60% - 80%), and 12 were in the advanced user group (80% - 100%). This aligns with the high educational level of the participants. To understand the similarities and differences in reading performance between the two language families and the impact of different first languages on reading the same second language (positively or negatively), we present the comparative data between Chinese and Dutch bilinguals in Table 1 by comparing GECO with GECO-CN.

[3] The participant was excluded because we suspected inattentive reading based on reading performance. The participant either scanned the sentences with surprisingly fast speed or read from the middle of the sentences, even in second language reading. Furthermore, the participant had very low comprehension scores in both languages and was unable to tell the storyline when chatting with the first author after each session. Thus, we exclude this participant from the analysis and corpus after thoughtful consideration to ensure a highly qualified database.

Table 1 *Descriptive Statistics of Subject Information*

	Chinese Bilinguals		Dutch Bilinguals		t Value	t Value	
	L1 (Chinese)	L2 (English)	L1 (Dutch)	L2 (English)	L1-L2	L1-L1	L2-L2
(%)	M[SD]	M[SD]	M[SD]	M[SD]	(Chinese-English)	(Chinese-Dutch)	(English-English)
Comprehension Score	74.67[11.74]	65.83 [13.27]	79.63[10.96]	78.95[12.54]	3.79[29]***	-1.50[40.43]	-3.49[40.08]**
Subjective Exposure	49.23[23.44]	45.03[22.71]	75[15.25]	25[15.25]	0.51[29]	-4.66[46.94]***	3.69[46.789]***
Lexical decision Score		41.73[16.84]	80.19[5.41]	56.84[11.12]			-3.74[45.49]***
Lextale Score		75.75[12.07]	92.43[6.34]	75.63[12.87]			0.03[36.59]
Spelling Score		60.24[12]	83.16[7.81]	69.92[8.74]			-3.26[45.97]**
HSK	95.47[3.01]						
Age of Acquisition		7.73[3.23]		11.26[2.47]			--4.32[45.19] ***

*p < .05. **p < .01. ***p < .001

The average mean[standard deviations] is in the first four columns, and the result[degrees of freedom] is in the last three columns. The t-test columns represent comparisons between Chinese and English scores among Chinese-English bilinguals, between Chinese and Dutch bilinguals in their L1, and comparisons in their second language.

Subjective exposure is the percentage of time that participants are currently exposed to a language environment, queried in the LEAP-Q questionnaire (Marian et al., 2007). Although English is the work/study environment for Chinese participants, Dutch, French, and German are the official languages in Belgium. Some participants were also exposed to language environments other than Chinese and English. Thus, the sum of some subjective exposures does not equal 100%.

Materials

Following the research of Cop, Dirix, Drieghe, & Duyck (2017), this work employed the detective story *The Mysterious Affair at Styles* (斯泰尔斯庄园奇案 in Chinese) written by Agatha Christie (1920) as reading material. The book was chosen after careful deliberation of copyright issues (free to use, also for further research), the length of reading the novel, the familiarity of the words in the novel, and the availability of multiple languages for future research and comparison (see details in Cop, Dirix, Drieghe, & Duyck, 2017).

The novel was divided into two parts, each presented in one of the two languages. The order of languages was counterbalanced between participants. Fifteen participants read Chapter 1-7 in L1 (Chinese) and Chapter 8-13 in L2 (English), while the other 15 participants read in reverse language order. Chinese text was displayed in simplified form. The Chinese version of the novel has 59,403 words with 5,053 unique types, and the English version has 56,841 words with 5,363 unique types. More detailed information of the novel is presented in Table 2.

Table 2 *Descriptive Statistics of the Chinese and English Versions of the novel "The Mysterious Case at Styles"*

	Chinese			English		
Number of words	59403			56841[4]		
Number of word types	5053			5363		
Number of nouns	9996			8911		
Number of noun types	1708			1812		
Number of sentences	5066			5242		
	M	**SD**	**Range**	**M**	**SD**	**Range**
Number of words per sentence	11.73	8.86	[1-78]	10.84	8.27	[1-70]
Word frequency	5.96	1.31	[1.47-7.70]	5.91	1.37	[1.17-7.67]
Word length	1.44	0.63	[1-6]*	4.31	2.41	[1-19]

***** Note that words and phrases are included when calculating the average word length in Chinese and English. In Chinese, the phrases are included because they cannot be segmented without changing the meaning (e.g., 没关系; NEVER MIND in English), as is the case in the *Modern Chinese Dictionary 7th Edition* (2016). In English, it is because of the abbreviations, e.g., Mary's.

Frequencies are Log10-transformed: SUBTLEX-CH for Chinese words (Cai & Brysbaert, 2010) and SUBTLEX-UK for English words (Van Heuven, et.al., 2014). The log frequency transformation method of the two languages is the same.

Apparatus

The equipment used to collect the eye movement data was the common desktop-mounted EyeLink 1000 Plus system (SR Research, Canada) using a 1 kHz sampling rate. Participants were required to use a chin- and headrest to minimise head movement. The Experiment Builder software package (SR Research Ltd.) was used for stimuli presentation. The text was presented in paragraphs on screen with no more than 120 words in English and 200 words in Chinese per paragraph. One screen was counted as one trial. Texts were triple-spaced and displayed in a style corresponding to the language. For example, the dialogues of different characters were presented in different paragraphs, in line with the Chinese writing style, although different from the English and Dutch versions (Cop, Dirix, Drieghe, & Duyck, 2017). The words were in 28-point Courier New font and presented in black against a light grey background, and 1.6 Chinese characters and two English letters subtended 1 degree of

[4] Note that there are some differences with GECO (Cop et al., 2017) in the reported number of words and nouns. In contrast to Cop et al. (2017), we did not apply any filters in the data report (e.g., in their published report of the dataset, some names and number words are not included; the complete dataset of the English part of both GECO and GECO-CN however contains the same material).

visual angle or 59 pixels. Although participants read the text binocularly, only the movements of one eye were recorded.

Presentation ® software (developed by Neurobehavioral Systems) was used to collect the data of the Lexical decision task. Presentation and EXCEL were used to conduct LexTALE before and during the COVID pandemic, respectively.

Procedure

The experiment consisted of four sessions of two hours each. Participants completed the study within a three-week time period. They had at least one day in between each session and a maximum of three sessions per week. Participants read Chapter 1-4 during the first session, Chapter 5-7 during the second session, Chapter 8-10 during the third session, and Chapter 11-13 during the fourth and final session.

Participants were invited to read the novel in a relaxed and natural way. They were instructed to try not to move their heads while reading in silence, and they could take a break whenever they wanted after finishing a trial. Any questions about the study were explained and answered by the researcher. The experiment took place in a quiet, dimly lit lab.

Printed summaries of previous chapters were provided to participants at the start of each session (with the exception of the first), helping them recall the previous storylines. The experiment began with an instruction presented on screen, followed by a 9-point calibration. Participants then read three or four paragraphs from *Alice in Wonderland* as a practice run and answered two multiple-choice questions about the story to get used to the test environment and procedure. After being familiarised with the experiment and experimental setup, participants started the main task. Recalibration was carried out before the start of each chapter and then approximately every 10-15 minutes regularly. Calibration was also performed again when participants moved their heads.

During the main task, participants could read at their own preferred speed by pressing the spacebar to control when to move on to the next part (they did not have the opportunity to revisit previous paragraphs, as was also the case in the original GECO study). There was a drift check after each trial. Participants could continue if the error was less than 0.5º, otherwise, there would be a recalibration.

After finishing each chapter, participants answered several (1-6) pencil-and-paper multiple-choice questions with four answer options about chapter content, ensuring they paid attention to the story rather than just reading without processing meaning (see scores in Table

1). The language of the questions was congruent to the language of the chapter, and the number of questions was proportional to the length of the chapter.

Word Segmentation

Chinese words are salient for native readers despite the lack of spaces in between. Chinese word concepts have been investigated by Chinese scholars back in the 1960s (e.g., Lu, 1964) and the definition agreed upon to determine word status now refers to the smallest linguistic unit with specific meanings that can be used independently, rather than simple combinations of the meaning of the characters. This entails that word boundaries in linguistics may differ from the reader's view, but are more analogous to alphabetic scripts.

Word segmentation is an indispensable step in ensuring comparability between Chinese and word-based languages. After fully considering the stringency of word segmentation[5], this study divided sentences manually into words according to the authoritative word dictionary—"*Modern Chinese Dictionary 7th Edition"* (2016) rather than by words listed in "*Lexicon of common words in contemporary Chinese"* (2008). The study furthermore followed commonly accepted rules for word segmentation (Fu, 1985; Ge, 2014; Liu, 2019), such as considering the inflection of the word as a single word (e.g., the word 笑笑 (smile) is the inflection of the word 笑 (smile) and is considered as a single word). For words with indistinguishable boundaries, suggestions were provided by the associate editor who participated in the compilation of "*Modern Chinese Dictionary 7th Edition*" (2016) through personal communication with the first author.

Results and Discussion

This methodological paper disclosing the corpus will report descriptive statistics on five basic reading time measures: a) First fixation duration (FFD), the duration of the first fixation on the current word; b) single fixation duration (SFD), the duration of the fixation on a word that is fixated only once; c) gaze duration (GD), the summed duration of all fixations on the current word before the eyes move on to the next (right-side) word; d) total reading time (TRT), the summed duration of all fixations on the current word; and e) go-past time (GPT), the summed durations of all fixations and regression to the previous (left-side) words since the current word is first fixated until the next (right-side) word is fixated. The distribution and descriptive statistics of these reading time measures are shown in Figure 1. This paper also discloses all data, freely available online (access link:

[5] The first author contacted Professor Su Xinchun of Xiamen University who participated in the compilation of "Lexicon of common words in contemporary Chinese" (2008) to understand the difference between this book and "Modern Chinese Dictionary 7th Edition" (2016).

https://osf.io/pmvhd/?view_only=77def2827a514254957cc846e14826cf), for further research. See Appendix A for details of supplementary materials.

Fixations shorter than 100 ms are not likely to reflect the processing of written language (Sereno & Rayner, 2003) and were therefore removed from the analysis. The analyses below were conducted using Rstudio software (Version 2021.09.1-372, developed by R Core Team) and report on the trimmed data, unless noted otherwise.

Reading Times Distributions

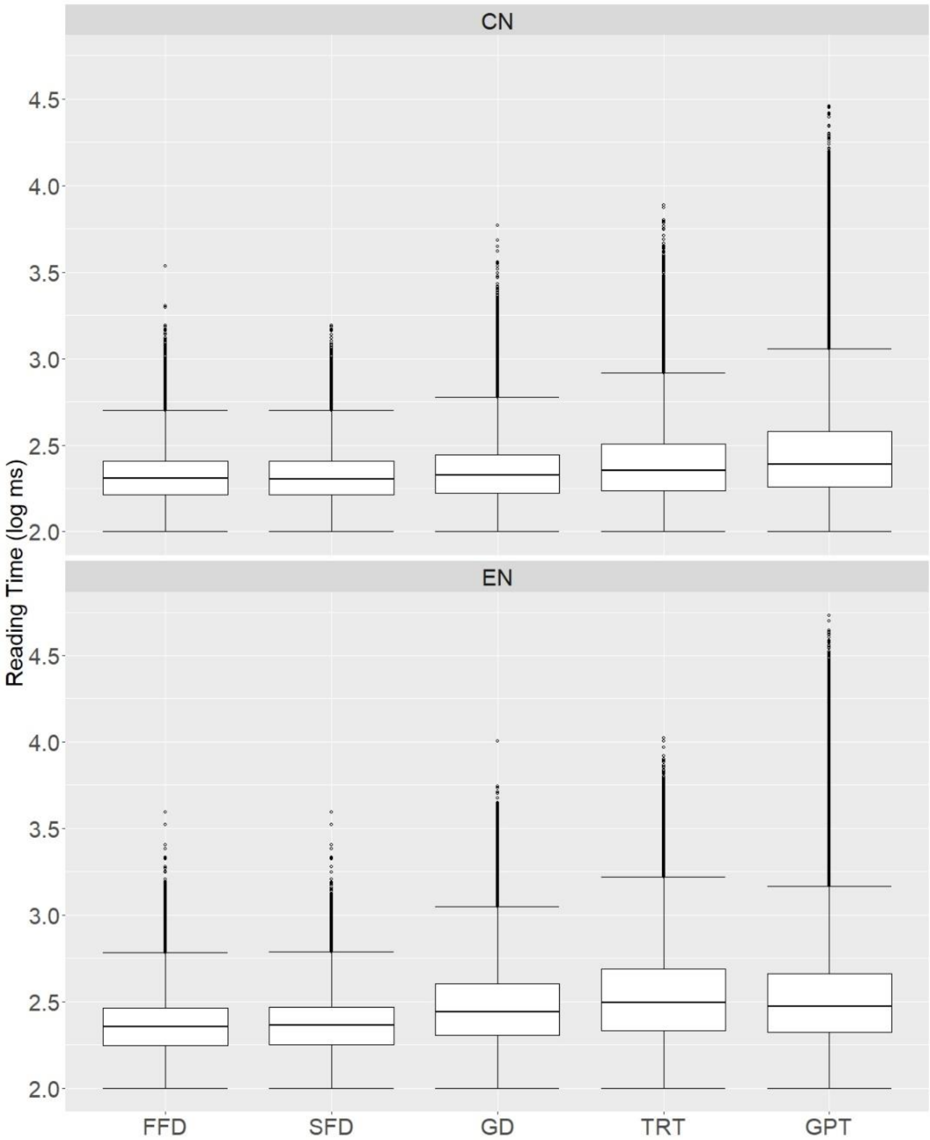
Excluded trials were the few trials accidentally skipped by participants (i.e. by pressing the spacebar by mistake) or trials for which the machine failed to detect eye movements because of technical malfunction (we removed 15 trials in total out of 19,140 trials). Boxplots show the log-transformed reading times, aggregated over participants (See Figure 1). A large number of positive outliers were found, which is consistent with GECO (Cop, Dirix, Drieghe, & Duyck, 2017). The median of all five reading time measures of the L2 (English) is slightly higher among Chinese-English bilinguals than among Dutch-English bilinguals (Cop, Dirix, Drieghe, & Duyck, 2017). Regarding the reading rate, Chinese bilinguals read about 466 wpm in L1. In contrast to previous studies that showed a comparable reading rate between Chinese (260 wpm) and English speakers (200–320 wpm; Brysbaert, 2019; but see Yen et al., 2011)[6], this work thus reports considerably higher rates. These very distinct results are unexpected yet plausible.

The Chinese reading rate from Brysbaert (2019) is based on reading sentences and texts in Traditional and Simplified Chinese. The reading rate of paragraphs, however, tends to be higher than that of single sentences (Radach et al., 2008), which is also supported by the higher skipping probability in paragraphs discussed below. In addition, paragraphs are likely to be smaller in font size than individual sentences when displayed on the screen. When the font size is smaller, the reader may process more information within a comparable perception span, thus showing a faster reading rate (also see Yen et al., 2011). The nature of the reading material (e.g., difficulty) and experimental condition may also have an effect. Reading a novel in Simplified Chinese in a natural way may be easier to read than the texts of previous research and faster than those in Traditional Chinese or under certain experimental conditions (e.g., when using the self-paced moving-window paradigm; Zhang & Perfetti, 1993 whose reading rate was incorporated into the analysis of Brysbaert, 2019). It is also possible that

[6] Note that in Brysbaert (2019), the author made an estimate of the average word length (set to 1.5) of several of the Chinese texts that were incorporated in the analysis. As the true word length is unknown, this might make the reported reading rate for Chinese somewhat unreliable. Our average word length is also somewhat shorter (i.e., 1.44, see Table 2).

most participants in this work were perhaps more proficient readers than in previous. Most likely, a combination of the factors mentioned here contributed to the high Chinese reading rate in this corpus.

Figure 1 *Boxplots of Reading Times*



Boxplots present the reading times of the first and second language of Chinese-English bilinguals. The reading times were log-transformed (*y*-axis, in milliseconds). The upper plot presents the reading times of the L1 (Chinese), while the lower plot indicates the reading times of the second language (English).

In L2, participants read at a rate of 166 wpm. The second language reading rate is similar to those observed in previous studies (139-174 wpm; Brysbaert, 2019), and is in line with the hypothesis that a lower L2 proficiency results in slower, less efficient (visual) word processing (Cop, Drieghe & Duyck, 2015; Diependaele et al., 2013).

For further analysis of the timed data, reading times exceeding 2.5 standard deviations of each participant's average in each language were considered outliers and discarded. Figure 2 shows the quantile-quantile plots of the reading times after log-transformation and outlier removal. The Lilliefors normality test (L) showed that the *p*-values of all reading measures were below .001, indicating that reading times deviated significantly from normal distributions, even after log transformations and outlier exclusion. The Pearson's Moment Coefficient of Skewness (G) showed that the distribution of all reading time measures was positively skewed (with right-side tail). These results are consistent with the GECO findings of Cop and colleagues (2017). The statistical values of the Lilliefors and Skewness tests are presented in the corresponding panels of Figure 2.

Description of Reading Times

Descriptives of the five reading time measures (i.e. FFD, SFD, GD, TRT, and GPT) are depicted in Table 3. Average reading times in L1 were statistically different from those in L2, and this for all measures. The average first fixation duration is 24.18 ms longer in L2 versus L1, and the difference in average total reading time reaches 113.01 ms. This is consistent with previous work documenting that bilinguals spend more time reading the weaker L2 (13 and 40 ms differences between L1 and L2 in the first fixation duration and the total reading time, respectively; Cop, Drieghe & Duyck, 2015). Furthermore, standard deviations of the reading time measures in L2 were greater than those in L1, indicating that L2 reading shows much greater variability than L1 reading. Interestingly, the average fixation times in L1 are highly correlated with fixation times in L2 (>.80 for all measures). This suggests that the reading speed in L1 is predictive for the reading speed in L2.

Fixation duration of the first language was similar across language groups. Except for the slight difference in GPT, Chinese readers exhibited a highly similar reading pattern compared to Dutch bilinguals and English monolinguals (Cop, Dirix, Drieghe, & Duyck, 2017), as shown in the reading time measures in Table 3. It is inconsistent with previous findings that Chinese readers have longer fixation durations than English monolinguals and European language bilinguals (e.g., Liversedge et al., 2016). It suggests that, on the one hand, there is a certain similarity in the speed of language processing despite the enormous differences between the different writing systems. On the other hand, bilinguals did not show a supposedly slower processing in their L1 due to bilingualism (but see Gollan et al.,2011) unless one assumes that language processing speed varies by language.

Interestingly, it seems that Chinese readers fixated on a Chinese word in a time comparable to previous studies on reading paragraphs (Liversedge et al., 2016; Feng et al., 2009) but somewhat faster than those on reading single sentences (Rayner et al., 2005; Yan et al., 2006). One possibility for this discrepancy between studies is that the reading patterns might be altered due to different reading materials. Readers may benefit more from an informative paragraph than an isolated one-line sentence when processing a word because, for instance, words could be more predictable in the former condition.

Table 3 *Descriptive Statistics of Five Reading Time Measures*

	Chinese bilinguals								Dutch bilinguals		
	Bilingual L1 (Chinese)			Bilingual L2 (English)			t Value (Chinese-English)	Correlation (Chinese-English)	Bilingual L2 (English)		
	M	SD	Range	M	SD	Range	t[DF]	t[DF]	M	SD	Range
First fixation duration	212.24	69.36	100-546	236.42	83.54	100-665	-12.49[29]***	0.91	222.00	74.00	101–536
Single fixation duration	211.69	68.95	100-541	238.86	82.99	100-617	-12.08[29]***	0.89	224.00	74.00	101–540
Gaze duration	229.50	90.58	100-734	312.80	161.26	100-1411	-17.68[29]***	0.86	250.00	105.00	101–877
Total reading time	260.82	133.94	100-1245	373.83	230.00	100-1584	-15.97[29]***	0.80	296.00	194.00	101–978
Go past time	325.26	265.50	100-3439	391.42	320.47	100-4230	-9.66[29]***	0.82	332.00	218.00	101–2,130

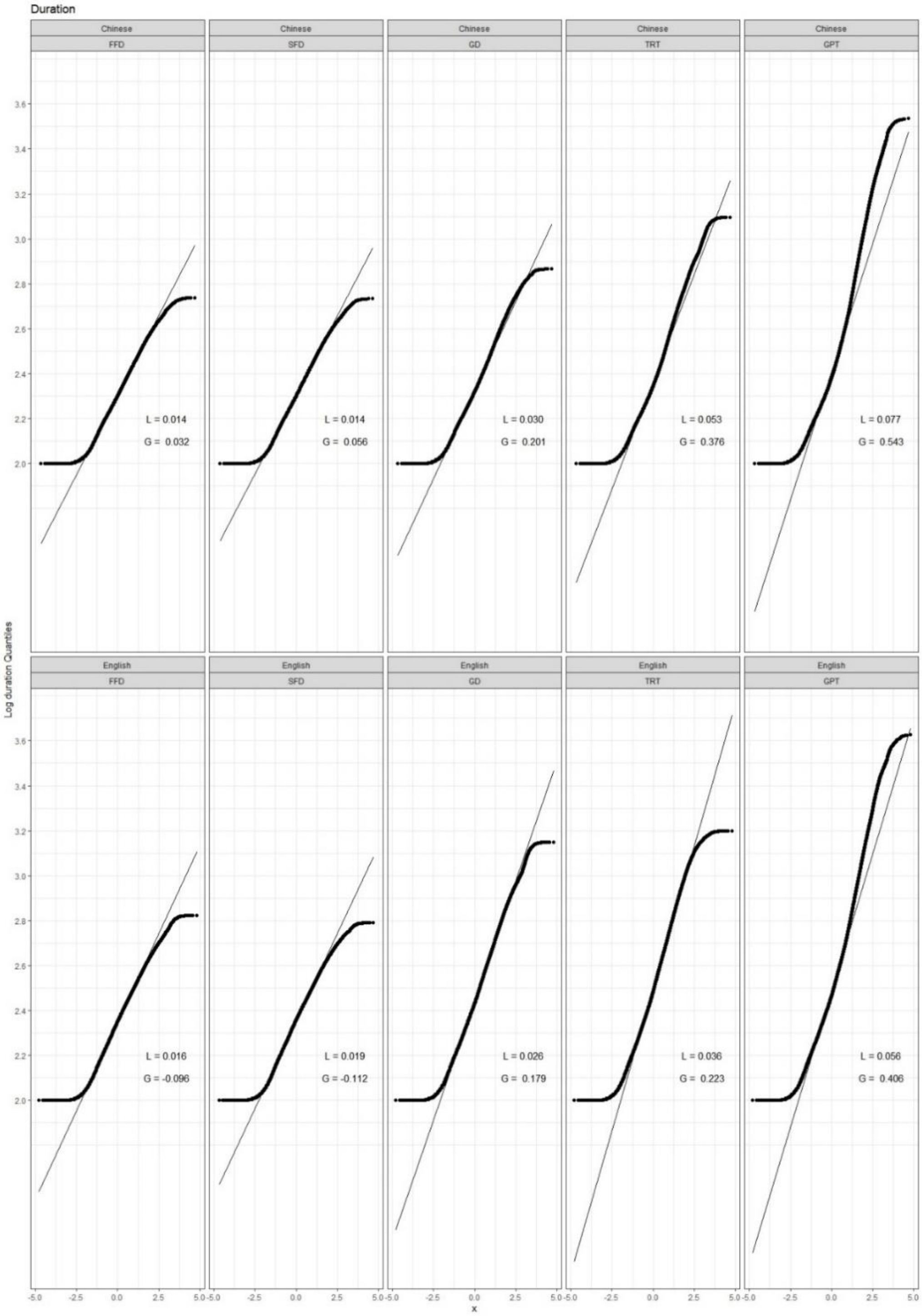
*** p <.001

Average (M), Standard Deviation (SD), and Range (minimum-maximum) of reading time measures for the first (Chinese) and second (English) languages of Chinese bilinguals and second (English) languages of Dutch bilinguals (Cop, Dirix, Drieghe, & Duyck, 2017). The t-test column (result[Degrees of Freedom]) shows comparisons between L1 and L2 reading times of Chinese-English bilinguals. The Correlation column presents the correlation between first and second language reading times in five reading time measures.

Interindividual Consistency of Reading Times

In order to test the reliability of the dataset, split-half correlations with a Spearman-Brown correction were conducted using the psych package (Version 2.1.9, Revelle, 2015). This analysis calculates the correlation between half of the participants for all stimuli in each language condition. As shown in Table 4, the consistency of the reading times was quite high and similar to those displayed in GECO (Cop, Dirix, Drieghe, & Duyck, 2017), confirming the reliability of the current corpus. However, reliability coefficients for L1 were much lower than those for L2 in this study, which is different from Dutch-English bilinguals in GECO, who showed similar values across languages. Further analysis showed that the average fixation counts and regression rates in Chinese reading were significantly lower than in English, $M_{L1} = 1.26$, $M_{L2} = 1.70$, $t = -13.20$, df = 29, $p < .001$ and $M_{L1} = 0.12$, $M_{L2} = 0.17$, $t = -6.73$, df = 29, $p < .001$, respectively, ruling out the possibility of comparatively lower consistency in Chinese reading due to the greater number of refixations and regressions.

Figure 2 *Quantile–quantile Plots of Reading Times*



Quantile–quantile plots present the reading times of the five measures in each language condition (Chinese in the upper figures and English in the lower figures). Reading times were trimmed and log-transformed. The statistical values of the Lilliefors test of normality (L) and the Pearson's moment coefficient of skewness (G) are presented in each condition. The L value corresponds to the deviations from the standard normal distribution. The higher the value, the larger the deviation. The G value corresponds to the skewness. The larger the positive value, the greater the positive skewness.

Table 4 *The Spearman-Brown Split-half Reliability Coefficients of the Five Reading Time Measures*

TYPES	Bilingual L1 (Chinese)	Bilingual L2 (English)
First fixation duration	0.69	0.75
Single fixation duration	0.7	0.89
Gaze duration	0.8	0.95
Total reading time	0.84	0.94
Go past time	0.79	0.92

The unbalanced consistency of reading times likely is an effect of smaller cross-lingual similarity between English and Chinese. Possible reasons for greater variations in Chinese reading times could be either inconsistent views of word boundaries among readers or much higher skipping probabilities in Chinese reading. The former possibility may result in various lengths of processed words, resulting in diversity in inter-individual consistency. The latter possibility will be discussed in the next section. In any case, the expected values for English confirm that this observation is due to language characteristics, and not to some unwanted participant factor (e.g., reading motivation, see also the high comprehension scores).

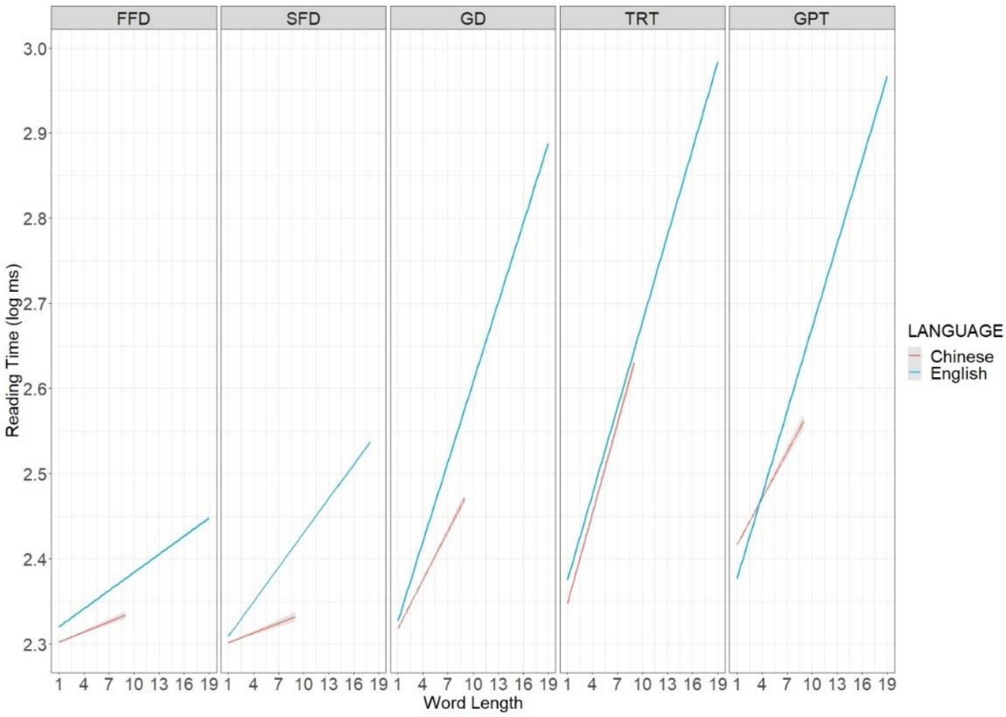
Frequency and Word Length

Frequency and word length may be the most important predictors of reading behaviour (Rayner, 2009). Participants recognise a word more quickly if it is a high-frequency word and/or a short word (Rayner & Duffy, 1986; Rayner et al., 2011). Figure 3 displays the effect of word length on five reading time measures. Although Chinese words were on average much shorter than English words, the exhibited patterns in fixation durations were similar, but larger in L2. The results seem consistent with previous work on the word length effect in Chinese (Zang et al., 2018) and in English (Rayner et al., 2011). However, it should be noted that Chinese word recognition is also affected by the number of strokes (Liversedge et al., 2014) and word frequency (Wei et al., 2013), as mentioned above. Increasing word length in Chinese generally increases the number of strokes and sometimes reduces word frequency. Thus, unlike in alphabetic languages, it is less persuasive to study Chinese word length effect without considering some language-specific factors.

Figure 4 shows the effect of word frequency on reading time measures. Fixation times decrease with increasing word frequency in both languages, consistent with previous studies of the frequency effect (Rayner & Duffy, 1986; Yu et al., 2020). The frequency effect found with words also supports the hypothesis that words might be the basic unit in Chinese reading (Li et al., 2014).

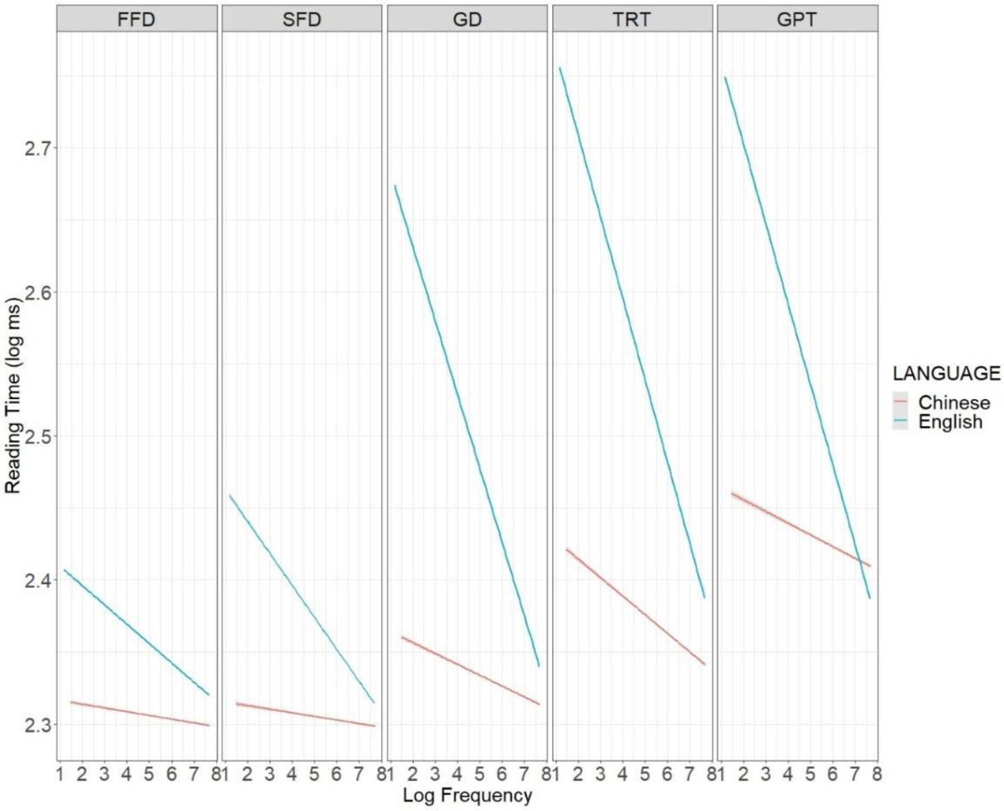
There are a number of interesting predictions about frequency effect on L1 and L2 based on the rank hypothesis (Murray & Forster, 2004). The hypothesis suggests that word recognition is a process that sequentially compares and validates an input with the stored orthographies that are organised into serial subsets or bins ranked by frequency, where the highest frequency has the highest rank. The comparison starts with the highest ranked entry in the bin, and its access speed influences reading speed. If the bins contain only words from one orthography, the frequency effects should be comparable since the lexical entries in L1 and L2 should be in roughly the same order in different bins. If the bins contain lexical entries from both languages, a larger L2 frequency effect is expected since L2 appears infrequently and thus ranks lower compared to L1. If the bins are shared only with languages of the same writing system, the frequency effect should be greater for Dutch-English bilinguals in L2 but comparable for Chinese-English bilinguals (see Duyck et al., 2008 for further discussion).

Figure 3 *Plots of the Word Length Effect on Reading Times*



It shows the word length effect on the reading times of the five measures when reading in the first (Chinese, red line) and second (English, blue line) languages. The grey shadow is the confidence interval.

Figure 4 *Plots of the Word Frequency Effect on Reading Times*



It shows the effect of word frequency on the five measurements of reading times in the first (Chinese, red line) and second (English, blue line) languages reading. The grey shadow is the confidence interval.

The frequency effect in L2 reading was much larger than in L1 reading by Chinese-English bilinguals in this work, arguing against the possibility that frequency ranked bins only contain one language, or that bins are writing system specific, unless one assumes that search speeds are language specific. Furthermore, the steeper frequency effect in L2 was consistent with previous studies investigating visual recognition (eye-tracking, Cop, Keuleers, et al., 2015; lexical decision, Duyck et al., 2008; word recognition, Diependaele et al., 2013) and language production (Gollan et al., 2008; but see Ivanova & Costa, 2008) in European language bilinguals. It shows that the frequency effect may generalise to L2 readers with a structurally different L1 writing system.

Skipping Probability

Word skipping probability is an important variable for understanding the ongoing reading process. It is affected by word length, frequency, and predictability constrained by previous information (Zang et al., 2018; Brysbaert & Vitu, 1998; Yan et al., 2006; Rayner & Raney, 1996; Rayner et al., 2005; Ehrlich & Rayner, 1981), and is a universal phenomenon across different language families (e.g., Liversedge et al., 2016; Rayner et al., 2005). Table 5 shows the average skipping probabilities of L1 and L2, while Figure 5 presents the effect of word length on skipping probability.

Table 5 *Skipping Probabilities*

	Bilingual L1 (Chinese)			Bilingual L2 (English)		
	M	**SD**	**Range**	**M**	**SD**	**Range**
First-pass skipping probability	0.72	0.07	0.56-0.83	0.34	0.09	0.16-0.49
Overall skipping probability	0.61	0.08	0.48-0.74	0.26	0.07	0.13-0.42

It presents the skipping probabilities of the first-pass (top) and the overall (bottom) for reading in Chinese (L1) and English (L2): Averages (M), Standard Deviations (SD), and Ranges.

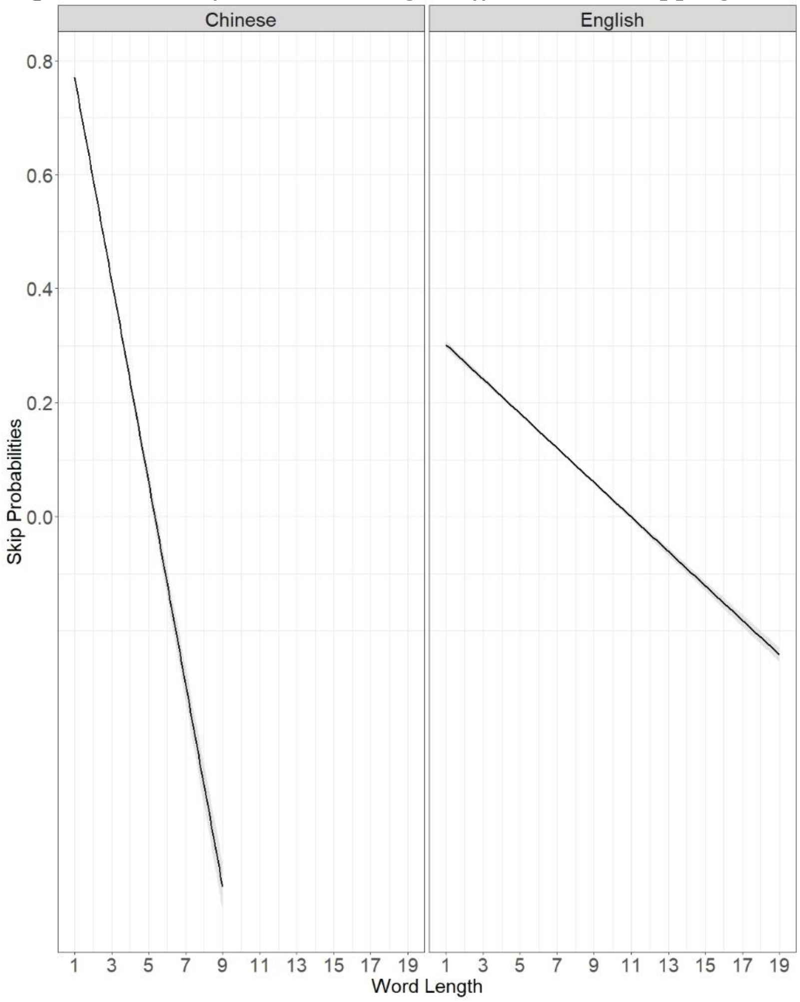
The skipping probability in Chinese reading on the first-pass was much higher than in English. Participants initially skipped about 70% of the words when reading Chinese and only about 30% when reading English. Overall, around 60% of the words have no fixation in Chinese reading and 25% in English. The skip proportion in English reading was similar to that of English monolinguals and Dutch-English bilinguals (Cop, Dirix, Drieghe, & Duyck, 2017), confirming reliability of that finding and of normal reading behaviour here. However, the skipping probability in Chinese reading is much different from previous research that did not study book reading. Previous studies reported Chinese readers skip around 3-25% of words when reading isolated sentences (Tsai et al., 2004; Rayner et al., 2005; Yan et al., 2006) and 47% when reading short paragraphs (Liversedge et al., 2016). This is well below the skipping probabilities in the current study, where readers were presented a continuing narrative rather than shorter texts. So, the trend with increasing skipping rates as text length increases holds.

Taking into account comprehension scores, it seems unlikely that Chinese readers compromised reading quality for a high skipping probability. Although their skipping rates were larger, they were significantly more accurate in Chinese than in English reading (See Table 1). They were also as accurate as Dutch-English bilinguals or English monolinguals (Cop, Dirix, Drieghe, & Duyck, 2017) in their respective native languages, while skipping rates for these two groups were only a third of the Chinese group.

The first possible explanation for this high skipping proportion in the current sample is related to the type of Chinese we employed. Whereas Simplified Chinese and Traditional Chinese are orthographically similar, the latter is sometimes more visually complex and often has more strokes for the same character (侦探 in Simplified Chinese, 偵探 in Traditional

Chinese, and DETECTIVE in English). Since visual complexity impacts fixation probability, readers are more likely to skip less complex Simplified Chinese words (Liversedge et al., 2014), as employed in the current study. Nevertheless, skipping rates in other studies using Simplified Chinese during sentence reading (e.g., Rayner et al., 2005) were still lower than in the current one.

Figure 5 *Plots of the Word Length Effect on the Skipping Probability*



It shows the effect of word length on the skipping probability when reading in the first (Chinese, left-side) and second (English, right-side) languages. The grey shadow is the confidence interval.

The second possible explanation for our findings therefore concerns predictability and top-down influences of the narrative. It was previously observed that paragraphs yield higher skipping rates (47%) than isolated sentences (3-25%). The reading material of the current study, however, is a novel with commonly used expressions and a coherent, continuing storyline. This further increases word and text predictability, beyond the values observed for paragraphs. It also illustrates the importance of basing our understanding of reading on natural texts in addition to shorter, experimental materials. Within Chinese reading, the observed values show a plausible evolution, although it still remains the case that, between languages, these values are higher than those observed for Dutch-English bilinguals and English monolinguals (Cop, Dirix, Drieghe, & Duyck, 2017), which shows that between-language differences remain important.

The third possibility concerns reading materials. Previous studies used controlled sentences composed of words with uncontroversial boundaries as test materials, resulting in an artificially longer average word length. The current work employed natural text with about 62% of single-character words (i.e. the most common word length encountered in real life). Given that short words are more likely to be skipped than long words (Zang et al., 2018), this may also have boosted high skipping probability.

Conclusion

This work presents the first eye-tracking corpus of natural reading of Chinese-English bilinguals. Considering that the majority of existing language processing models are based exclusively on alphabetic languages, this corpus is a crucial addition to the literature, as it enables examining the diversity and generalisation of these models. Following up on the success of the first GECO[7] (Cop, Dirix, Drieghe, & Duyck, 2017), primary potential research questions could be investigated by analysing data of this corpus, which is why we made it freely available online. This corpus provides data in both languages of the same group of participants. Researchers in related research fields can use this corpus to explore broad aspects, such as the reading performance of each language at different levels (e.g., word or syntactic level), the impact of systematically different L2 learning on reading in the L1 and influences in the reverse direction, eye-movement control, and L2 education.

The current paper provides a general overview of the reading performance of Chinese-English bilinguals in their two languages. Chinese bilinguals showed similar fixation durations to Dutch bilinguals and English monolinguals in their respective native languages, rather than being slower, as shown in previous studies (Liversedge et al., 2016; Gollan et al.,2011). Chinese readers also exhibited much faster reading speed and surprisingly higher skipping probability when reading the novel in their native language than was the case in previous studies (e.g., Brysbaert, 2019; Cop, Dirix, Drieghe, & Duyck, 2017; Liversedge et al., 2016). Unlike the original GECO, where the reading time consistency of the two languages is similar, this work showed somewhat lower reading time consistency in L1, yet still very high. The difference between the two language groups may be due to the disparity

[7] 152 Google Scholar citations in the first five years of publication, measured May 19th, 2022

in writing systems, for instance, the existence of word boundary demarcation rather than individual differences. Consistent with previous research, Chinese-English bilinguals spent significantly more time reading in their L2 than in L1, showing that language processing is more laborious in the less proficient language (Cop, Dirix, Drieghe, & Duyck, 2017; Cop, Drieghe & Duyck, 2015). In addition, Chinese bilinguals also exhibited larger frequency effects in L2, similar to what was observed in Dutch-English bilinguals (Cop, Keuleers, et al., 2015).

Differences with earlier isolated sentence reading studies (e.g., in skipping rates), once again, highlight the importance of natural reading materials. Including unmodified test materials is indeed effortful in Chinese reading experiments, mainly due to the unclear concept of words. However, limiting the diversity of test material to avoid controversy over word boundaries may not be ideal, as artificial test materials have limitations. Some of the well-known factors involved in the reading discussed above, such as frequency (Yan et al., 2006; Yu et al., 2020), word length (Zang et al., 2018), and the number of strokes (Liversedge et al., 2014) may show somewhat different effects in experimental studies and natural reading corpus. The results based on artificial reading materials may confuse researchers and may even lead to serious deviations in understanding the Chinese reading process (Dirix et al., 2019; Kuperman, Drieghe, et al., 2013).

Another notable finding from this work is the influence of L1 on L2 processing. This work, along with previous GECO (Cop, Dirix, Drieghe, & Duyck, 2017) data, shows that L2 performance in bilinguals depends on the language family of their L1 and L2. Compared with their Dutch-English counterparts, Chinese-English bilinguals began acquiring their L2 relatively early, reported significantly greater L2 exposure in their environment, and were as proficient L2 speakers, as measured by LexTALE (see Table 1). Still, they obtained significantly lower L2 spelling and lexical decision scores than Dutch participants and they spent longer time reading in their L2. Indeed, language proficiency consists of multiple aspects, and a single task can only examine part of them. For example, the LexTALE and lexical decision tasks measure vocabulary knowledge, and the spelling task investigates the spelling ability. Different groups may perform dissimilarly in distinct language proficiency aspects (e.g., someone with dyslexia may score high on the LexTALE but low on the spelling and speeded lexical decision tasks; for reasons, see Callens et al., 2012). The relatively lower scores shown in the speeded lexical decision task, compared to the un-speeded LexTALE, may be due to the speed-accuracy trade-off strategy (see Table 1). This strategy may be more influential for Chinese bilinguals who need more time to process their L2 than their Dutch counterparts. Collectively, these findings may indicate that the similarity between L1 and L2 affects L2 processing at the word, syntax, and even the comprehension level.

What should be noted is that although the first author (Native Chinese Speaker) has made considerable efforts to revise the Chinese translation to ensure equivalent translation while following the natural Chinese writing style, there are still differences between Chinese and English in terms of expression and writing style. Indeed, translation traces are inevitable due to the gigantic difference between the two languages. However, the difference between Chinese and English in written style may affect the reading comprehension of Chinese readers even when reading in their native language.

To conclude, this eye-tracking corpus of natural reading with its high ecological validity is an essential source for investigating actual reading performance. Although laboratory test methods may indeed aid our understanding of specific and isolated processing factors, this corpus can present the bird's eye view of the processes involved in reading, including their mutual influence and coordination, and it can shed light on potential undiscovered perspectives for further research. Compelling computational models (e.g., E-Z Reader Model, Rayner, Li & Pollatsek, 2007; CRM, Li & Pollatsek, 2020) related to reading should be able to explain the abundance of phenomena encountered in natural reading.

References

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), 1277–1287.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive psychology*, 17(3), 364-390.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, 109, Article 104047. https://doi.org/10.1016/j.jml.2019.104047.

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. *In Eye guidance in reading and scene perception* (pp. 125-147). Elsevier Science Ltd.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE*, 5, e10729. doi: 10.1371/journal.pone.0010729

Callens, M., Tops, W., & Brysbaert, M. (2012). Cognitive profile of students who enter higher education with an indication of dyslexia. *PloS one*, 7(6), e38081.

Chen, H-C, Song H, Lau W. Y, Wong K. F. E, & Tang S. L. (2003). Developmental characteristics of eye movements in reading Chinese. In C. McBride-Chang, & H-C Chen (Eds.), *Reading development in Chinese children* (pp. 157–169). Westport, CT: Praeger.

Chen, H. C., & Tang, C. K. (1998). The effective visual field in reading Chinese. *In Cognitive processing of the Chinese and the Japanese languages* (pp. 91-100). Springer, Dordrecht.

Chen, H. C., Vaid, J., & Wu, J. T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes*, 24(7-8), 967-982.

Chuang, Y. Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., & Baayen, R. H. (2021). The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior research methods*, 53(3), 945-976.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1), 204-256.

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, 49(2), 602-615.

Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (2017). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, 20(4), 747-769.

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PloS one*, 10(8), e0134008.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic bulletin & review*, 22(5), 1216-1234.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109, 193–210. doi:10.1016/j.cognition.2008.07.008

Dictionary Editorial Office of Linguistic Research Institute, Chinese Academy of Social Sciences (2016). *Modern Chinese Dictionary (7th Edition)*. Beijing, China: The Commercial Press.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66(5), 843-863.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, 5(3), 175-197.

Dirix, N., & Duyck, W. (2017). The first-and second-language age of acquisition effect in first-and second-language book reading. *Journal of Memory and Language*, 97, 103-120.

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior research methods*, 51(6), 2800-2816.

Dirix, N., Cop, U., Drieghe, D., & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(6), 887–915.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic bulletin & review*, 15(4), 850-855.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, 20(6), 641-655.

Epelboim, J., Booth, J. R., & Steinman, R. M. (1994). Reading unspaced text: Implications for theories of reading eye movements. *Vision research*, 34(13), 1735-1766.

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.

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological science*, 22(6), 829-834.

Frank, S. L., Monsalve, I. F., Thompson, R. L., & Vigliocco, G. (2013). Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior research methods*, 45(4), 1182-1190.

Fu, H. Q. (1985). *Modern Chinese Vocabulary*. Peking University Press

Ge, B.Y. (2014). *Modern Chinese Lexicology (3rd Edition)*. Beijing, China: The Commercial Press.

Chinese Proficiency Test. (n.d.) *HSK Level VI*. http://www.chinesetest.cn/gosign.do?id=1&lid=0#

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of memory and language*, 58(3), 787-814.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186–209. https://doi.org/10.1037/a0022256

Hollenstein, N., Chersoni, E., Jacobs, C. L., Oseki, Y., Prévot, L., & Santus, E. (2021, June). CMCL 2021 shared task on eye-tracking prediction. *In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 72-78).

Hollenstein, N., Rotsztejn, J., Troendle, M., Pedroni, A., Zhang, C., & Langer, N. (2018). ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Scientific data*, 5(1), 1-13.

Hollenstein, N., Troendle, M., Zhang, C., & Langer, N. (2019). ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation. *arXiv preprint arXiv:1912.00903.*

Huang, L., & Li, X. (2020). Early, but not overwhelming: The effect of prior context on segmenting overlapping ambiguous strings when reading Chinese. *Quarterly Journal of Experimental Psychology*, 73(9), 1382-1395.

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), 20–34. https://doi.org/10.1037/0096-1523.24.1.20

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta psychologica*, 127(2), 277-288.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision research*, 45(2), 153-168.

Kennedy, A., Pynte, J., Murray, W. S., & Paul, S. A. (2013). Frequency and predictability effects in the Dundee Corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, 66(3), 601-618.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12–35.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 802–823. doi:10.1037/a0030859

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *The Quarterly Journal of Experimental Psychology*, 63(9), 1838-1857.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, 66(3), 563-580.

Kuperman, V., Siegelman, N., Schroeder, S., Alexeeva, A., Acartürk, C., Amenta, S., Bertram, R, Bonandrini, R, Brysbaert, M, Chernova, D, Fonseca, S, Dirix, N, Duyck, W, Fella, A, Frost, R, Gattei, C, Kalaitzi, A, Lõo, K, Marelli, M, . . . Usual, K. A. (2022). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus (MECO). *Studies in Second Language Acquisition*, 1-35.

Laurinavichyute, A. K., Sekerina, I. A., Alexeeva, S., Bagdasaryan, K., & Kliegl, R. (2019). Russian Sentence Corpus: Benchmark measures of eye movements in reading in Russian. *Behavior research methods*, 51(3), 1161-1178.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, 44(2), 325-343.

Lexicon of Common Words in Contemporary Chinese Research Team. (2008). *Lexicon of common words in contemporary Chinese*. Beijing, China: The Commercial Press.

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), 895–913.

Li, X., Huang, L., Yao, P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology*, 1(3), 133-144.

Li, X., & Pollatsek, A. (2011). Word knowledge influences character perception. *Psychonomic Bulletin & Review*, 18(5), 833-839.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, 127(6), 1139–1162

Li, X., Gu, J., Liu, P., & Rayner, K. (2013). The advantage of word-based processing in Chinese reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 879–889

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

Li, X., & Su, X., (2022). *Lexicon of common words in contemporary Chinese Second Edition*. Beijing, China: The Commercial Press.

Li, X., Zang, C., Liversedge, S. P., & Pollatsek, A. (2015). The role of words in Chinese reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 232–244). New York: Oxford University Press. doi:10.1093/oxfordhb/9780199324576.013.14

Liu P-P, Li W-J, Lin N, Li X-S (2013) Do Chinese Readers Follow the National Standard Rules for Word Segmentation during Reading? *PLoS ONE* 8(2): e55440. doi:10.1371/journal.pone.0055440

Liu, Y. H. (2019). *Practical Modern Chinese Grammar Third Edition*. Beijing, China: The Commercial Press.

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, 147, 1-20.

Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition*, 22(3-4), 441-457.

Lu, Z.W., (1964). *The Word Formation of Chinese*. Science Press.

Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior research methods*, 50(2), 826-833.

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1046-1059

Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, 50, 940–967.

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), 375–407.

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

Miller, L. M. S., Cohen, J. A., & Wingfield, A. (2006). Contextual knowledge reduces demands on working memory during reading. *Memory & cognition*, 34(6), 1355-1367.

Mitchell, J., Lapata, M., Demberg, V., & Keller, F. (2010, July). Syntactic and semantic factors in processing difficulty: An integrated measure. *In Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 196-206).

Murray, W. S., & Forster, K. I. (2004). Serial Mechanisms in Lexical Access: The Rank Hypothesis. *Psychological Review*, 111(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721.

Pan, J., Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2021). The Beijing Sentence Corpus: A Chinese sentence corpus with eye movement data and predictability norms. *Behavior Research Methods*, 1-12.

Parshina, O. (2020). Bilingual Reading Fluency and Prediction: Heritage Language Versus Second Language (Doctoral dissertation, City University of New York).

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, 144(1), 48–76. https://doi.org/10.1037/bul0000124

Perea, M., & Acha, J. (2009). Space information is important for reading. *Vision Research*, 49(15), 1994-2000.

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), 43–59.

Pynte, J., & Kennedy, A. (2006). An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. *Vision Research*, 46(22), 3786-3801.

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.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology*, 62(8), 1457-1506.

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., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, 3(2), 245-248.

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. Wiley Interdisciplinary Reviews: *Cognitive Science*, 1(6), 787-799.

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.

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., 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., 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., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and aging*, 21(3), 448-465.

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, 27(6), 537-544.

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision research*, 47(21), 2714-2726.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 514–528.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81(2), 275–280. https://doi.org/10.1037/h0027768.

Reichle, E. D., & Yu, L. (2018). Models of Chinese reading: Review and analysis. *Cognitive Science*, 42, 1154-1165.

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.

Revelle, W. (2015) psych: Procedures for personality and psychological research. Northwestern University, Evanston, Illinois, USA. Retrieved from https://CRAN.R-project.org/package=psych

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in cognitive sciences*, 7(11), 489-493.

Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., Ahn, H., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Fonseca, S., Dirix, D., Duyck, W., Fella, A., Frost, R., Gattei, C., Kalaitzi, A., Kwon, N., Lõo, K., . . . Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods, 1-21*.

Su, Y. F., & Samuels, S. J. (2010). Developmental changes in character-complexity and word-length effects when reading Chinese script. *Reading and Writing*, 23(9), 1085-1108.

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(3), 761–775.

Tsai, J. L., Lee, C. Y., Tzeng, O. J., Hung, D. L., & Yen, N. S. (2004). Use of phonological codes for Chinese characters: Evidence from processing of parafoveal preview when reading sentences. *Brain and language*, 91(2), 235-244.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological science*, 20(8), 923-927.

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of memory and language*, 39(3), 458-483.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

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.

Whitney, C. (2011). Location, location, location: How it affects the neighborhood (effect). *Brain and Language*, 118, 90–104. doi:10.1016/j.bandl.2011.03.001

Wilkinson, G. S., & Robertson, G. (2006). *Wide Range Achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.

Woolley, G. (2011). Reading comprehension. *In Reading Comprehension* (pp. 15-34). Springer, Dordrecht.

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. *Quarterly Journal of Experimental Psychology*, 63(4), 705-725.

Yang, H. M., & McConkie, G.W. (1999). Reading Chinese: Some basic eye-movement characteristics. In H.-C. Chen (Ed.), *Reading Chinese script: A cognitive analysis* (pp. 207–222). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Yen, N. S., Tsai, J. L., Chen, P. L., Lin, H. Y., & Chen, A. L. (2011). Effects of typographic variables on eye-movement measures in reading Chinese from a screen. *Behaviour & Information Technology*, 30(6), 797-808.

Yu, L., Liu, Y., & Reichle, E. D. (2020). A corpus-based versus experimental examination of word-and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of Experimental Psychology: General*.

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2018). Investigating word length effects in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 44(12), 1831–1841. https://doi.org/10.1037/xhp0000589

Zhang, S., & Perfetti, C. A. (1993). The tongue-twister effect in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1082–1093. https://doi.org/10.1037/0278-7393.19.5.1082

Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics*, 36(3), 277-284.

CHAPTER 3. Word Frequency Effects Differ Across Writing Systems: The Word and Character Frequency Effect in Chinese Reading[8]

Abstract

The word frequency effect (FE), where high-frequency (HF) words are processed faster than low-frequency (LF) ones, has been extensively studied in alphabetic writing systems. The effect has also been observed in Chinese reading, a language that differs enormously from alphabetic languages, not only in appearance but also in the nature of the words. In the present study, we investigated the word and character FEs by analyzing reading data from an eye-tracking corpus in which participants read an entire novel (GECO-CN). The results show that as character frequency in Chinese increases, the facilitative word FE tends to flatten or even reverse, and vice versa. These findings reveal that even if there is a similar effect pattern of the word FE between the Chinese and alphabetic languages, the underlying processes are actually different. In addition, it also provides a plausible explanation for the inconsistent character effects found previously.

Keywords: Frequency effects, Chinese reading, Eye-movements

[8] Sui, L., Woumans, E., Duyck, W., Dirix, N., (submitted) Word Frequency Effects Differ Across Writing Systems: The Word and Character Frequency Effect in Chinese Reading. *Language Learning*.

The word FE is the phenomenon where processing speed is affected by the frequency of word occurrence, with faster processing of HF words (Rayner & Duffy, 1986). It is one of the most potent effects observed in language research to date, independent of other confounding factors, such as word length, number of phonemes, age of acquisition, and orthographic neighborhood (Brysbaert et al., 2018). Our current understanding of the FE is mainly based on findings from alphabetic languages, whose words are visually salient by spatial separation (e.g., Rayner & Duffy, 1986). The question is whether the same applies to nonalphabetic writing systems, such as Chinese, which is qualitatively distinct from alphabetic languages in both writing and pronunciation. Early efforts have examined the FE of Chinese words using categorical designs (sets of HF and LF word stimuli), showing a negative correlation between processing speed and word frequency, similar to that found in alphabetic languages (Ma et al., 2015). However, a peculiar convention of Chinese words, namely the character property, poses a great challenge to the feasibility and reliability of studying the Chinese word FE separate from character effects.

Chinese words are composed of a limited number of orthographically independent characters. The words themselves, however, are not salient in sentences due to the lack of interspace for demarcating word boundaries. This difference from alphabetic languages results in many disparities, namely, larger influences of character properties on reading. Some research even suggests that characters are the basic units in Chinese reading (e.g., Hoosain, 1992). Although more evidence suggests that word properties influence Chinese reading to a higher degree, and that words are actually the fundamental reading process units (e.g., Yang et al., 2012), it remains true that characters also appear to affect word recognition (e.g., Yan et al., 2006).

Yet, some of their specific effects remain unclear and inconclusive due to inconsistent results found in limited empirical evidence (e.g., character FE, Li et al., 2014; Yan et al., 2006). Nevertheless, a number of Chinese reading models also assume a role for character processing in word recognition (e.g., Li & Pollatsek, 2020). If language-specific factors (i.e., character properties) influence word identification, there should be at least some differences compared with alphabetic languages in the underlying reading process, possibly even accompanied by distinct cognitive mechanisms. For instance, as Chinese words are composed of a limited number of characters, a word in itself may not appear often, but its characters may have frequent occurrences. The Chinese word FE might then be absent or interact with character frequency.

To our knowledge, none of the previous work has questioned whether the similar word frequency phenomena observed in Chinese and English reading might stem from fundamentally different reading processes. Indeed, the word FE found in alphabetic language is so robust and reliable that it is assumed to be universal across all world languages. And, indeed, previous studies on Chinese reading found similar results (e.g., Ma et al., 2015). However, differentiating between and understanding Chinese character and word FEs can advance our specific knowledge of Chinese reading, inspire further research and interpretation, and refine models that have already been and are being developed to account for Chinese reading (e.g., Li & Pollatsek, 2020). Importantly, it is also crucial to assess the specificity and universality of the word FE across languages in light of theories and models which apply the effect to all writing systems.

Existing evidence for Chinese FEs in reading is mainly based on single words (e.g., Mattingly & Xu, 1994) and isolated controlled sentences manipulating the target word (e.g., low-constraint sentences which have avoided word ambiguity, sometimes even used the same sentence frame differing only in the target word. e.g., Cui et al., 2021). Yet, word identification performance under these conditions cannot adequately and fairly reflect the performance in reading coherent paragraphs of a certain length, which are the most common reading materials in everyday life (Dirix et al., 2019). This is important as text-level context generates semantic, grammatical, and lexical expectations that may influence the reading of subsequent words. Moreover, the difference in reading performance between artificial and natural reading conditions can be expected to be more severe in Chinese reading, where word segmentation is an additional process to be performed, compared with in alphabetic language where word segmentation is entrenched. This work, therefore, aims to clarify the issues of character and the word FEs in natural reading by using frequencies as continuous variables.

The Chinese writing system

Chinese is a logographic language composed of box-shaped characters constructed by a number of strokes under certain rules, different from alphabetic languages that employ letters. Character complexity varies based on the number of strokes, with more strokes resulting in greater visually complex and longer recognition times (Liversedge et al., 2014). Chinese words are composed of a limited number of characters (about 6,000 characters, Li et al., 2022), most of which can be part of different words and appear in different positions or constitute a word. The average length of Chinese words is comparatively shorter than English words. One- and two-character words represent 97.2% of written words in Chinese (Li & Pollatsek, 2020).

Modern Chinese texts are written vertically from left to right, similar to English. However, different from many alphabetic languages, there are no physical cues between words in written Chinese (e.g., blank spaces). In contrast, characters are spatially discrete from each other. The absence of delineated word boundaries makes the segmentation of words less apparent than characters. Moreover, the concept of word did not appear in Chinese until the twentieth century (e.g., Sui et al., 2022). Even now, most dictionaries are character and not word dictionaries. Furthermore, the less well-defined word concepts and the lack of clear word boundaries have sometimes led Chinese readers to have divergent segmentation decisions on identical strings, resulting in disagreement on the number of characters that constitute a word (Liu et al., 2013).

However, although Chinese is an unspaced script with no explicitly marked word boundaries, words are considered as the basic processing units according to the existing findings (Li et al., 2013, 2014; but see Hoosain, 1992). Some studies have shown that disrupting word processing, such as inserting blank spaces within words or between characters (Bai et al., 2008), masking one of the characters that constitute a word using the moving window paradigm (Li et al., 2013), or separating a word and placing the characters on different lines (Li et al., 2012) can interfere with reading performance. In contrast, spacing between words elicited a similar (Bai et al., 2008) or a facilitative performance in reading times (e.g., Oralova & Kuperman, 2021) or in ambiguous string conditions (Hsu & Huang, 2000) compare to conventional, unspaced characters. Furthermore, word properties (e.g., word length and frequency) explain more variance in reading than character properties (e.g., character frequency; Li et al., 2014).

The above evidence collectively demonstrates the importance of words as the primary processing units for efficient sentence reading (e.g., Li et al., 2014; Yang et al., 2012). Still, the constituents of words appear to affect Chinese word recognition (Yan et al., 2006). As mentioned, characters, most of which can act independently, are the components of Chinese words, somewhat similar to the single units of a compound word in alphabetic languages. Whereas in alphabetic languages, some evidence shows that the frequency of letters (e.g., New & Grainger, 2011), syllables (e.g., Conrad et al., 2009), and single units of a compound word (e.g., Kuperman et al., 2008) affects word recognition.

Hence, a logical assumption is that, despite the disparity between the writing systems, the morphological constituents of words in Chinese are activated in conjunction with word processing (e.g., Yan et al., 2006; Tse & Yap, 2018), as found for compound words in alphabetic languages (e.g., Kuperman et al., 2008, 2009). Existing research has confirmed this assumption, showing that words with LF single constituents take longer to read than those with HF ones (e.g., Hasenäcker & Schroeder, 2019; Zhou et al., 2018).

Word and character FEs in eye movements

Eye movement measures are employed to elucidate the underlying processes in word reading. The two basic components of eye movements are saccades and fixations (Rayner, 2009). The former refers to the action of quickly moving the eyes to a point, while the latter refers to the moment when the eyes concentrate on a point. Moreover, several timed measures can be calculated, based on the duration of fixations: a) first fixation duration (FFD), the duration of the first fixation on a word; b) single fixation duration (SFD), the duration when the word is fixated only once; c) gaze duration (GD), the duration of all fixations on a word before the next (right-side) word in a sentence is fixated; and d) total reading time (TRT), the duration of all fixations on the word. Measures such as FFD and GD belonging to first-pass time are generally considered "early" measures and are supposed to reflect processes at the initial stages of word identification. In contrast, measures such as TRT that involve the second-pass time are referred to as "late" measures (Boston et al., 2008). Additionally, there is a dichotomous measure that refers to the probability of skipping a word in the first-pass reading, the skipping probability.

Chinese word properties seem to affect reading in similar ways with other highly different scripts when looking at eye movement measures. Evidence has shown that influential factors such as word frequency (e.g., Ma et al., 2015; Slattery et al., 2007), word length (Zang et al., 2018; Rayner et al., 2011), and predictability (Rayner et al., 2005, 2011) in Chinese have similar effects on both number and durations of fixations as in alphabetic language reading. Readers make shorter fixations on more frequent (Liu et al., 2016; Wei et al., 2013), shorter (Zang et al., 2018), and highly predictable words (Rayner et al., 2005).

However, the unique features of Chinese orthography could have substantial influences on reading processes. As mentioned, Chinese words are composed of a limited number of characters. The frequency of a word may be low, but its constituent characters may be highly frequent. If HF components accelerate reading times for LF words, then the difference in reading times between HF and LF Chinese words may be negligible. A crucial question is whether the robust word FE found in alphabetic language can also be observed in a completely disparate script. Below, we will elaborate on the empirical findings with regard to word and character frequency in Chinese reading.

The word FE in Chinese reading

Early efforts have investigated the word FE in Chinese sentence reading by means of eye-tracking (e.g., Liversedge et al., 2014; Zhou et al., 2018). These studies primarily examine the performance of target (content) words embedded in isolated low-constrained sentences. Target words are often categorized into 'HF' and 'LF', and controlled for some influential factors like length and complexity (the number of strokes in a character, e.g., Yu et al., 2021). Most of them found significantly shorter fixation durations for HF words (Cui et al., 2021; Liu et al., 2019; Ma et al., 2015; Li et al., 2014; Wei et al., 2013; Yu et al., 2021; Yan et al., 2006; Zhou et al., 2018), but there are a few notable deviations from these results (Liversedge et al., 2014).

When reading one-character target words embedded in sentences, the word FE seems less consistent. In Liversedge et al. (2014), the main effect of word frequency was not significant in fixation durations (i.e., SFD, FFD, and GD) but did show up in the skipping probability. HF words are thus processed as fast as LF words but are skipped more often. The interaction between word frequency and character complexity was significant in SFD and FFD, but not in GD and the skipping probability, showing that words with LF and high character complexity are processed slower in early stages of word recognition.

In contrast, Zang et al. (2016) observed the word FE in GD and the skipping probability, even using the same test material as Liversedge et al. (2014). Readers spent more time gazing at LF one-character words and skipped them less often. Zang et al. (2016) also found a significant interaction between word frequency and complexity, showing shorter fixations for HF, less complex words. Yet, the interaction was observed in GD but not in FFD and SFD, contrary to that of Liversedge et al. (2014). The discrepancy between the two studies may be due to low statistical power, with 2880 datapoints of young adults recruited in the former study compared to 5120 in the latter.

When reading a multi-character word in a sentence, it was consistently found across eye-tracking measures that HF words are recognized faster than LF words (in two-character words: Liu et al., 2016; Liu et al., 2019; Ma et al., 2015; Yan et al., 2006; Yu et al., 2021; in three-character words: Zhou et al., 2018; for all word lengths: Li et al., 2014; Yu et al., 2021). Although Wei et al. (2013) failed to find a reliable word FE for two-character words in FFD (i.e., only marginally significant in the item analysis and non-significant in the subject analysis), the authors do observe it in GD.

In sum, the word FE appears to be robust in multi-character words in Chinese reading, with a pattern similar to that of alphabetic languages (e.g., Li et al., 2014). Yet, since word and character frequencies inevitably have certain collinearity, it remains to be determined whether the observed word FEs indeed rely on the word frequency, or if character frequency also plays a role. To study whether and how characters influence word recognition, a few studies have investigated two-character words which can tease apart the influence of character and word frequencies, in comparison with single character words, on word recognition.

The character FE in Chinese reading

Some existing research has investigated the effect of character frequency on eye movement behavior by studying two-character target words embedded in a single sentence (Yan et al., 2006; Yu et al., 2021) and some even with the same sentence frame (Cui et al., 2021; Cui et al., 2013; Ma et al., 2015). In addition to word frequency, these studies also manipulated first (C1) and second character (C2) frequency (Yan et al., 2006; Cui et al., 2021) or C1 frequency only (controlling for C2 frequency; Yu et al., 2021). Surprisingly, different results emerged among the few existing studies, with some observing a facilitative (Yan et al., 2006; also see Mattingly & Xu, 1994; Tse & Yap, 2018 in lexical-decision task) or an inhibitory FE of the C1 (Yu et al., 2021; Cui et al., 2021). Others found no influence of character frequency (Li et al., 2014; Ma et al., 2015; Cui et al., 2013).

Yan et al., (2006) observed significantly shorter fixation times in words with frequent C1s in FFD, SFD, GD but not in TRT. The frequency of the C2 did not have an effect ($p > .05$). The character FE was negligible at HF words but appeared at LF words. The interaction between word and C1 frequency was significant in FFD but not in SFD, GD and TRT, and between word and C2 frequency was significant in GD, but a hint in FFD. In contrast, Cui et al. (2021) found a reversed, inhibitory character FE when analyzing LF two-character words, showing that the higher the frequency of the C1, the longer the fixation duration on the entire word. However, the C2 frequency did not affect fixation duration on the word, consistent with what Yan et al. (2006) observed.

Yu et al. (2021) also observed an inhibitory character FE when analyzing target words, with longer FFDs on words with HF initial characters. Although a facilitative character FE emerged in TRT when analyzing all words in a sentence, the authors explained that it could be due to uncontrolled collinearity. Moreover, the interaction between word and character frequencies of target words was not significant, contrary to the previous results (e.g., Yan et al., 2006). Yu et al., (2021) explained that the discrepancy between the reported character FEs could be due to the predictability of upcoming words. High-constraint sentences can narrow the number of lexical candidates to those compatible with the sentence context, thereby attenuating lexical competition and leading to the facilitative character FE (Yu et al., 2021). It should be noted, however, that the cloze predictability in both their studies was negligibly low (M = 0.1%, estimated by 80 participants in Yu et al., 2021, and M = 1.5%, estimated by 10 participants in Yan et al., 2006).

Finally, some studies did not observe a reliable character FE (Cui et al., 2013; Li et al., 2014; Ma et al., 2015). Notably, Cui et al. (2013) show that although the C1 of a word had no main effect on word fixation duration, fixations on HF C1s were shorter than those on LF ones when sharing the same C2. Furthermore Ma et al. (2015) found that the fixations of the pre-target word decrease with the increase of the C1 frequency of the target word. In summary, the existing controlled experiments show that the word FE is only reliable in multi-character words, not in single-character words, whereas the character FE varies in two-character words. The question remains whether these phenomena can be replicated in natural reading where word segmentation is required and may affect word recognition.

Word segmentation during Chinese Reading

Chinese word recognition in sentence reading may be affected by word segmentation, thus affecting recognition performance (e.g., how character and word frequencies affect word recognition). As stated, Chinese words are considered as the critical units determining reading efficiency, and Chinese readers are accustomed to reading unspaced scripts. Segmenting sentences that do not carry word boundary information into words is essential for word identification. However, word segmentation in Chinese reading is far more than segmenting the contiguous words into individuals, as Chinese characters can appear at various positions in words. That is, a character in text strings may act as a one-character word or form a word with its predecessor(s) or successor(s). As a result, a continuous text can sometimes be segmented into words in several ways.

To process multiple characters in a row, readers may need to process the un-fixated character to the right of the fixation point to some extent, possibly benefiting from sufficient parafoveal processing (Yang et al., 2012; see Ma et al., 2015 for parafoveal words affecting reading performance). They can then decide, contingent on word knowledge and sentence context, how to segment into words (Huang & Li, 2020). Word segmentation in unspaced Chinese text has been studied using spatially overlapping strings (Huang & Li, 2020; Ma et al., 2014), which usually have three characters. The intermediate character can form two (distinct) words with the first and third character, respectively. Evidence shows that readers' performance is more likely to be affected by the properties of the possible words that constitute the overlapping ambiguous string (Huang et al., 2021). If the activated word candidates appear more frequently, yet are implausible to the sentence context, participants are more likely to spend a longer time fixating on the ambiguous part (Ma et al., 2014).

These results can be well explained by the word segmentation hypothesis proposed by the Chinese Reading Model (CRM; Li & Pollatsek, 2020). The model demonstrates that all characters (e.g., ABC) within the perceptual span are processed in parallel, activating all the words they can compose (e.g., AB and BC). Since some characters are themselves a word, single-character words are also activated. The activated words then compete with each other for selection (also see Ma et al., 2014; Li et al., 2009). Multiple-character words (e.g., AB) have advantages over the words that constitute them (e.g., A and B), as they receive feedforward activation from all their characters (e.g., A and B), whereas embedded words (e.g., A or B) receive only from themselves. Additionally, the frequency of activated words affects competition. HF words are more likely to win the competition. Once a word unit wins the competition, word segmentation occurs concurrently with recognition.

Interestingly and noteworthy, one could assume that character properties may cause multiple influences on the reading times of Chinese words based on above hypothesis. That is, all activated characters activate and facilitate identifying their constituent words (including one-character word). The higher the frequency of a character, the stronger the activation of the candidate words, and the less time it takes to retrieve the words. Yet, the stronger the activation of a one-character word, the more it interferes with the target (multi-character) word and the longer it takes to process at this stage. Although these model assumptions are not yet appropriately empirically supported and need further verification and investigation, it could indicate that the character FE and even word recognition processes are influenced by word segmentation.

The present study

Previous studies on Chinese visual word recognition have observed facilitatory word FEs in multi-character words, inconsistent patterns in one-character words, and diverse character FEs (facilitatory, inhibitory, or even no effect). Yet, the existing research mainly consists of small-scale controlled studies, often with arbitrary categorizations of the frequency variables. The reported eye-tracking studies above mainly investigated target words embedded in isolated, controlled sentences, limited both in the number of stimuli and contextual sentence diversity. Such manipulations severely narrow the variations that occur naturally in written language and may not provide a comprehensive picture of the word and character FEs or their interaction with other word characteristics.

Inspired by the existing evidence, the theoretical inferences, and the above reasons, this study aims to investigate the character and word FEs in Chinese natural reading. The prerequisite for understanding the word FE is to know how character properties affect word reading performance. Thus, firstly, we will elucidate the effects of character frequency on word recognition in natural reading and hope to take a step toward clarifying their discrepant results in previous research. Secondly, we will investigate whether the character effect is independent of or interacts with word frequency. If this language-specific factor in Chinese is related to word frequency, and even if a similar effect occurs in Chinese and alphabetic scripts, the underlying processes of the word FE could be qualitatively different. Thirdly, we will examine whether there is a word FE, independent of character frequency, in Chinese natural reading. Importantly, this work will be the first to assess the continuous effects of frequency of both word and all its constituent characters in Chinese text reading, rather than using dichotomous frequency categories (also see Tse & Yap, 2018 for lexical-decision tasks).

To assess continuous effects of character and word frequency in natural reading, we will employ data from the Chinese Ghent Eye-tracking Corpus (GECO-CN; Sui et al., 2022), a high-quality corpus with over a million datapoints. It provides the statistical power to reliably detect (minimal) effects and interactions, and can help gain further insight into word and character FEs. Additionally, the corpus entails data from readers processing the content of a continuous narrative (i.e., a fiction novel). Furthermore, GECO-CN has a diverse range of word stimuli, and thus a wide range of word and character frequencies that can be investigated as continuous predictors. It can minimize the probability that observed phenomena were due to stimulus selection or reduced statistical power, reliability, or inappropriate rejection of the null hypothesis due to categorizing continuous variables (Balota et al., 2004). For more details on the Method (i.e., Participants, Materials, procedure, and Analysis), please see the APPENDIX B.

Results

First fixation duration

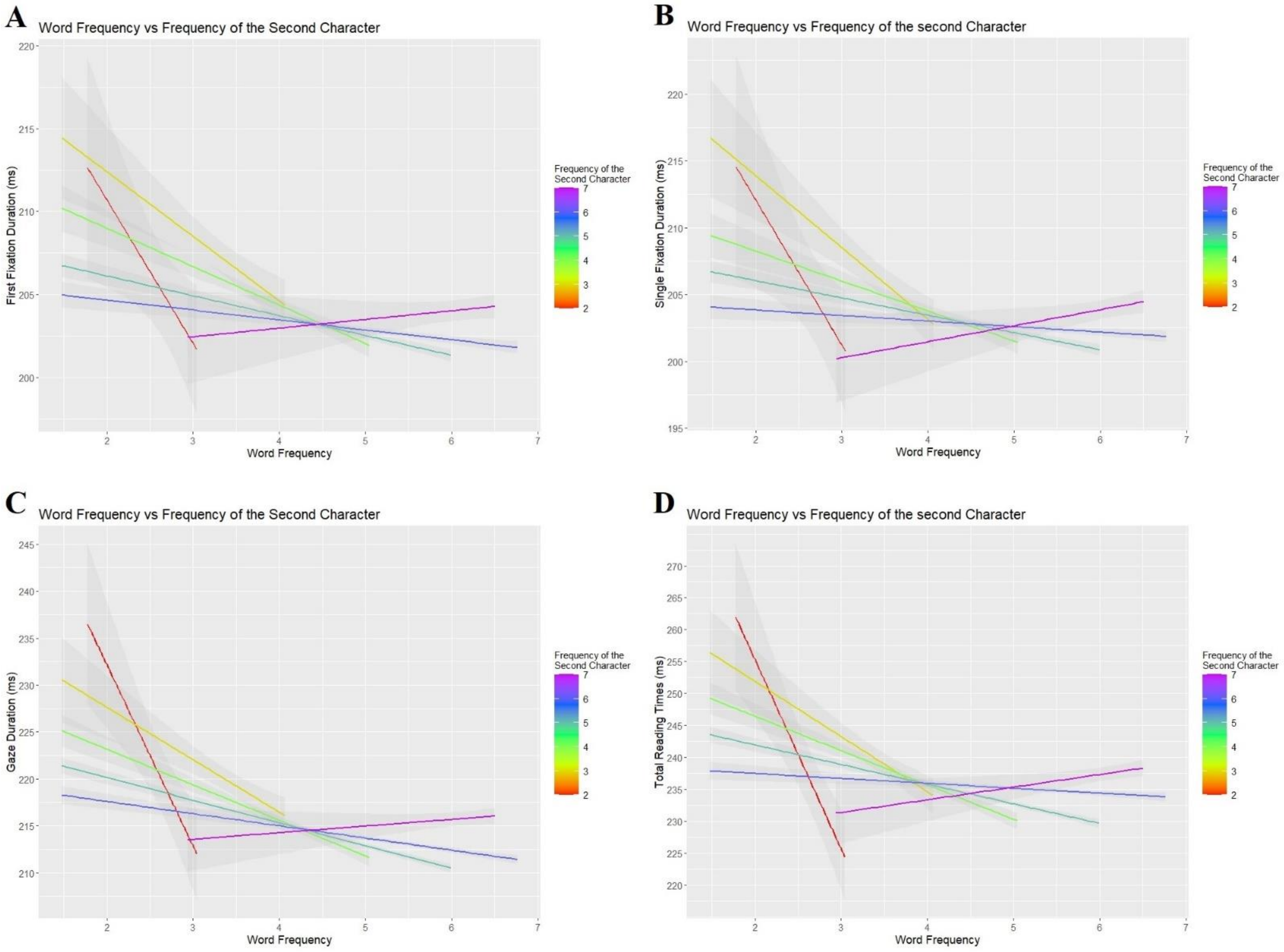
The main effects of repetition and C2 complexity were significant (see Table 1). Words with fewer occurrences or complex C2s take longer to process. The main effects of word and C2 frequencies were also significant. At the reference levels, the higher the word frequency, or the lower the C2 frequency, the shorter the FFDs. Yet, the significant interaction between the two effects indicated that the pattern and magnitude of word FEs changes with increasing C2 frequency and vice versa (see Fig 1.A and Fig 2.A). The FFD on HF words is shorter than on LF words when the C2 frequency is less than 6.19 (see Table A2 and Fig 1.A). However, the word FE is absent when the C2 frequency is greater than this value. Likewise, when word frequency is less than 2.25, reading times decrease with increasing character frequency, showing a facilitative character FE (see Fig 2.A). The FE disappears afterward until word frequency is up to 4.23. The direction of the character FE becomes reversed, with a longer FFD for HF C2s with increasing word frequency.

Table 1 *Analyses of Fixation-Duration Measures*

	Predictors	**Estimate**	**Std. Error**	**t value**	**p value**	**VIF**
First Fixation Duration	(Intercept)	4.648868	0.012514	371.483	<.001***	
	Word Frequency	-0.003585	0.001084	-3.306	<.001***	1.685
	Second-Character Frequency	0.005209	0.001508	3.454	<.001***	1.873
	Second-Character Strokes	0.001829	0.000281	6.516	<.001***	1.182
	Repitation	-0.000126	0.000045	-2.824	.005**	1.350
	Word Frequency:Second-Character Frequency	0.004129	0.001112	3.713	<.001***	1.217
Single Fixation Duration	(Intercept)	4.6485423	0.01265	367.473	<.001***	
	Word Frequency	-0.004719	0.0011455	-4.119	<.001***	1.491
	Second-Character Frequency	0.0052607	0.0016748	3.141	.002**	1.838
	Second-Character Strokes	0.0018508	0.0003127	5.919	<.001***	1.174
	Word Frequency:Second-Character Frequency	0.0037987	0.0012122	3.134	.002**	1.139
Gaze Duration	(Intercept)	2.700000	0.003767	716.619	<.001***	
	Word Frequency	-0.001530	0.000375	-4.076	<.001***	1.932
	First-Character Frequency	-0.001115	0.000444	-2.511	.012*	1.509
	Second-Character Frequency	0.002270	0.000499	4.553	<.001***	1.958
	Second-Character Strokes	0.000520	0.000091	5.731	<.001***	1.181
	Repitation	-0.000040	0.000014	-2.806	.005**	1.351
	Word Frequency:Second-Character Frequency	0.001793	0.000360	4.985	<.001***	1.217
Total Reading Time	(Intercept)	2.158000	0.002300	938.315	<.001***	
	Word Frequency	-0.000880	0.000229	-3.840	<.001***	1.751
	First-Character Frequency	-0.000863	0.000285	-3.029	.002**	1.507
	Second-Character Frequency	0.001493	0.000318	4.696	<.001***	1.933
	Second-Character Strokes	0.000230	0.000058	3.968	<.001***	1.175
	Word Frequency:Second-Character Frequency	0.001111	0.000224	4.973	<.001***	1.141

*p < .05. **p < .01. ***p < .001 (*p* values were calculated using *lmerTest* package)

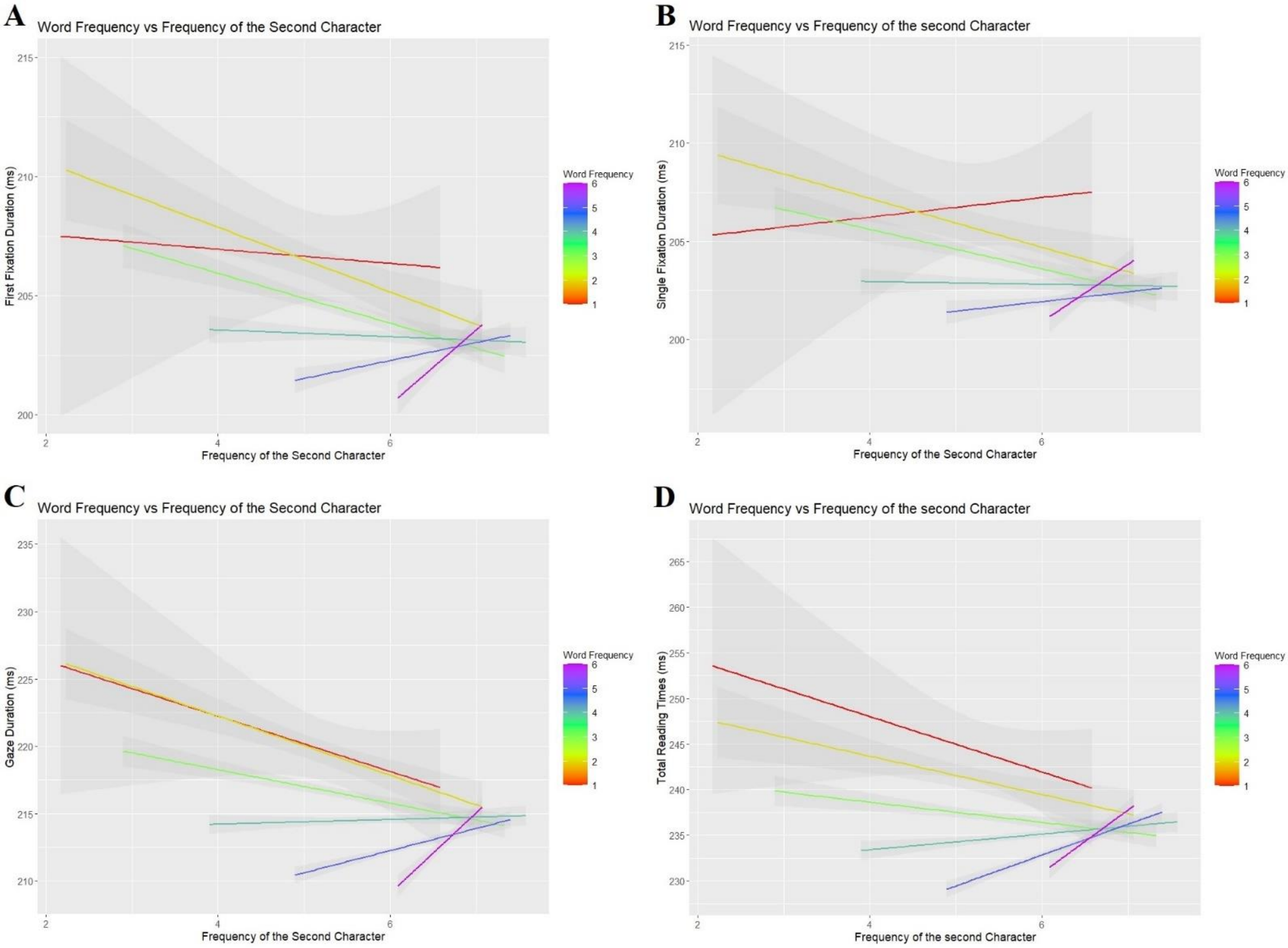
Figure 1. Plots of the effect of second-character frequency on word frequency in predicted first fixation duration (A), single fixation duration (B), gaze duration (C), and total reading times (D). The grey shadow is the confidence interval.



Single Fixation Duration

The SFD pattern was similar to that of FFD (see Table 1). Except for the repetition, the word and its C2 FEs, and their interaction, and the effect of the complexity of the C2 was still present. Words with fewer strokes have shorter SFD, while the FE of the word or C2 depends on the other. The word FE shows a negative pattern when the C2 is less than 6.46 and disappears when greater than it (see Table A2 and Fig 1.B). Similarly, when the word frequency exceeds 4.26, there is an inhibitory character FE (see Fig 2.B). The effect disappears when the word frequency is less than 4.26. Although character frequency is visually facilitative trended when word frequency is low, it is not significant.

Figure 2. Plots of the effect of word frequency on second-character frequency in predicted first fixation duration (A), single fixation duration (B), gaze duration (C), and total reading times (D). The grey shadow is the confidence interval.



Gaze Duration

At the word level, the main effects of frequency and repetition were significant (see Table 1). Repetition has a conventional facilitative influence on GD. At the character level, the main effects of the C1 frequency, the C2 frequency, and the C2 complexity were significant. The higher the frequency of the C1 or the fewer strokes the C2 has, the faster the word is processed. The effect of word frequency on fixation times is influenced by the C2 frequency, as their interaction was significant.

The GD decreased with increasing word frequency when the C2 frequency was less than 6.28, after which the word FE was absent until it reached 7.45 (see Table A2 and Fig 1.C). When the character frequency was above 7.45, the inhibitory word FE appears with longer GD for HF words, different from that in FFD and SFD. Likewise, when the word frequency was less than 2.73, GD decreased with increasing C2 frequency (see Fig 2.C). In contrast, when it exceeded 4.08, words with a HF C2 took longer to read than those with LF characters. The character effect disappeared when word frequency was at a moderate level (between 2.73 and 4.08).

Total Reading Time

The effects of word frequency, C2 frequency, their interactions, and C2 complexity were significant, consistent with the three measures described above (see Table 1) and had similar effect patterns. Furthermore, there was a statistically significant facilitative FE of C1, as in GD. Similar to those found in GD, facultative, flattening, and inhibitory word FEs were observed when the C2 frequency was below 6.24, between 6.24 and 7.31, and above 7.31, respectively (see Table A2 and Fig 1.D). Also, the character FE appears only when word frequency is below 2.73 and above 4.08, showing facilitative and inhibitory trends, respectively (see Fig 2.D).

Discussion

This study investigated the impact of word and character frequencies on eye movements in natural Chinese reading. Our first goal was to understand the effect of character frequency on visual word processing. The results showed that the C1 has a facilitative effect while the C2 has an overall inhibitory effect. Our second goal was to investigate potential interactions between character and word frequencies. We found an interaction between word and C2 frequencies, confirming the assumption that the underlying processes and manifestations of the word FE in Chinese reading differ from alphabetic languages. Our third goal was to study whether the word FE was present in natural reading after accounting for character frequency. Evidence shows that the facilitative word FE persists. In the next section, we address the word- and character-level results in detail and discuss their theoretical relevance for Chinese reading.

Character FE

First character

Our first key finding was the effect of C1 frequency on word processing times. Word fixation duration decreased as the C1 frequency increased. The facilitative FE we found seems consistent with Yan et al. (2006) but contrary to what was reported by Cui et al. (2021) and Yu et al. (2021). Although previous studies obtained inconsistent FE patterns of C1, when only analyzing the target word, they all observed it in FFD but not in TRT. Contrary to previous experiments, our findings demonstrate that the C1 frequency affects the early and late stages of the word recognition process, but not at the very early stage, as we found it in GD and TRT but not in FFD and SFD.

One possible reason for not having detected the C1 FE in the fairly early stage could be due to parafoveal processing. As illustrated, segmenting continuous text into words is necessary for reading in Chinese. To decide on word boundaries, readers may benefit from the parafoveal processing on the un-fixated character to the right of the fixation point (Yang et al., 2009). The C1 of a word is likely to have been processed to some extent in the last fixation, affecting its influence on the current fixation. This postulation is supported by previous findings in which the lexical properties of the subsequent word affect the current word processing (Li et al., 2014). It could explain the absence of a) the C1 FE in the very early measures; b) the C1 complexity effect and the interactions between the C1 and word frequencies or C1 complexity throughout the time course of word processing in this work (see Table 1).

What differentiates previous work from the current study regarding C1 processing is that we studied natural reading in context. Some previous studies presented target words either embedded in sentences within the same frame (Cui et al., 2021), or with identical beginnings preceding them (Cui et al., 2013), or with certain rules of preceding and following verbs and commas, respectively (Yan et al., 2006). Such manipulations may weaken or avoid word segmentation. Readers may segment a sentence into words the first time it appears and are very likely to retrieve the segmentation from working memory when encountering it the next time, thus eliminating the need to fulfill word segmentation once again. As a result, readers may rely less on parafoveal processing for word segmentation than in natural reading.

Indeed, in our study sentences are neither controlled nor manipulated, allowing readers to read as they would in daily life. Therefore, it is not surprising that the C1 effect appeared in very early measures in controlled studies but not in ours. This finding is also in line with studies that report a low degree of correspondence between paradigms in psycholinguistics (with lowest correspondence on early eye-tracking measures; see Dirix et al., 2019). The C1 effect found in our work may not reflect the complete character processing due to the parafoveal processing discussed above, which might (at least partially) explain the difference with previous findings where the C1 may have been full processed (e.g., Cui et al., 2021). Below, we will discuss possible reasons for the different results between the previous work and ours that can be ruled out.

One could argue that the facilitative character FE found in our own and Yan et al. (2006)'s work is an artifact of predictability, as Yu et al. (2021) suggested. Indeed, word anticipation could reduce competition between the target word and the candidate words (including one- and multiple-character words) whose meanings are implausible in the sentence context, resulting in target word reading facilitation. However, since no empirical evidence nor theoretical assumptions have suggested that predictability affects the first but not the C2, we expected a facilitatory effect of the C2 frequency as well. Yet, we observed facilitative as well as absent and reversed FEs of C2, which will be discussed in detail below. Apparently, the predictability explanation can account for some of our results, but not all.

Second character

The second remarkable finding of this work is the reliable C2 FE in all reading time measures. Previous studies only found a FE of the C1 (e.g., Yan et al., 2006; Cui et al., 2021). This work, however, is the first to observe the significant main effect of the C2, demonstrating that its frequency affects both the early and late stages of the word recognition process. In addition to the possible reasons for the parafoveal processing discussed above, it is worth noting that Yu et al. (2021) only manipulated the frequency of the C1 and controlled that of the C2, making an estimation of the C2 FE impossible. In Yan et al. (2006), the authors did not detect a main effect of the C2 frequency but observed a significant interaction between the C2 and word frequency, consistent with the current findings.

Before proceeding, it is important to clarify two things: (a) whether the underlying processes differ for C1 and C2 reading; (b) whether the two characters in a word influence each other. If one of (a) and (b) is fulfilled, it would be inappropriate to compare the FE of C2 either to the C1 or to studies where the C1 has been fully processed. Cui et al. (2021) suggested that since HF C1s appear more often as the initial character of words and have larger morphological family sizes, they are less predictive for the C2 in two-character words in comparison to LF C1s. Thus, the fixation duration on the C2 should be longer, consistent with what they found.

From this, they explained that the inhibitory C1 FE they found in LF words was a carry-over effect due to morphological family size. Hence, according to Cui et al. (2021), the processing of the C2 is affected by the frequency of the C1 and may have a different underlying process than the C1. However, if C2 processing is largely constrained by the characteristics of the C1, the properties of the C2 should have minimal effects on word reading. That is, if two different C1s have similar morphological family sizes, the processing time of the accompanying C2s, regardless of their frequency and complexity, should be similar.

However, the interaction between the C1 and C2 frequencies was not reliably significant in the current and previous work. Instead, the current study demonstrated reliable FEs of C2 in all time measures, arguing against the above inference. Additionally, Yu et al. (2021) argued on the basis of a post hoc analysis that, although statistical power was small, the morphological family size might not be responsible for the inhibitory character FE they observed. Collectively, the frequency of the C1 seems unlikely to influence the C2 processing or to be different from its process.

Interestingly, our findings regarding the fully processed C2 frequency seem to be consistent with previous results that contradict each other (e.g., Yan et al., 2006; Yu et al., 2021). The facilitative character FE of shorter fixation duration for HF characters observed in LF words (see Table A2 in Appendix B) was consistent with Yan et al. (2006). In contrast, the inhibitory pattern found in HF words was similar to that of Yu et al. (2021) and Cui et al., (2021), showing that higher character frequency was related to longer word fixation duration. The character FE was absent in medium-frequency words, congruent with the findings of Yan et al. (2006) and Cui et al. (2021) in their HF condition. This may explain why some studies failed to obtain the character FE (e.g., Li et al., 2014).

The variation of the character FE reiterates the importance of having a wide range of test stimuli and using them as continuous factors. Otherwise, the obtained results may reflect only part of the effect and lead to unnecessary contradiction and confusion, as shown by previous evidence. So far, there are some explanations for the character FE. Cui et al. (2021) indicated that the inhibitory character FE was due to the morphological family size, which we have discussed above, and argued against it being the main reason for the character FE. The Chinese E-Z Reader (CEZR) proposed by Yu et al. (2021) involves an interesting alternative, independent of morphological or neighborhood family size. We will explore in detail whether CEZR and CRM (Li & Pollatsek, 2020) can explain the current results after discussing the word FE.

Word FE

Consistent with previous studies on Chinese (e.g., Wei et al., 2013) and alphabetic language reading (e.g., Cop et al., 2015), this work also observed a significant facilitatory main effect of word frequency across all measures. HF words had shorter fixation durations than LF words at the reference level. Furthermore, the word FE was found to be modulated by character frequency: the facilitative word FE decreases and even reverses as the character frequency increases. These findings suggest that word frequency affects word recognition and C2 frequency throughout the processing stage.

The facilitative word FE appears when the C2 frequency is less than 6.2 (the highest character frequency in Cai & Brysbaert (2010) is about 7.64). Apparently, it is the most frequently observed pattern in Chinese reading (e.g., Cui et al., 2021; Wei et al., 2013). Additionally, the fixation duration on words with HF characters is often shorter than those with LF characters (see Figure 1). Thus, experiments that categorize frequency variables are likely to find the facilitative word FEs only, as is often observed (e.g., Ma et al., 2015). It also reiterates the importance of studying continuous variables to detect the full picture of the effect. When the C2 frequency exceeds about 6.2, the word FE is no longer significant due to its strong influence. It could clarify the previously found seemingly counterintuitive results, where weak or absent word FE may be due to the influence of the character frequency rather than low statistical power or design flaws (e.g., Liversedge et al., 2014).

At extremely high C2 frequency (> 7.3), there was a tendency for an inhibitory word FE, which only occurs in GD and TRT (see Figure 1 and Table A2). Indeed, such an effect is expected because word FE should change as C2 frequency increases in a similar way to how word frequency affects character FE. Yet, the number of stimuli under this condition is limited even in a large-size database. Further investigation is needed to verify the stability and reproducibility of the inhibitory word FE. Notably, various patterns of the word and character FEs are unlikely to be fully noticed in isolated word recognition. The predetermined word length allows for the quick exclusion of word candidates of different lengths (e.g., one-character words), thereby limiting the influence of character frequency on word retrieval. It may explain the observation of only facilitative word and character FEs in the lexical-decision task (Tse & Yap, 2018), suggesting that artificial tasks may not reflect the natural reading process fairly (also see Dirix et al., 2019).

Our findings regarding the word FE have several theoretical implications. First, it strengthens the argument that words are indeed an important processing unit in Chinese reading, although their word boundaries are less explicit than in characters (Li et al., 2014). Their properties influence eye-movement measures during reading. Second, the facilitative word FE observed when its character frequency is not particularly high reflects, to some extent, the similarity of word retrieval processes between Chinese and alphabetic language, as the effect should primarily reflect the role of word frequency in cases where the character FE is limited. Third, the interaction between word and character frequencies confirms our hypothesis that the nature and manifestations of the word FE in Chinese reading differ from those in alphabetic languages. Moreover, their interaction allowed us to learn the extent of their influence on each other. Given the word FE is only affected by relatively HF characters, one can infer that the influence of word frequency is greater than that of character frequency.

One might argue that, in some cases, the Chinese word processes might resemble that of alphabetic languages. In fact, Cui et al. (2021) indicated that HF words could be processed as a holistic unit, and character property only plays a role in LF words. If so, HF Chinese words may have a process similar to that of alphabetic languages. While Cui et al. (2021)'s assumption could explain the absence of character FE in HF words found previously (e.g., Yan et al., 2006; Cui et al., 2021), it cannot account for the reversed character FE we found. Therefore, word processing is unlikely to vary with its frequencies.

Additionally, bilingualism may be argued to affect the exploration of FEs in Chinese reading. However, previous research has shown that word FEs are affected by language proficiency rather than bilingualism, indicating no notable distinction between monolinguals and bilinguals in native language processing (Cop et al., 2015). Furthermore, no empirical or theoretical supports suggests that bilingualism influences character FEs. Additionally, since English is a compulsory course in China starting no later than junior high school, recruiting young adults who only speak Mandarin is a considerable challenge, even in China. Therefore, there is no reason to assume that the current results are qualitatively different from previous related findings. Nonetheless, future research is needed to investigate these issues.

Implications for Chinese reading models and research

A reputable Chinese reading model should be able to explain existing results. To do so, it should posit a role for character processing and incorporate its influence on word frequency into word processing. However, some existing models are designed for character identification (e.g., split-fovea model, Hsiao & Shillcock, 2004, 2005) and cannot explain the widely observed word FE (e.g., Li et al., 2014). Conversely, some are designed to illustrate word processing without character processes engaged (e.g., extended E-Z Reader model, Rayner et al., 2007). These are apparently at odds with the observed character FE (e.g., Yan et al., 2006). There are also models based on word identification and including the role of characters, namely the CRM (Li & Pollatsek, 2020) and the CEZR (Yu et al., 2021). While they can explain the word and character FEs found previously well, they fail to explain the novel results we observed.

In CEZR (Yu et al., 2021), word processing is related to how words are segmented. If the processing time of the first unidentified character exceeds a certain threshold, as may be the case with a LF C1 in a two-character word, the word-identification system initially infers it as a single-character word with a shorter fixation. In contrast, if the processing time does not exceed the threshold and is less than the time to process the first and C2s, for instance in HF initial character words, the word-identification system infers it as a two-character word with longer fixation. Although these hypotheses may seem counterintuitive, they can explain the obtained inhibitory character FE. The authors argued that their model predicted the facilitative effect for non-target words while the inhibitory effect for carefully controlled target words. The failure to observe the facilitative effect in the analysis of target words is due to the strong collinearity of words and character frequencies. The model appears to be able to explain the standard word FE as well as both facilitative and inhibitory character FEs obtained in this work. However, the conditions under which different character FEs occur are not entirely understood. Additionally, it seems unable to explain the C2 FE we found, as it stated the role of the C1 but did not specify if and how the C2 impacts word recognition. Therefore, it is even less able to explain the interaction between the C2 and the word frequencies this study observed.

In contrast, CRM (Li & Pollatsek, 2020) was able to account for the effects of the C2 properties since it indicates parallel character processing. However, since it does not elucidate the role of character frequency on word frequency, it is unable to fully explain the interactions reported here unless further assumptions are proposed. As discussed, the model assumes that character frequency affects word recognition in a hybrid manner. HF characters can greatly facilitate the retrieval of the words they constitute, but also induce more protracted interference as single-character words in lexical competition (also see Li et al., 2022). The amount of word processing time is affected by the magnitude of the character impact at different stage.

If one further hypothesizes that character frequency interacts with word frequency at the word retrieval stage, as an extension of CRM (Li & Pollatsek, 2020), then our results can be well explained. That is, when word frequency is low, word retrieval is difficult and time-consuming. It can benefit more from its HF components. The amount of facilitation from the HF character could be more influential than the prolonged time of the interference it causes. Its counteracted effect may be greater than that of LF characters, therefore demonstrating a facilitative character in LF words, as found in this work.

Conversely, as the word frequency increases, the word is easier to access from memory and therefore benefits lesser from HF characters. The facilitation of HF characters may not differ much from that of LF characters, yet it still causes more interference in lexical competition. Its interference may affect word processing more than its facilitation, with a greater residual effect than LF characters. As a result, the fixation durations of words with LF characters are shorter than those with HF characters, consistent with the inhibitory character frequency we found in HF words. Whereas in medium-frequency word, the counteracted effect of facilitation and interference of HF characters may be quantitatively similar to those of LF characters. Thus, the processing times of HF and LF characters are not statistically different, explaining the absent character FE this study found.

Likewise, LF characters cause limited facilitation and interference. Therefore, the higher the word frequency, the faster the word is processed. Conversely, HF characters cause high interference but greatly facilitate word retrieval for LF words and have very limited effects on HF words. After counteracting the influence between facilitation and interference, the word FE could illustrate a flattened or even a reversed trend. These assumptions could explain our finding of observing facilitative and absent word FEs and provide a theoretical hypothesis for the inhibitory trends when its characters are fairly HF.

Of course, there is another inference that can explain the current findings. While the LF multiple-character word retrieval is time-consuming, HF characters can quickly access one-character words. One-character words that are not plausible with the sentence context can be quickly excluded, resulting in reconsidering multiple-character words. Thus, the LF multiple-character words with HF characters should be processed faster than those with LF ones. In contrast, HF multiple-character words can be quickly retrieved and compete with single-character words activated by HF characters but not with LF single-character words. Therefore, HF multiple-character words with HF characters are expected to be processed slower than those with LF characters. Future research needs to verify how exactly word and character frequencies are engaged in Chinese reading, as they are apparently vital for understanding Chinese reading.

Nevertheless, the novel findings in this work can be an important new “benchmark” with significant implications for assessing theoretical assumptions of Chinese reading. Future model developments should be able to simulate the findings reported here, derived from a large corpus of natural reading. Future research on Chinese reading and aiming to seek computational models or mechanism should be cautious in proposing assumptions and interpretations, even if the results they find are similar across different writing systems.

Conclusion

This work, we believe, is the first to propose and demonstrate with evidence that the observed word FE in Chinese is, in fact, fundamentally different from that in alphabetic languages, both in its manifestation and in its underlying processes. It is also the first to test the continuous effect of Chinese word and character frequencies across the entire ranges in text reading. In the present study, the word FE exhibited facultative and flattening patterns with the change of character frequency, demonstrating that word frequency interacts with character frequency. Additionally, it observed all the seemingly contradictory character FEs (e.g., facultative, flattening, and inhibitory) found in previous research (e.g., Li et al., 2014; Yan et al., 2006), indicating the different appearance of the character FE is due to the variation in word frequency. Finally, our findings emphasize the importance of using natural reading material to explore effects as continuous variables in statistically powerful research to reveal the full picture of effects.

References

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), 1277–1287.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*, 133(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001

Boston, M.F., Hale, J.T., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*. 2(1): 1-12.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45-50.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5, e10729. https://doi.org/10.1371/journal.pone.0010729

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and bigrams: Orthographic redundancy and syllabic units affect visual word recognition at different processing levels. Journal of Experimental Psychology: Human Perception and Performance, 35(2), 461–479. https://doi.org/10.1037/a0013480.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, 22(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Cui, L., Wang, J., Zhang, Y., Cong, F., Zhang, W., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology*, 74(4), 610-633.

Cui, L., Yan, G., Bai, X., Hyönä, J., Wang, S., & Liversedge, S. P. (2012). Processing of compound-word characters in reading Chinese: An eye-movement-contingent display change study. *Quarterly Journal of Experimental Psychology*, 66(3), 527–547. https://doi.org/10.1080/17470218.2012.667423.

Dictionary Editorial Office of Linguistic Research Institute, Chinese Academy of Social Sciences (2016). *Modern Chinese Dictionary* (7th Edition). Beijing, China: The Commercial Press.

Dirix, N., & Duyck, W. (2017). The first-and second-language age of acquisition effect in first-and second-language book reading. *Journal of Memory and Language*, 97, 103-120.

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior Research Methods*, 51(6), 2800–2816.

Hasenäcker, J., & Schroeder, S. (2019). Compound reading in German: Effects of constituent frequency and whole-word frequency in children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(5), 920–933. https://doi.org/10.1037/xlm0000623

Hoosain, R. (1992). Psychological Reality of the Word in Chinese. In H. C. Chen, & O. J. L. Tzeng (Eds.), *Language Processing in Chinese* (pp. 111-130). Amsterdam, NY: North-Holland. http://dx.doi.org/10.1016/S0166-4115(08)61889-0

Hsu, S.-H., and Huang, K.-C. (2000). Effects of word spacing on reading Chinese text from a video display terminal. *Perceptual and Motor Skills*, 90, 81–92. doi: 10.2466/pms.2000.90.1.81

Hsiao, J. H., & Shillcock, R. (2004). Connectionist modelling of Chinese character pronunciation based on foveal splitting. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society* (pp. 601–606). Erlbaum.

Hsiao, J. H., & Shillcock, R. (2005). Differences of split and non-split architectures emerged from modelling Chinese character pronunciation. In B. G. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty Seventh Annual Conference of the Cognitive Science Society* (pp. 989–994). Erlbaum

Huang, L., & Li, X. (2020). Early, but not overwhelming: The effect of prior context on segmenting overlapping ambiguous strings when reading Chinese. *Quarterly Journal of Experimental Psychology*, 73(9), 1382-1395.

Huang, L., Staub, A., & Li, X. (2021). Prior context influences lexical competition when segmenting Chinese overlapping ambiguous strings. *Journal of Memory and Language*, 118, 104218.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23, 1089–1132. https://doi.org/10.1080/01690960802193688.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 876–895. https://doi.org/10.1037/a0013484.

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), 895–913.

Li, X., Gu, J., Liu, P., & Rayner, K. (2013). The advantage of word-based processing in Chinese reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 879–889

Li, X., Huang, L., Yao, P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology*, 1-12.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, 127(6), 1139–1162.

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

Li, X., Zhao, W., & Pollatsek, A. (2012). Dividing lines at the word boundary position helps reading in Chinese. *Psychonomic bulletin & review*, 19(5), 929-934.

Liu P-P, Li W-J, Lin N, Li X-S (2013) Do Chinese Readers Follow the National Standard Rules for Word Segmentation during Reading? *PLoS ONE*, 8(2): e55440. doi:10.1371/journal.pone.0055440

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, 1008–1025.

Liu, Y., Yu, L., Fu, L., Li, W., Duan, Z., & Reichle, E. D. (2019). The effects of parafoveal word frequency and segmentation on saccade targeting during Chinese reading. *Psychonomic bulletin & review*, 26(4), 1367-1376.

Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition*, 22(3-4), 441–457. https://doi.org/10.1080/13506285.2014.889260

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. Journal of Experimental Psychology: Human Perception and Performance, 40(3), 1046–1059. https://doi.org/10.1037/a0035389.

Ma, G., Li, X., & Rayner, K. (2015). Readers extract character frequency information from nonfixated-target word at long pre-target fixations during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 41(5), 1409-1419.

Mattingly, I. G., & Xu, Y. (1994). Word superiority in Chinese. *Advances in the Study of Chinese Language Processing*, 1, 101–111.

New, B., & Grainger, J. (2011). On letter frequency effects. *Acta Psychologica*, 138(2), 322-328.

Oralova G and Kuperman V. (2021). Effects of Spacing on Sentence Reading in Chinese. *Frontiers in Psychology*. 12:765335. doi: 10.3389/fpsyg.2021.765335

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.

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., 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., 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., Sereno, S. C.,Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, 4(special issue), 21–49.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 514–528. https://doi.org/10.1037/a0020990

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11), 489–493. https://doi.org/10.1016/j.tics.2003.09.010

Slattery, T.J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition*, 35, 1283-1292.

Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods*, 1-21.

Tse, C. S., & Yap, M. J. (2018). The role of lexical variables in the visual recognition of two-character Chinese compound words: A megastudy analysis. *Quarterly Journal of Experimental Psychology*, 71(9), 2022-2038.

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.

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.

Yang, J., Staub, A., Li, N., Wang, S., & Rayner, K. (2012). Plausibility effects when reading one- and two-character words in Chinese: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(6), 1801–1809. https://doi.org/10.1037/a0028478

Yang, J., Wang, S., Xu, Y., & Rayner, K. (2009). Do Chinese readers obtain preview benefit from word n + 2? Evidence from eye movements. *Journal of Experimental*

Psychology: Human Perception and Performance, 35(4), 1192–1204. https://doi.org/10.1037/a0013554

Yu, L., Liu, Y., & Reichle, E. D. (2021). A corpus-based versus experimental examination of word- and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of Experimental Psychology: General*, 150(8), 1612–1641. https://doi.org/10.1037/xge0001014

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2018). Investigating word length effects in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 44(12), 1831–1841. https://doi.org/10.1037/xhp0000589

Zang, C., Zhang, M., Bai, X., Yan, G., Paterson, K. B., & Liversedge, S. P. (2016). Effects of word frequency and visual complexity on eye movements of young and older Chinese readers. *Quarterly Journal of Experimental Psychology*, 69(7), 1409–1425. https://doi.org/10.1080/17470218.2015.1083594

Zhou, J., Ma, G., Li, X., & Taft, M. (2018). The time course of incremental word processing during Chinese reading. *Reading and Writing*, 31(3), 607-625.

CHAPTER 4. The Word Frequency Effect in First and Second Language Reading by Chinese and Dutch Bilinguals[9]

Abstract

High-frequency (HF) words are processed faster than low-frequency (LF) words, a phenomenon known as the word frequency effect (FE). Although the effect has been studied in various writing systems as well as in first (L1) and second language (L2) reading, existing theoretical hypotheses about the FE are mainly based on findings in alphabetic languages. To date, no study investigated whether the learning, lexical entrenchment and rank hypotheses regarding of FE apply to Chinese-English bilinguals. The present study is therefore the first to compare the FEs in Chinese- and Dutch-English bilinguals during natural paragraph reading in L1 and L2, using eye-tracking measures. The results showed that Chinese bilinguals exhibited a larger FE in L1 than in L2. In addition, they have a smaller L1 FE and much steeper L2 FE curves than Dutch bilinguals. These findings are not entirely consistent with the existing FE hypotheses, and the present study discusses theoretical accounts in light of the results we found.

Keywords: Frequency effects, reading, bilingualism, Eye-movements

9 Sui, L., Woumans, E., Duyck, W., Dirix, N., (submitted) The Word Frequency Effect in First and Second Language Reading by Chinese and Dutch Bilinguals. *Journal of Memory and Language*.

The word FE refers to the phenomenon that words with a higher frequency of occurrence are processed faster than those that appear less often. It has been well-studied in monolinguals and bilinguals of alphabetic languages (e.g., English) and is one of the strongest factors affecting word processing (Brysbaert et al., 2016). The effect is also evident in reading Chinese (e.g., Li et al., 2014), a writing system that systematically differs from that of alphabetic languages in terms of spelling and pronunciation. However, recent evidence shows that although there is an overall positive FE in Chinese reading, the underlying processes are, in fact, different from those in English (Sui et al., submitted). Consequently, the magnitude of the Chinese FE in natural reading may be different from that of alphabetic languages. In addition to such between-group comparisons across languages, an interesting line of research has also compared FEs within readers, between L1 and L2, in research on alphabetical languages. Typically, the FE is larger in L2 than in L1 (Duyck et al., 2008 in lexical decision task; Cop et. al., 2015; Whitford & Titone, 2012, 2017 in eye-tracking).

However, none of the studies have compared L1 and L2 FEs in natural reading among bilinguals with different L1 writing systems. Given the lack of empirical evidence, the existing assumptions of FEs proposed based on alphabetical languages, such as the learning, lexical entrenchment, and rank hypotheses, have yet to be verified for their applicability to Chinese-English bilinguals. This work, therefore, aims to investigate whether readers with distinct L1s (e.g., Chinese and Dutch) and the same L2 (e.g., English) have comparable size of FEs in L1 and L2, respectively. In addition, it will compare the FEs of L1 and L2 within the group of Chinese-English bilinguals. Investigation of these questions firstly allows us to examine related theories based primarily on alphabet reading research (e.g., learning hypothesis) and assess their universality. Secondly, it allows an understanding of the similarities and differences between different L1 writing systems and whether the nature of the L1 writing system affects the processing of the L2.

Besides the above cross-lingual complications, it is also important to consider the (experimental) context in which words appear. Word recognition in natural text is influenced by a wide range of contextual influences (e.g., syntactical or semantic expectations), and consequently, reading in isolation may differ from natural reading (Kuperman et al., 2013; Dirix et al., 2019). Indeed, the FE observed in tasks where target words are presented in isolation (e.g., the lexical decision task) appears to be larger in comparison to target words embedded in sentences (e.g., in eye-tracking research; Dirix et al., 2019). Apparently, studying the FEs of words in sentences, which closely resembles reading in everyday life, is essential for understanding language processing, especially in Chinese reading. It is because words are important units in reading, and their boundaries are often not clearly defined, making word segmentation essential for reading Chinese sentences but not in isolated word conditions (see Sui et al., submitted, for discussion). In the current study, we therefore investigate FEs by comparing two eye-tracking corpora, GECO (Ghent Eye-tracking Corpus, Cop et al., 2017a) and GECO-CN (Ghent Eye-tracking COrpus for Chinese-English bilinguals, Sui et al., 2022), which recorded eye-movement data for L1s and L2s (English) in paragraph reading for native Chinese and Dutch bilinguals, respectively.

Eye-tracking is a popular method used to study the underlying processes involved in sentence reading by monitoring the eye movements of the reader while reading. This approach provides a range of eye movement measures (Rayner, 2009), such as saccades (the action of rapidly moving eyes to a new point) and fixations (the duration of eyes fixating on a specific point). The fixation duration includes multiple reading measures, such as a) first-fixation duration (FFD), the duration that the reader first fixates on a word, b) gaze duration (GD), the summed duration that fixates on a word in the first pass, and c) total-reading times (TRT), the summed durations of all fixations and re-fixations on a word. The first two are generally viewed as early measures, whereas the last one is considered a late measure (e.g., Clifton et al., 2007). Skipping probability refers to whether the word is skipped during the reading, not just in the first pass.

As another methodological issue, some studies investigate word frequency as a categorical variable, although it naturally occurs as a continuous variable (e.g., Li et al., 2014). However, categorizing continuous variables can result in reduced statistical power and reliability, inappropriate rejection of the null hypothesis, and failure to capture the variation of the effect (Balota et al., 2004). Here, the large amount of target words in the two eye-tracking corpora allow us to assess word frequency as a continuous variable. In the following part, we will begin with brief summaries of the existing findings on the L1 FE for distinct writing systems, i.e., alphabetic languages and Chinese. Then, we will review the key results on L2 FEs and discuss theoretical issues regarding FEs in bilinguals. Finally, we will report the analysis of this research and discuss the main findings obtained.

L1 frequency effect in Alphabetic languages

The FE, the difference in processing times between LF and HF words, has been studied extensively in L1 reading of alphabetic languages (e.g., Cop et al., 2015; Rayner & Raney, 1996; Whitford & Titone, 2017). The effect is one of the most potent phenomena (it explains over 30% of the variance in lexical decision mega-studies; Keuleers et al., 2012; Yap & Balota, 2009; Brysbaert et al., 2016; Keuleers et al., 2010a; Ferrand et al., 2010) and is robust in both monolinguals and bilinguals (Cop et al., 2015). Numerous reading experiments have shown that when reading in the first or dominant language, alphabetic language readers spend more time fixating on LF words and are less likely to skip them than HF words (e.g., Duyck et al., 2008 in lexical decision; Cop et al., 2015 in sentence reading; see Rayner, 2009 for a review). The FE appears to be modulated by the degree of language exposure: readers with more language exposure exhibit a smaller FE (e.g., Ashby et al., 2005). In addition, English monolinguals and alphabetic language bilinguals (e.g., Dutch-English) exhibit comparable FEs in L1 reading (or dominant language; Cop et al., 2015 in sentence reading; Diependaele et al., 2013 in lexical decision).

Furthermore, unskilled readers exhibited larger FEs compared to skilled readers, with steeper curves at LF words (Kuperman & Van Dyke, 2013). Apparently, the limited exposure to a language appears to negatively affect exposure to LF words since such readers are likely to have a limited vocabulary and may opt for easier materials (i.e., with fewer LF words; Brysbaert et al., 2017). Consequently, their exposure to HF words should be similar to readers with extensive language exposure but considerably less to LF words. As a result, the difference in reading times between HF and LF words decreases with increased language exposure, leading to a reduced FE, congruent with the existing findings.

L1 FE in Chinese writing systems

Chinese is a logographic language that is qualitatively distinct from alphabetic languages. Chinese characters are written in strokes, and are the components of words. In Chinese, there are about 5,000 commonly used characters and they can constitute more than 56,000 words. The most encountered word type is two-character words, while the commonly used word tokens are one-character words, i.e., the characters themselves. One- and two-character words account for the majority of commonly used Chinese words (97.2%, Li & Pollatsek, 2020). Obviously, Chinese words are, on average, much shorter than those of alphabetic languages. Another major difference between Chinese and alphabetic languages is that the words of the former are not visually separated in sentences, whereas the latter contains spaces between words. That is, a character might be a single-character word or form a word with its preceding or following character in a Chinese sentence. Since the word is an important processing unit in Chinese reading (for the discussion, see Li et al., 2014; Sui et al., 2022), word segmentation is challenging, but undoubtedly necessary for Chinese sentence reading.

The evidence shows that despite the lack of visual demarcation between words in Chinese sentences, most research has observed conventional word FEs (e.g., Sui et al., submitted; Li et al., 2014). The word FE of Chinese single-character words is inconsistent, with some studies showing a significant main effect (Zang et al., 2016) and others failing to find it (Liversedge et al., 2014). Liversedge et al. (2014) did not observe a main effect of word frequency but did observe a significant interaction between frequency and word (character) complexity (i.e., number of strokes). That is, the fixation duration was longer for LF, complex words. In contrast, the results of the FE for multi-character words were consistent, with shorter reading times (e.g., Li et al., 2014; Ma et al., 2015; Sui et al., submitted) and higher skip rates for HF words (e.g., Cui et al., 2021; Liu et al., 2019; Yan et al., 2006). In general, the FEs found in Chinese sentence reading are concordant with those reported in alphabetic languages.

Note that most studies investigating the FE on eye movements in Chinese reading only studied target words (primarily two-character content words) embedded in a single manipulated low-constrained sentence (e.g., Li et al., 2014; but see Sui et al., submitted in paragraphs), with many even using the same sentence frames that differed only in the target words (e.g., Cui et al., 2013, 2021), in order to minimize sentence context effects. Furthermore, most research investigated word FEs using dichotomous frequency categories (e.g., Li et al., 2014; Cui et al., 2021), i.e., categorizing continuous frequencies. Only one analyzed them as continuous variables (i.e., Sui et al., submitted). Hence, the word FEs observed in some existing research may be adversely affected or even biased by employing these manipulations. In addition, the effect of language exposure on the FE does not seem to apply to Chinese readers, unlike alphabetic languages. So far, only Sui et al. (submitted) have considered language proficiency (a proxy of language exposure) when studying the Chinese word FE. Surprisingly, we did not find an effect of language proficiency on the word FE.

L2 frequency effect

An increasing number of studies have investigated whether a FE also occurs in L2 reading. Evidence has shown that unbalanced bilinguals usually have a larger FE in L2 than in L1 reading (e.g., Duyck et al., 2008 in lexical decision; Cop et al., 2015; Whitford & Titone, 2012, 2017 in eye movements). When language exposure is included as a predictor in the analyses, the difference between the FEs in L1 and L2 reading becomes negligibly small in the lexical decision tasks (Brysbaert et al., 2017) but not in eye-movement studies (Cop et al., 2015), where FEs remain larger in the L2 than in the L1. Cop et al. (2015) explained that the distinct results observed in different experiments may be due to the usage of disparate methods. The eye-movement measures are, not surprisingly, more complex and time-sensitive than the reaction times obtained in lexical decisions.

Notably, however, the findings of a larger FE in L2 originated primarily from the exploration of alphabetic language pairs (e.g., Cop et al., 2015). Only a few studies have explored FEs in within-group comparisons of bilinguals with non- and alphabetic language pairs, namely Hebrew-English (e.g., Mor & Prior, 2020). Mor and Prior (2020) found a larger FE in L2 than in L1 among unbalanced Hebrew-English bilinguals using a continuous variable analysis of word frequency in a lexical decision task. Still, these pioneer findings need to be further explored in both different scripts (such as Chinese) and with different experimental paradigms (such as natural reading). Furthermore, the existing FE hypotheses proposed based on findings of alphabetic languages remain to be verified for the speculation on the L2 word frequency between bilinguals with disparate L1 writing systems due to the lack of empirical evidence. Below, we will discuss the existing hypotheses regarding FEs.

FE hypotheses

The learning hypothesis is generally considered to explain the FE. It suggests that repeated exposure to an item could lower recognition threshold (e.g., the logogen model of Morton, 1970) or raise baseline activation (e.g., Monsell, 1991, cited from Cop et. al., 2015). Hence, HF words, which have a higher rate of exposure, are processed faster than LF words. In addition, this hypothesis involves the asymptotic learning function, which posits that as the occurrences of words increase (i.e., as word frequency increases), the facilitation effect of learning on its performance gradually diminishes, resulting in a corresponding decrease in processing time until it remains constant (also see Duyck et al., 2008; Murray & Forster, 2004). Therefore, word recognition times should correlate negatively with word frequency in a nonlinear, logarithmic way.

The FE can also be explained by the lexical entrenchment hypothesis, which highlights the strength of lexical representations in memory. Frequent exposure to a word leads to more entrenched representations, resulting in faster and more accurate processing compared with LF words. Given that unbalanced bilinguals are generally less exposed to their L2, the objective frequency of their L2 should be lower than that of their L1. Both theoretical hypotheses predict a larger FE in L2 than in L1, consistent with the existing findings (for detailed discussion, see Duyck et al., 2008). Interestingly, they also predict that once the L1 and/or L2 exposure is similar in balanced bilinguals, the L1 and/or L2 FEs of the two groups should be similar in size, regardless of their writing systems (e.g., Chinese and Dutch).

Alternatively, another possible explanation is the rank hypothesis (Murray & Forster, 2004) that was extended to bilingual readers (Duyck et al., 2008). It suggests that the lexicon is organized into frequency-ordered bins while searching is done sequentially, starting with

HF words. For bilinguals, the bins are either language-specific (i.e., L1 or L2) with specific scanning speeds or shared by all known languages, whose processing time increases nonlinearly with decreasing word frequency. If, as suggested by the preceding extended hypothesis, the lexicon is language-specific and the scanning speed of L2 is longer than that of L1, then bilinguals with similar frequency rankings should exhibit comparable FEs even if the language writing systems are different. If, as bilingual word recognition suggests (Brysbaert & Duyck, 2010), lexicons are not language-specific, then bilinguals with similar word frequency rankings should have comparable FEs, regardless of language dominance or writing systems. One might also assume interactions with cross-lingual similarity, for instance that the bins are shared only when L1 and L2 employ the same writing system (e.g., Dutch-English). In this case, bilinguals with different scripts (e.g., Chinese-English) may still have separate, and thus smaller bins for their L1 and L2, resulting in smaller L1 and L2 FEs than those that share bins. The L1 and L2 FE could also be more similar in size, as lexicons are separated for the different scripts (for a discussion, see Duyck et al., 2008). If different-script bilinguals have longer scan speed in L2 than the same-script bilinguals, as assumed above, L2 FEs of different- and same-script bilinguals may also be similar in size.

To summarize, all the above assumptions predict that language exposure moderates the word FE, regardless of writing systems or language dominance (L1 or L2). When language proficiency is included in the analysis, Chinese-English and alphabetic language bilinguals should exhibit similar L1 FEs (except for the last extended rank hypothesis, which also predicts a larger FE in Chinese bilinguals). Additionally, unbalanced bilinguals with less L2 exposure should have a larger FE in L2 reading, regardless of language pairs, either due to the relatively more asymptotic learning function in their LF words or relatively weak lexical representations or to the well-behind location of LF L2 words or longer L2 scanning speed in the frequency-ranked bins. Balanced bilinguals with similar exposure to both languages should have similar FEs in their L1 and L2. In addition, various hypotheses generate different predictions regarding the L2 FEs for bilinguals with distinct L1 wiring systems. The learning, lexical entrenchment, and frequency-ranked (assuming that bins are language-specific or shared among languages) hypotheses suggest that the L2 FE should not be affected by the L1 writing systems. Instead, the frequency-ranked hypothesis, which assumes that bins are only shared by alphabetic languages and are language-specific for different scripts, indicates that the L2 FE may vary with the L1 writing system and that different-script bilinguals should exhibit smaller L2 effects than those with the same script.

Current study

Comparisons of FEs in natural reading between Chinese-English and alphabetic languages bilinguals in their L1 and L2 and compare L1 and L2 Fes within Chinese bilinguals are of theoretical importance. First, they are necessary to evaluate the universality of assumptions and predictions of FE and reading theories. Second, they can shed light on whether word processing differs between L1s with diverse writing systems and whether the L2 reading is affected by the L1 writing system. By doing so, one can provide a plausible explanation for the seemingly counterintuitive results that may be found in different groups of bilinguals.

However, to date, no studies have compared the FEs of Chinese-English bilinguals with those of the same alphabet bilingual reading. Indeed, studying FEs in natural reading across-group of bilinguals is a considerable challenge. One reason is that data collection among bilinguals with disparate L1s is challenging (e.g., preparing materials) and time-consuming, especially when aiming for a dataset with sufficient power. In addition, cross-experiment comparisons are generally not convincing in investigating FE differences in reading across bilinguals unless carefully matched. One major reason is that differences in materials affect reading performance as discussed above. Yet, studying this effect in isolated conditions is not ideal, as the observed phenomenon cannot fully reflect the performance in natural reading, especially for Chinese-English bilinguals who need to perform word segmentation, which may affect word recognition in Chinese sentence reading (see discuss above).

Hence, the present study aims to investigate the FEs of bilinguals with different L1 writing systems and the same L2, i.e., Chinese- and Dutch-English bilinguals, in tL1 and L2sreading by measuring their eye movements. Our first interest is to understand whether the L1 FE of non-alphabetic (i.e., Chinese) is comparable to that of alphabetic languages (i.e., Dutch) and whether language exposure explains the variation in FEs. Our second interest is to compare the L2 FE between different bilinguals and whether it differs depending on the L1 writing systems. Our third interest is to verify whether the FE in L2 is larger than in L1 for unbalanced Chinese-English bilinguals (see Appendix C for the results of this analysis; note that Cop et al., 2015 have explored the within-group comparisons for Dutch bilinguals). We will further consider language exposure, which is known to influence FEs in alphabetic languages (e.g., Brysbaert et al., 2017), by examining whether this influence applies to different writing systems, and whether it can explain group differences across bilinguals.

We will compare eye-movement data from two large corpora, GECO (Dutch-English bilinguals; Cop et al., 2017a) and GECO-CN (Chinese-English bilinguals; Sui et al., 2022), in which unbalanced bilinguals read different language versions of an entire novel in paragraphs. Readers read half of the novel in their L1 and the other half in L2. The novel has approximately 5,000 sentences and contains a wide range of word stimuli, and thus word frequencies, in each language. Logically, the linguistic properties these two datasets involve should be comparable and not interfere with the comparison between the bilingual groups. In addition, these corpora shared identical experimental procedures. Both also corpora provide Lextale scores (Dutch and English; HSK score for Chinese), which reflect language proficiency by examining the vocabulary size, which we will use as a proxy of language exposure.

Method

Participants and Materials

GECO

GECO (Cop, Dirix, Drieghe, & Duyck, 2017) is an eye movement corpus where 19 Dutch-English bilinguals and 14 British English monolinguals read an entire novel (*The Mysterious Affair at Styles* by Agatha Christie) while their eye movement behaviour was measured. Dutch natives read half of the novel — in Dutch and the other half in English, whereas monolinguals read the entire book in English. For the present study, only the bilingual data was used. For further information on the corpus, we refer the reader to Cop et al. (2017).

GECO-CN

GECO-CN is a dataset consisting of eye movement data from 30 Chinese-English bilinguals. It follows the identical experiment procedure and uses the same reading materials as the original GECO (Cop et al., 2017a). Participants read half of the novel in Chinese and the other half in English. They also complete a series of language proficiency tests in both languages. For more details, we refer the reader to Sui et al. (2022).

Analysis

This study only investigated content words (for Chinese-English bilinguals, 511,157 data points in Chinese and 442,638 in English; for Dutch-English bilinguals, 275,458 data points in Dutch and 264,634 in English) excluding all cognates, as these orthographically and semantically overlapping equivalents may confound the investigation of the FE. The present work classified a word as a cognate if its Levenshtein distance between the two languages was greater than or equal to 0.7 (in orthography; 5.19% of words in Dutch and 7.29% in

English; also see Da Silveira & van Leussen, 2015). Cognates were only present among the Dutch-English text. Furthermore, the first and last words of a line and fixations of less than 100 ms were removed from the analysis, as the former could reflect the sentence wrap-up effect (e.g., Rayner et al., 1989; 10.31% in Chinese and 16.97% in English for Chinese-English bilinguals; 17.3% in Dutch and 16.8% in English for Dutch-English bilinguals), while the latter fixations are considered too short to reflect word processing (e.g., Sereno & Rayner, 2003).

This experiment used R software (version 3.4.1; R Core Team, 2017) to perform linear mixed-effects models (LMMs) from the lme4 package (Version 1.1–12). We conducted separate analyses for L1 and L2 and considered important psycholinguistic predictors as control variables. In each model, predictor variables included group (categorical, Chinese vs. Dutch bilinguals), word frequency (continuous), word length (continuous), proficiency of the language congruent with the model (continuous; if the model investigates the performance of the L1, it is the L1 proficiency), and the chronological of word repetition in sessions (continuous). Additionally, we examined various eye movement measures as dependent variables, including FFD, GD, and TRT (e.g., Clifton et al., 2007) and skipping probability. The random effects were participant and the word token. The predictors were all centered, whereas the dependent variables were Box-Cox transformed. Such transformation normalized the distribution without changing its functional relationship. In each reading time measure, fixation durations differing by more than 2.5 standard deviations per individual and per language were discarded.

Word length is one of the important factors affecting frequency performance. Yet, the average length of Chinese words is much shorter than that of alphabetic language words. Thus, this work made some adjustments by proportioning word length in Chinese and Dutch. For example, the longest Chinese words in GECO-CN were the six-character words. The length of a one-character word then became 1/6, and the length of a three-character word became 1/2. The method was used for Dutch word length rescaling as well.

Notably, since both bilingual groups had English as their L2 and completed the English LexTALE, L2 word length and language proficiency were not rescaled. In addition, this work used the same log10-transformed Zipf (frequency) based on SUBTLEX-CH (Cai & Brysbaert, 2010), SUBTLEX-NL (Keuleers et al., 2010b), and SUBTLEX-UK (Van Heuven et al., 2014) as frequencies for Chinese, Dutch, and English words, respectively. We also employed the car package (Version 3.0-12) to calculate the Variance Inflation Factor to estimate the multicollinearity of coefficients in each regression model. A VIF greater than 5 or 10 was considered as moderate or severe multicollinearity, respectively (also see Dirix & Duyck, 2017).

Results

Bilingual L1: Chinese versus Dutch

First Fixation Duration

The FFD in L1 reading did not differ significantly between the Chinese and Dutch bilingual groups (see Table 1). Both bilingual groups showed an overall FE, with shorter FFDs for HF words, and the effect was significant larger in the Dutch bilinguals. There was also a word length effect, where Chinese and Dutch bilinguals overall spent less time fixating on short words than on long, with similar effect sizes for both groups. However, frequency and word length interacted significantly in both groups. The FE became larger as word length increased, and increased more in Dutch than in Chinese bilinguals (see Fig.1). Language proficiency did not influence word fixation durations of either Chinese or Dutch bilinguals, but it did affect the FE in Dutch bilinguals (see Fig 2). The effect size of word frequency decreases with the increase of language proficiency in Dutch bilinguals. The word repetition effect had inconsistent results between the two groups. The effect was found only in Dutch bilinguals, with a decrease in fixation duration as the times of word repetitions increased.

Gaze Duration

Chinese bilinguals did not differ from Dutch bilinguals in terms of GD (see Table 1). Both frequency and word length had significant effects in the Chinese group, showing shorter GDs on higher frequency or shorter words. However, these effects differed between Chinese and Dutch groups, with a steeper frequency and word length effect for Dutch bilinguals. Notably, there was a significant interaction between frequency and word length in Chinese bilinguals, which differed significantly from Dutch bilinguals. The FE increased with word length, and this increase is greater in Dutch bilinguals, as shown in Figure 1. Unlike what was observed for FFD, language proficiency of Chinese and Dutch bilinguals did not affect GD or the FE (see Fig.2). The word repetition effect was present in Dutch bilinguals and negatively correlated with fixation durations, but not in Chinese bilinguals.

Table 1

Comparative Analyses of Fixation Duration Measures and Skipping Probabilities in First Language Reading between Chinese- and Dutch-English Bilinguals

	Predictors	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**VIF**
First Fixation Duration	**(Intercept)**	**4.420000**	**0.012510**	**353.389**	**< 0.001*** **	
	GroupDutch	-0.006788	0.020490	-0.331	0.742	1.13
	Frequency	**-0.002032**	**0.000620**	**-3.278**	**0.001** **	3.944
	Word Length	**-0.016320**	**0.007143**	**-2.285**	**0.022***	4.397
	L1 Proficiency	0.000104	0.003982	0.026	0.979	4.112
	Repetition	0.000012	0.000007	1.684	0.092.	2.087
	GroupDutch:Frequency	**-0.005959**	**0.000965**	**-6.174**	**< 0.001*** **	4.464
	GroupDutch:Word Length	0.009347	0.009995	0.935	0.350	5.215
	GroupDutch:L1 Proficiency	-0.000767	0.004675	-0.164	0.870	3.949
	Frequency:Word Length	**-0.013210**	**0.004298**	**-3.074**	**0.002** **	4.285
	Frequency:L1 Proficiency	-0.000032	0.000126	-0.258	0.797	4.978
	GroupDutch:Repetition	**-0.000030**	**0.000014**	**-2.06**	**0.039***	2.044
	GroupDutch:Frequency:Word Length	**-0.014350**	**0.005586**	**-2.568**	**0.010***	5.09
	GroupDutch:Frequency:L1 Proficiency	**0.000350**	**0.000143**	**2.455**	**0.014***	4.992

	Predictors	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**VIF**
Gaze Duration	**(Intercept)**	**2.698000**	**0.004325**	**623.81**	**< 0.001*** **	
	GroupDutch	-0.001164	0.007084	-0.164	0.870	1.13
	Frequency	**-0.001100**	**0.000216**	**-5.091**	**< 0.001*** **	3.938
	Word Length	**0.024900**	**0.002492**	**9.991**	**< 0.001*** **	4.371
	L1 Proficiency	-0.000102	0.001377	-0.074	0.941	4.112
	Repetition	0.000004	0.000002	1.738	0.082.	2.089
	GroupDutch:Frequency	**-0.002119**	**0.000336**	**-6.31**	**< 0.001*** **	4.448
	GroupDutch:Word Length	**0.009374**	**0.003482**	**2.692**	**0.007** **	5.165
	GroupDutch:L1 Proficiency	**-0.000152**	**0.001616**	**-0.094**	**0.925**	**3.949**
	Frequency:Word Length	**-0.007469**	**0.001513**	**-4.938**	**< 0.001*** **	4.21
	Frequency:L1 Proficiency	0.000077	0.000043	1.796	0.072.	4.969
	GroupDutch:Repetition	**-0.000012**	**0.000005**	**-2.46**	**0.014***	2.045
	GroupDutch:Frequency:Word Length	**-0.008119**	**0.001966**	**-4.131**	**< 0.001*** **	4.983
	GroupDutch:Frequency:L1 Proficiency	0.000066	0.000049	1.354	0.176	4.984

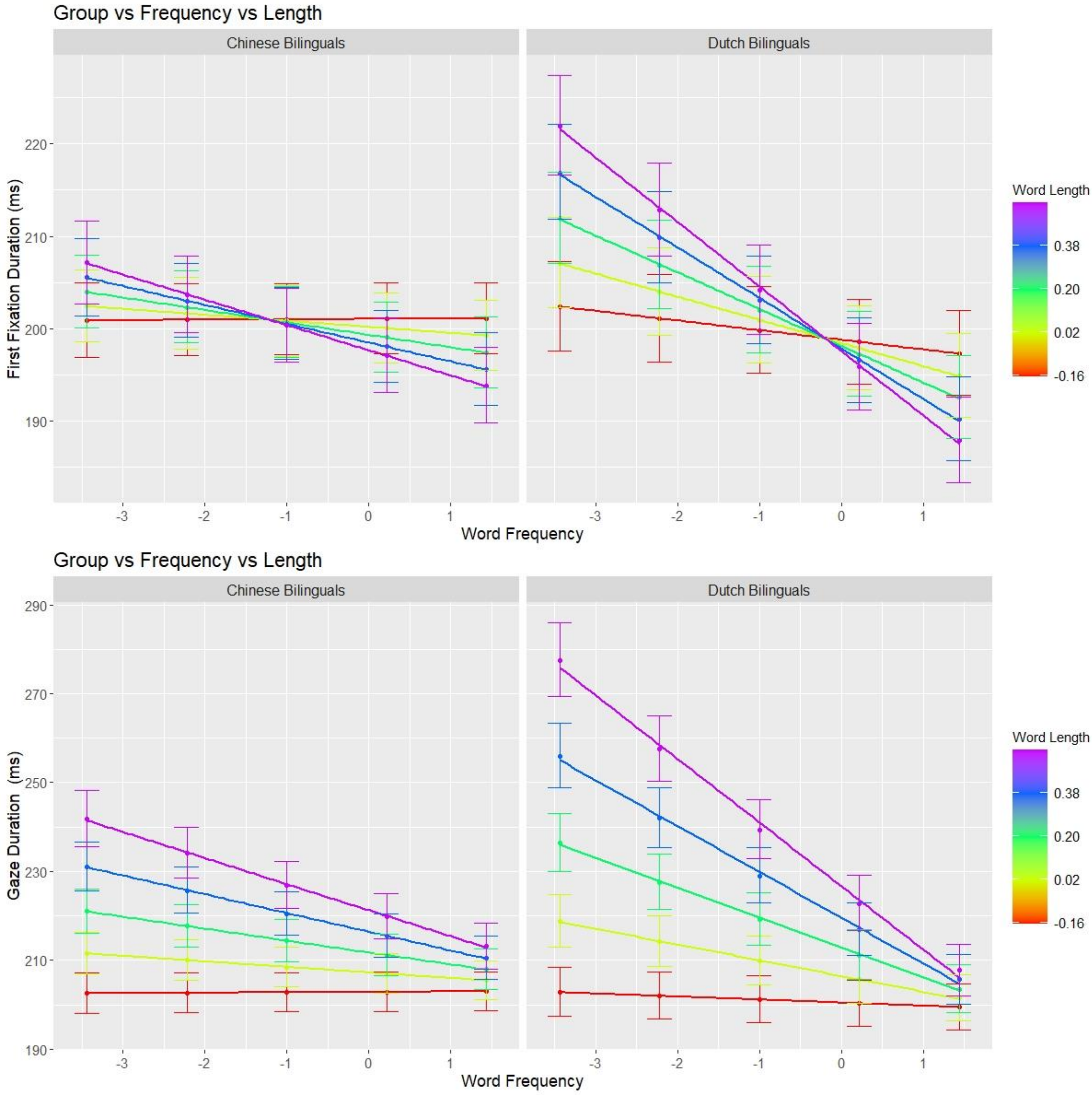
	Predictors	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**VIF**
Total Reading Time	**(Intercept)**	**2.236000**	**0.002739**	**816.247**	**< 0.001*** **	
	GroupDutch	0.001979	0.004486	0.441	0.661	1.13
	Frequency	**-0.000761**	**0.000147**	**-5.193**	**< 0.001*** **	3.942
	Word Length	**0.036540**	**0.001694**	**21.572**	**< 0.001*** **	4.39
	L1 Proficiency	-0.000103	0.000872	-0.118	0.906	4.113
	Repetition	**0.000007**	**0.000002**	**4.064**	**< 0.001*** **	2.091
	GroupDutch:Frequency	**-0.001228**	**0.000227**	**-5.404**	**< 0.001*** **	4.449
	GroupDutch:Word Length	**-0.006887**	**0.002362**	**-2.916**	**0.004** **	5.185
	GroupDutch:L1 Proficiency	-0.000004	0.001024	-0.004	0.997	3.949
	Frequency:Word Length	-0.000182	0.001034	-0.176	0.861	4.235
	Frequency:L1 Proficiency	0.000027	0.000029	0.938	0.348	4.957
	GroupDutch:Repetition	**-0.000019**	**0.000003**	**-5.584**	**< 0.001*** **	2.047
	GroupDutch:Frequency:Word Length	**-0.012810**	**0.001339**	**-9.567**	**< 0.001*** **	4.999
	GroupDutch:Frequency:L1 Proficiency	**0.000064**	**0.000032**	**1.985**	**0.047***	4.973

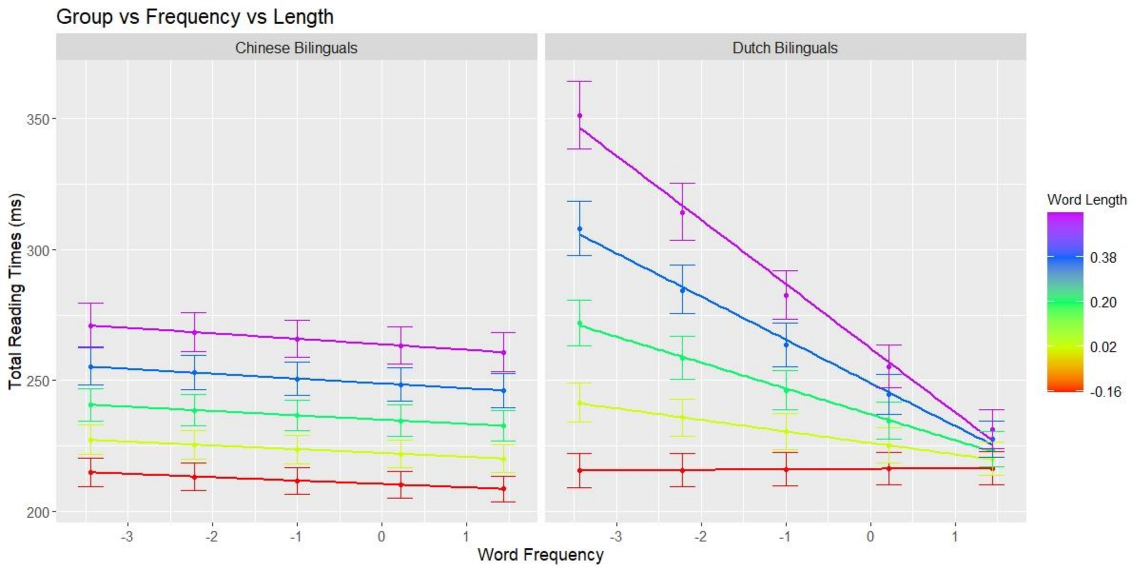
	Predictors	Estimate	Std. Error	z value	Pr(>|z|)	VIF
Skipping probability	**(Intercept)**	**0.251100**	**0.077160**	**3.254**	**0.001****	
	GroupDutch	**-1.256000**	**0.126000**	**-9.968**	**< 0.001*****	1.002
	Frequency	**0.023670**	**0.005184**	**4.565**	**< 0.001*****	2.937
	Word Length	**-5.489000**	**0.061780**	**-88.837**	**< 0.001*****	2.756
	Repetition	**-0.000447**	**0.000055**	**-8.145**	**< 0.001*****	1.993
	GroupDutch:Frequency	**0.090710**	**0.009541**	**9.507**	**< 0.001*****	3.286
	GroupDutch:Word Length	**-0.539000**	**0.104500**	**-5.157**	**< 0.001*****	3.212
	Frequency:Word Length	**-0.101100**	**0.042260**	**-2.391**	**0.017***	2.182
	GroupDutch:Repetition	**0.000853**	**0.000123**	**6.952**	**< 0.001*****	1.945
	GroupDutch:Frequency:Word Length	**-0.164000**	**0.063860**	**-2.568**	**0.010***	2.293

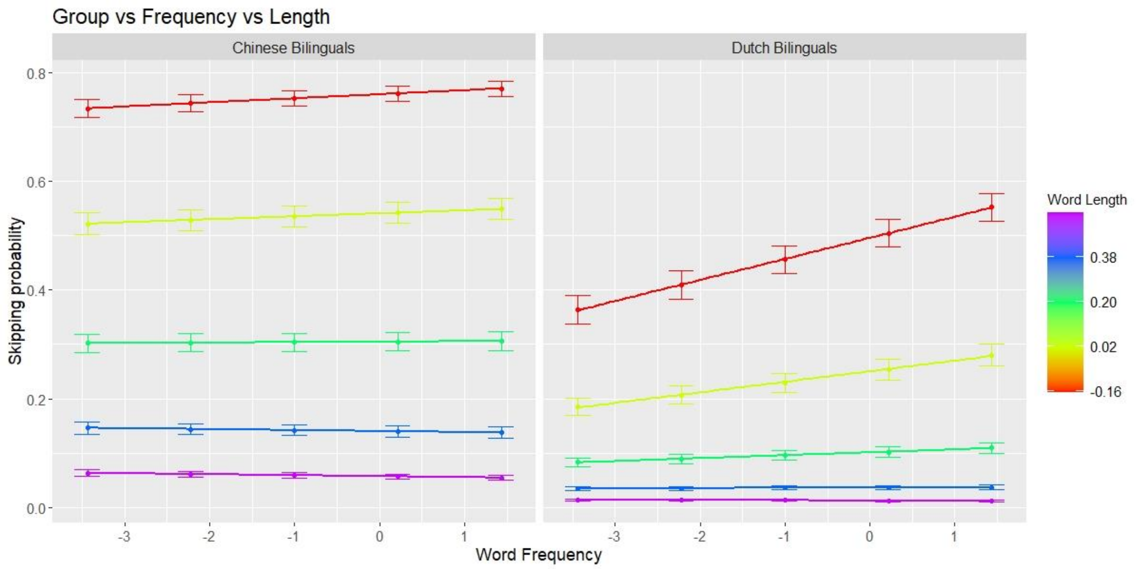
*p < .05. **p < .01. ***p < .001

Estimate = Estimates; SE = standard errors; *t* = t-values; ***p*** = *p*-values (calculated using lmerTest package); VIF = variance inflation factor.

Figure 1 Three-way Interaction Plots between the Group, Word Length, and Frequency in First Language Reading.



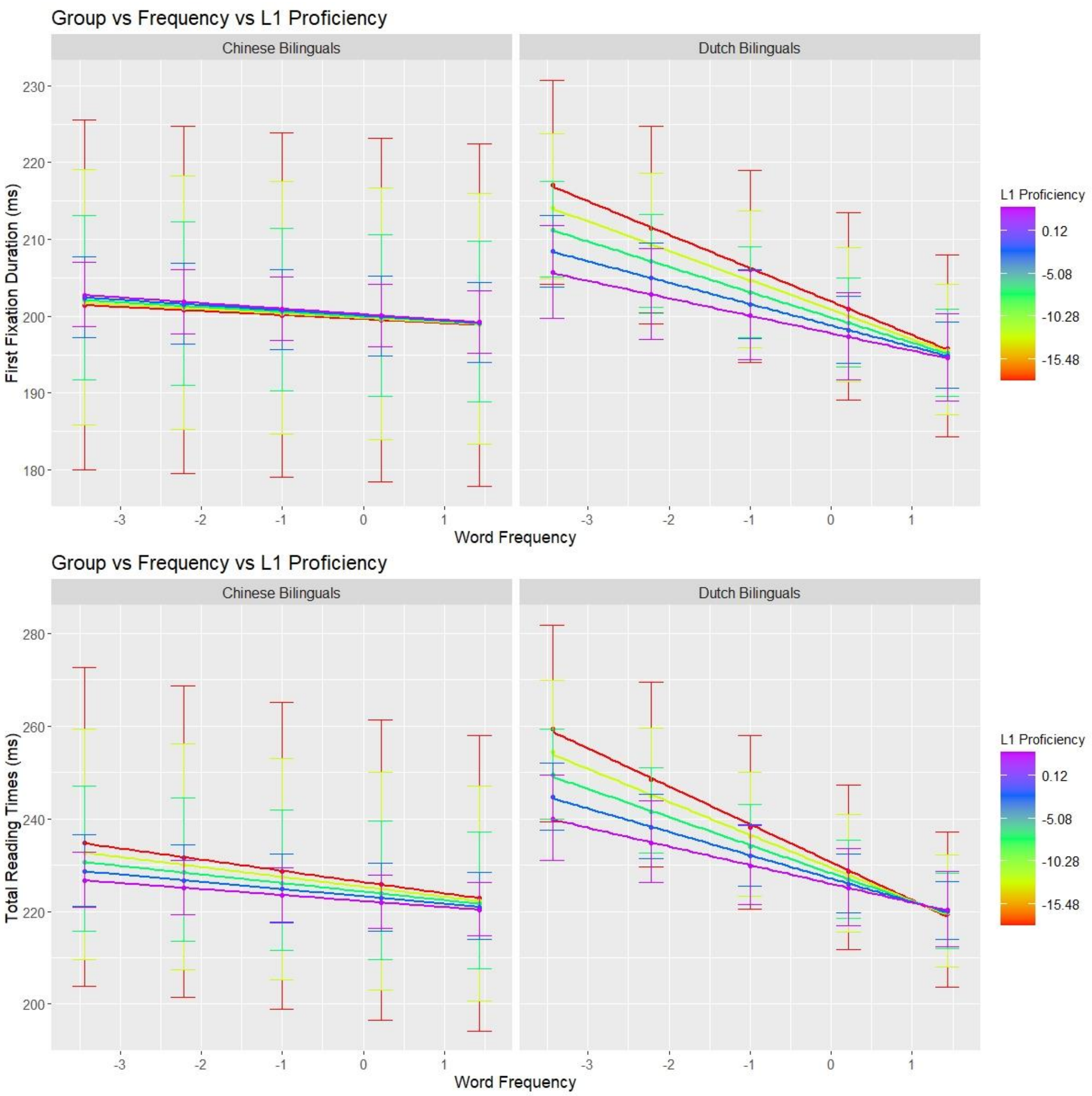




Total Reading Time

Overall, there was no significant difference in TRTs between Chinese and Dutch bilinguals (see Table 1). The frequency and word length effects were significant in Chinese and Dutch bilinguals, with a significantly smaller frequency and larger word length effect for Chinese compared to Dutch bilinguals. The interaction between frequency and word length and between frequency and L1 proficiency was not evident in Chinese bilinguals, significantly different from Dutch bilinguals (see Fig.1). Word length and language proficiency had a greater effect on LF words than on HF ones, thus showing a larger FE in long words or among readers with low proficiency levels in Dutch bilinguals (see Fig.2). The repetition effect was significant in both groups, inhibitory in Chinese and facilitatory in Dutch bilinguals.

Figure 2 Three-way Interaction Plots between the Group, First Language Proficiency, and Frequency in First Language Reading.



Skipping probability

The skip probability of Chinese bilinguals was significantly higher than that of Dutch bilinguals (see Table 1). Frequency and word length effects were significant in both groups. The higher the frequency, or the shorter the word length, the higher the skipping probability. The FE was larger in Dutch bilinguals, while that of word length was larger in Chinese. There were significant interactions between frequency and word length in both groups (see Fig.1). The FE was larger in short words and in Dutch bilinguals. The repetition effect was significant in Chinese bilinguals, whose performance differed markedly from that of Dutch bilinguals. The former group showed an inhibitory effect, with a low skipping probability for words that were repeated more often, whereas the latter group showed a facilitative effect, with higher word repetition related to higher skip rates.

Bilingual L2: English versus English

First Fixation Duration

Chinese bilinguals showed significantly longer FFD in L2 reading than Dutch bilinguals (see Table 2). They also showed significant frequency and word length effects, larger than those observed in Dutch bilinguals. The interaction between frequency and word length was significant in Chinese bilinguals, and different from that in Dutch bilinguals (see Fig.3). As word length reduced, the decrease in fixations for HF words was greater for Chinese bilinguals than for LF words, while the decrease in LF words was greater for Dutch bilinguals. Language proficiency had no effect on Chinese bilinguals, whereas it interacted with frequency among Dutch bilinguals (see Fig.4). The FE decreased with increasing language proficiency, with higher frequency words being affected more than lower frequency words. In addition, both groups spent more time reading frequently repeated words than infrequently repeated words, and the effect was similar between them.

Table 2

Comparative Analyses of Fixation Duration Measures and Skipping Probabilities in Second Language Reading between Chinese- and Dutch-English Bilinguals

	Predictors	Estimate	SE	*t*	*p*	VIF
First Fixation Duration	**(Intercept)**	**6.267000**	**0.025290**	**247.844**	**< .001 *****	
	GroupDutch	**-0.101400**	**0.040610**	**-2.498**	**.02 ***	1.00
	Frequency	**-0.028750**	**0.001148**	**-25.044**	**< .001 *****	3.78
	Word Length	**0.015660**	**0.000618**	**25.328**	**< .001 *****	3.68
	L2 Proficiency	-0.002214	0.002130	-1.039	.30	1.71
	Repetition	**0.000057**	**0.000019**	**2.948**	**.00 ****	2.36
	GroupDutch:Frequency	**0.008806**	**0.001807**	**4.873**	**.00 *****	3.48
	GroupDutch:Word Length	**-0.008814**	**0.001012**	**-8.707**	**< .001 *****	3.62
	GroupDutch:L2 Proficiency	0.001527	0.003311	0.461	.65	1.71
	Frequency:Word Length	**0.004098**	**0.000340**	**12.049**	**< .001 *****	2.34
	Frequency:L2 Proficiency	-0.000005	0.000053	-0.085	.93	1.58
	GroupDutch:Repetition	0.000012	0.000032	0.383	.70	2.28
	GroupDutch:Frequency:Word Length	**-0.005291**	**0.000558**	**-9.486**	**< .001 *****	2.34
	GroupDutch:Frequency:L2 Proficiency	**-0.000229**	**0.000088**	**-2.598**	**.01 ****	1.58

	Predictors	Estimate	SE	*t*	*p*	VIF
Gaze Duration	**(Intercept)**	**3.81E+00**	**1.10E-02**	**347.163**	**< .001 *****	
	GroupDutch	**-9.67E-02**	**1.76E-02**	**-5.493**	**.00 *****	1.00
	Frequency	**-0.028130**	**0.000503**	**-55.904**	**< .001 *****	3.71
	Word Length	**0.020720**	**0.000276**	**74.981**	**< .001 *****	3.72
	L2 Proficiency	-0.000942	0.000923	-1.02	.31	1.71
	Repetition	**0.000027**	**0.000007**	**3.936**	**.00** ***	1.53

	GroupDutch:Frequency	0.017360	0.000743	23.363	< .001 ***	3.21
	GroupDutch:Word Length	-0.011020	0.000431	-25.547	< .001 ***	3.55
	GroupDutch:L2 Proficiency	0.000494	0.001435	0.344	.73	1.71
	Frequency:Word Length	0.000783	0.000150	5.206	.00 ***	2.23
	Frequency:L2 Proficiency	0.000148	0.000022	6.834	.00 ***	1.59
	GroupDutch:Frequency:Word Length	-0.001457	0.000229	-6.375	.00 ***	2.04
	GroupDutch:Frequency:L2 Proficiency	-0.000271	0.000035	-7.649	.00 ***	1.59

	Predictors	Estimate	SE	*t*	*p*	VIF
Total Reading Time	(Intercept)	3.675000	0.010680	344.146	< .001 ***	
	GroupDutch	-0.077840	0.017150	-4.539	.00 ***	1.00
	Frequency	-0.028860	0.000541	-53.307	< .001 ***	3.78
	Word Length	0.017560	0.000297	59.081	< .001 ***	3.79
	L2 Proficiency	-0.000741	0.000899	-0.823	.41	1.71
	Repetition	0.000020	0.000007	2.759	.01 **	1.53
	GroupDutch:Frequency	0.015520	0.000790	19.647	< .001 ***	3.27
	GroupDutch:Word Length	-0.005946	0.000458	-12.983	< .001 ***	3.61
	GroupDutch:L2 Proficiency	-0.000279	0.001398	-0.199	.84	1.71
	Frequency:Word Length	0.000929	0.000161	5.767	.00 ***	2.25
	Frequency:L2 Proficiency	0.000187	0.000022	8.496	< .001 ***	1.59
	GroupDutch:Frequency:Word Length	-0.002101	0.000243	-8.66	< .001 ***	2.06
	GroupDutch:Frequency:L2 Proficiency	-0.000181	0.000036	-4.999	.00 ***	1.59

	Predictors	Estimate	SE	*z value*	*p*	VIF
Skipping probability	(Intercept)	-2.565	0.09327	-27.498	< .001 ***	
	GroupDutch	0.9059	0.1496	6.057	< .001 ***	1.00
	Frequency	0.1218	0.009138	13.327	< .001 ***	4.77
	Word Length	-0.4628	0.005604	-82.588	< .001 ***	4.48
	L2 Proficiency	0.01801	0.00784	2.297	.0216 *	1.71
	Repetition	-0.0001921	0.00007779	-2.469	.0135 *	1.47
	GroupDutch:Frequency	-0.02233	0.01213	-1.841	.0656 .	4.24
	GroupDutch:Word Length	0.105	0.007585	13.849	< .001 ***	4.24
	GroupDutch:L2 Proficiency	-0.009839	0.01218	-0.808	.4192	1.71
	Frequency:Word Length	-0.09443	0.002817	-33.521	< .001 ***	2.15
	Frequency:L2 Proficiency	-0.002984	0.0004415	-6.757	< .001 ***	2.25
	GroupDutch:Frequency:Word Length	0.04805	0.003937	12.206	< .001 ***	2.04
	GroupDutch:Frequency:L2 Proficiency	0.007032	0.000596	11.806	< .001 ***	2.22

*p < .05. **p < .01. ***p < .001

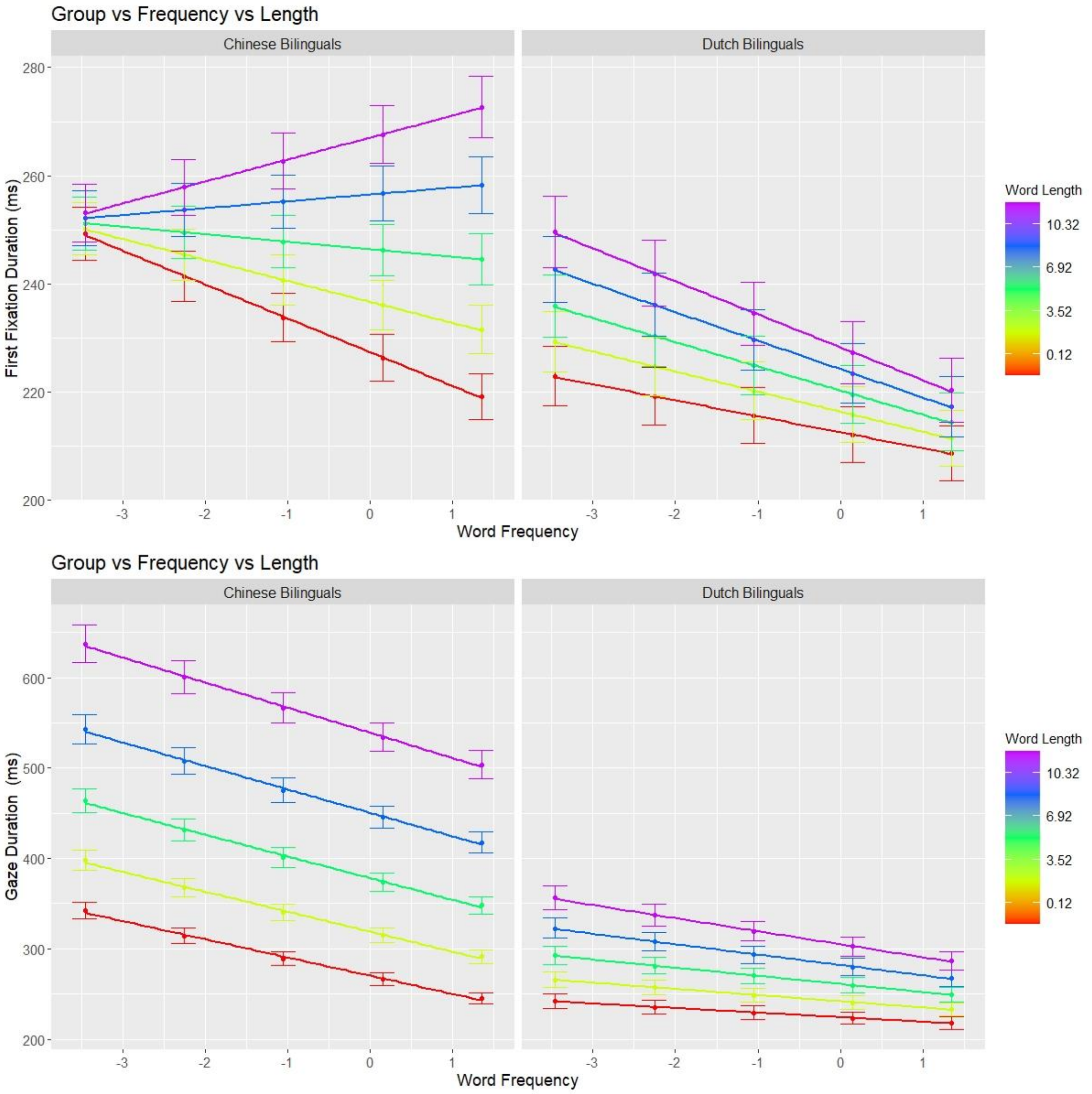
Estimate = Estimates; SE = standard errors; *t* = t-values; ***p*** = *p*-values (calculated using lmerTest package); VIF = variance inflation factor.

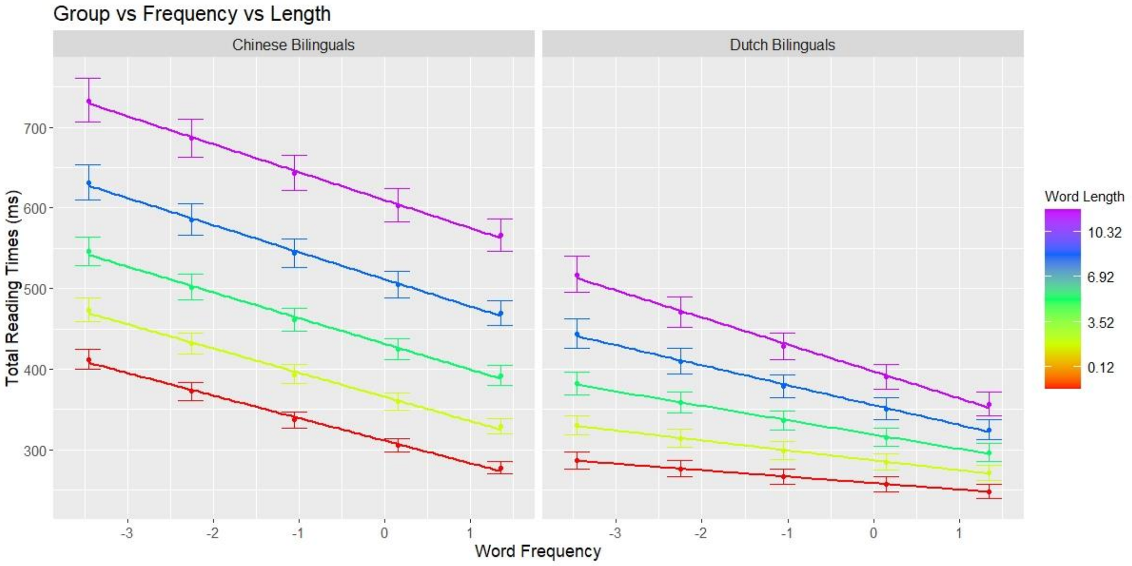
Gaze Duration

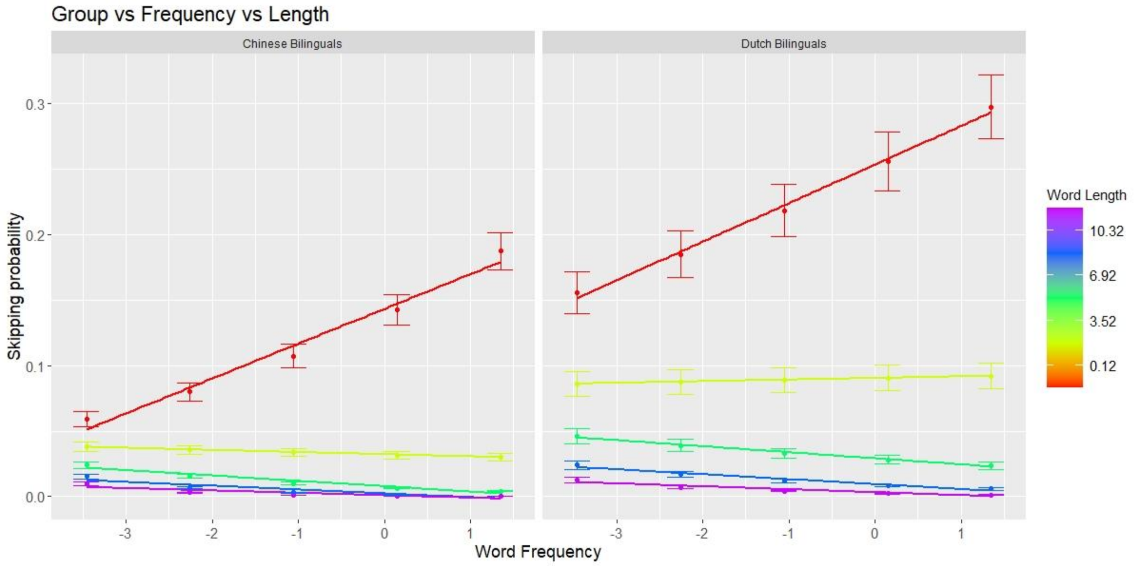
Similar to what was found for FFD, Chinese bilinguals spent more time on L2 reading than Dutch bilinguals, as shown in Table 2. Frequency and word length effects were significant in Chinese bilinguals and were statistically larger than those in Dutch bilinguals (see Fig.3). Both groups exhibited a significant interaction between frequency and word length, with Chinese bilinguals showing a smaller decrease in the FE as word length reduced.

Language proficiency had no effect on fixation duration, but did affect the FE in both groups, with a more pronounced decrease in FEs for Chinese than for Dutch bilinguals as language proficiency increased (see Fig.4). In addition, different from the results in FFD, as language proficiency increased, the Chinese group showed a greater decrease in GD for LF words, whereas the Dutch group showed a greater decrease for HF words.

Figure 3 Three-way Interaction Plots between the Group, Word Length, and Frequency in Second Language Reading.



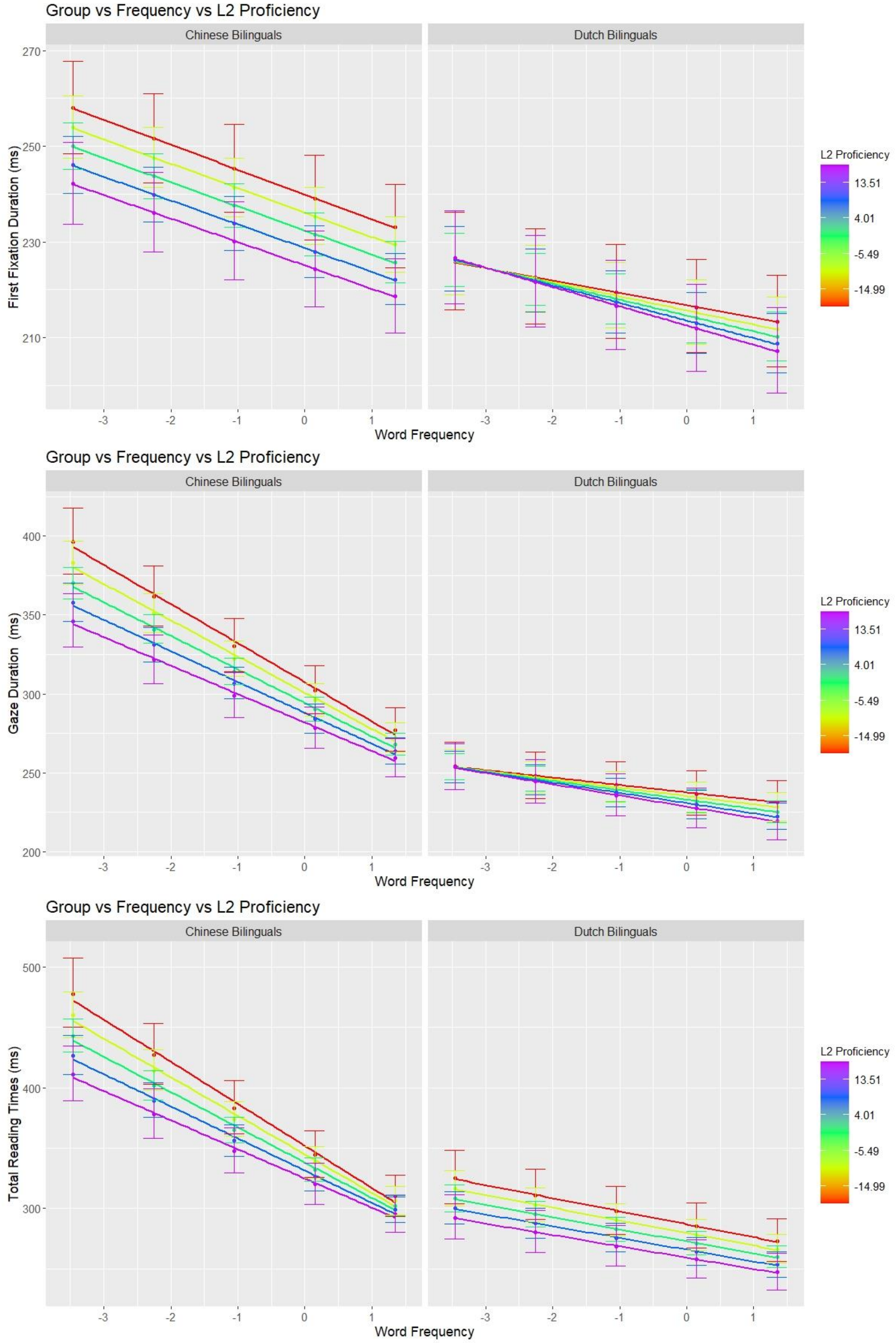


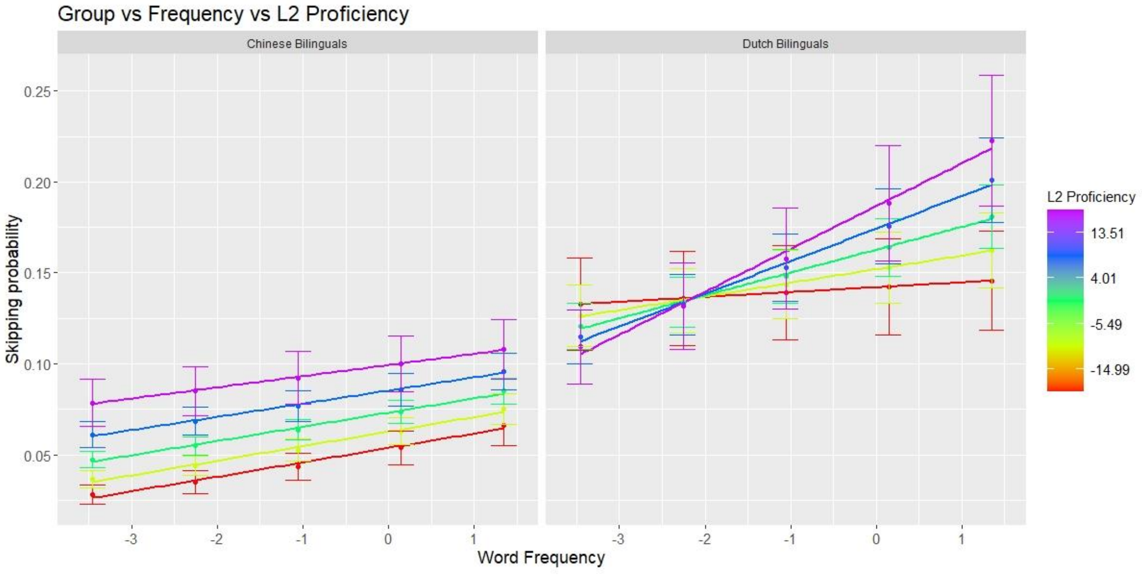


Total Reading Time

The Chinese group spent more time reading in L2 than the Dutch bilingual group (see Table 2). Frequency and word length in the two groups were negatively and positively correlated with TRTs, respectively, with smaller effect sizes in Dutch bilinguals. There was a significant interaction between frequency and word length in both groups, with the FE increasing with word length and to a greater extent in Dutch bilinguals, especially in LF words (see Fig.3). The effect of L2 proficiency on fixation duration was not evident in Chinese and Dutch bilinguals, as in FFD and GD (see Fig.4). However, it interacted with frequency in Chinese but not in Dutch bilinguals. Highly proficient readers exhibit a smaller FE, mainly manifested in the greater influence on the fixation duration of LF words. Both groups spent more time reading the more repeated words, and to a similar extent.

Figure 4 Three-way Interaction Plots between the Group, Second Language Proficiency, and Frequency in Second Language Reading.





Skipping probability

The main effect of the Group was significant, with Chinese bilinguals exhibited a lower skipping probability than Dutch bilinguals, different from that found in the L1 reading (see Table 2). The frequency, word length, and language proficiency were significant in Chinese bilinguals. Higher frequency, L2 proficiency, or shorter word length being associated with higher skip rates. They had a smaller FE than Dutch bilinguals but similar frequency and language proficiency effects. The interaction between frequency and word length was significant in Chinese bilinguals and differed significantly from that in Dutch bilinguals. As word length increased, the larger FE in short words decreased more in Chinese than in Dutch bilinguals (see Fig.3). Frequency was significantly interacted with language proficiency in both groups. As language proficiency increased, the FE decreased for Chinese bilinguals and increased for Dutch, manifesting in higher skipping probability for LF and HF words, respectively. The word repetition effect was significant and did not differ in the two groups. The more times a word is repeated, the lower the skipping probability.

Discussion

This work compared the FE between bilinguals with the same and different scripts in L1 and L2 reading as well as between L1 and L2 in Chinese bilinguals. Our three objectives were to examine a) whether the L1 FEs are similar in size across writing systems, b) whether the L2 FEs differ across readers with different L1 writing systems, and c) whether the L2 FE is larger than that of L1 in Chinese-English bilinguals. Language proficiency, which affects the FE, was also taken into account to ensure that if there were differences in the FEs between the two groups, they were not due to differences in language proficiency. Below, we will discuss the comparative results of FEs in L1 and L2 reading between the Chinese and Dutch groups. As the cross-group comparisons will address the L1 and L2 performance of Chinese-English bilinguals, we will briefly discuss their comparison between L1 and L2 FEs in the first subsection to avoid repetition. We will then relate these empirical observations to the predictions of the different theoretical accounts of the FE.

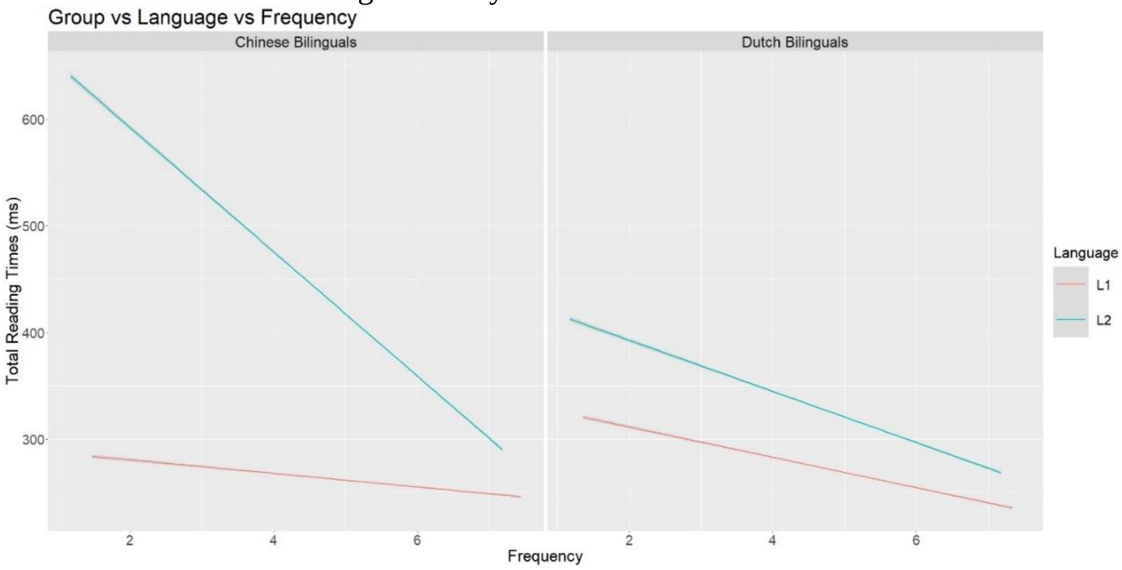
Bilingual L1: Chinese versus Dutch

In contrast to previous studies that found Chinese readers have longer fixation durations in single-sentence reading (Liversedge et al., 2016; Rayner et al., 2005; but see Sui et al., 2022), this work shows that they read their L1 as quickly as alphabetic language readers and with a much higher skipping probability. That is, Chinese readers have a much higher reading rate than Dutch bilinguals when reading texts of comparable length in L1. The divergent findings from previous studies may be due to the different nature of the reading material (e.g., single sentences vs. paragraphs; controlled sentences vs. natural sentences; for a discussion, see Sui et al., 2022). Here, we used a very natural form of reading, with meaningful, contextualized materials (a book).

Although the average length of Chinese words is much shorter than that of alphabetic languages, a Chinese word length effect is still observed in all reading measures as well as in skipping probabilities. FEs increased with word length in all reading measures for Dutch bilinguals but were only in FFD and GD for Chinese bilinguals, not in TRT. It indicates that the interaction between frequency and word length occurs only in the early stages of word recognition in Chinese bilinguals rather than throughout the time course as in Dutch bilinguals. As word length decreases, the FE, or be more specific, the fixation duration of LF words, decreases more rapidly in Dutch bilinguals than in Chinese bilinguals.

The L1 FEs of alphabet readers appear to be influenced by language proficiency rather than language quantity (monolinguals or bilinguals, Cop et al., 2015; Diependaele et al., 2013) or language (Dutch or French, Diependaele et al., 2013). The logical inference from this is that readers with similar proficiency levels should exhibit similar L1 FEs. Yet, Dutch bilinguals exhibited larger FEs than Chinese bilinguals in all reading measures and skipping probabilities in this study (see Fig.1 & Fig.5), inconsistent with some of the FE hypotheses discussed earlier, which we will discuss in detail below. In addition, the FEs in the L1 of Chinese-English bilinguals were significantly different from those in the L2, with larger effects in the L2 in all reading time measures (and the skip probability (see the results in Appendix C and Table A.1), congruent with previous findings (e.g., Cop et al., 2015; Mor & Prior, 2020). HF words had shorter fixation times or higher skip probability. The language proficiencies of Dutch bilinguals influenced their FE in all reading time measures. Readers with high language proficiency spent less time on LF words than those with low language proficiency. These findings are consistent with Kuperman & Van Dyke (2013)'s explanation that the FE is larger in those with less exposure to language, mainly because they have less exposure to LF words. However, for Chinese bilinguals, the interaction was marginally significant in the GD but not in FFD, TRT and skip probability.

Figure 5 Graph of raw total reading times for word frequency effects in L1 and L2 reading for Chinese and Dutch bilinguals. Grey shadows are confidence intervals.



The smaller FE for Chinese readers in skipping probabilities may be explained by a ceiling effect, as their skip rate reaches a surprisingly high 0.6. Regarding the relatively smaller FE in the time measures, there are several possible explanations: Firstly, certain language-specific factors, namely, character complexity, may affect FE. However, if this is the case, the word FE should be amplified rather than reduced, as words of high complexity often have lower frequencies. Furthermore, previous research has shown that there is no interaction between character complexity and word frequency, arguing against the assumption (Sui et al., submitted). Secondly, the number of words in languages may affect the FE. If there are significantly fewer words in Chinese than in Dutch, then Chinese words are likely to occur more frequently, resulting in a reduced FE. However, the number of commonly used words in Chinese and Dutch (about 56,000 words in Chinese and 54,319 in Dutch, Li & Su, 2022; Keuleers et al., 2015) and the frequency distribution of the analyzed data (see Fig.A.3) do not seem differ significantly, collectively arguing against this possibility.

Thirdly, the word frequency may have a more limited effect on reading in Chinese than in Dutch. Given that word frequency interacts with the frequency of its constituent characters in Chinese reading (for a discussion, see Sui et al., submitted), its impact on word recognition could be attenuated by character frequency. If, as previous work suggested, HF characters may have a greater inhibitory effect on HF words but a greater facilitative effect on LF words, it can explain the small difference in fixation duration between HF and LF words in Chinese reading and the limited influence of language proficiency on FE among Chinese bilinguals. However, it is worth noting that the influence of character frequency on word frequency has only been explored in two-character words. It is unclear if this effect applies to other word lengths, especially single-character words with high collinearity between character and word frequencies. In addition, word length in Chinese is determined by the number of characters. That is, words of different lengths have different numbers of character frequencies. It poses a challenge in investigating the effect of character frequency on word FE across different word lengths. Further research is needed to investigate whether the character frequency influence is responsible for the smaller word FE in Chinese with solid bases and well-designed.

Fourthly, the shorter word length constrains the degree of variation in FE. Undoubtedly, FE interacts with word length, where FE decreases as word length reduces, and shorter words have less visual complexity. Since Chinese words are generally much shorter than alphabetic language words, it is not surprising that their FEs are smaller and less affected by language proficiency, given the limited variations in FE. However, Chinese characters are composed of strokes, and their visual complexities far exceed those of the alphabet. That is, the visual complexity of a short Chinese word may not necessarily be lower than that of a long alphabetic language word (i.e., number of letters). Hence, whether the effect of word length on FEs is similar in Chinese and Dutch and whether it can explain the smaller FEs in Chinese reading requires further verification.

Bilingual L2: English versus English

In contrast to the findings in the L1 reading, Chinese bilinguals have longer fixation durations on all measures and lower skipping probabilities than Dutch bilinguals in the L2 reading. What differs more is that the FE was significantly larger for Chinese than for Dutch bilinguals throughout all time measures. Although the effect was smaller in Chinese bilinguals in the skipping probability, it was due to a floor effect rather than a ceiling effect. As Figure 3 and 5 illustrates, Chinese bilinguals are somewhat slower than Dutch bilinguals, also when reading HF words. That is, even when two groups have similar proficiency in the

L2 and read the same material in the same L2, bilinguals whose languages are from different writing systems are less efficient at visual word processing than those from one writing system.

One possible explanation for the findings is that Chinese bilinguals have relatively less exposure to alphabetic writing system (i.e., letters and their specific combinations in orthographic structures such as bigrams or trigrams). Indeed, Dutch and Chinese bilinguals were found to have comparable proficiency level in English based on the LexTALE scores (see Sui et al., 2022), and so should their exposure to the L2. Yet, it is worth noting that Dutch and English have many similarities in writing compared to Chinese and English. They use the same Latin alphabet and even share some of the underlying orthographic structures. Therefore, in this particular context it might not be the exposure to English as L2, but the exposure to *alphabetic languages* which explains the reported differences between the Dutch and Chinese group in the FE. The impact of the potential facilitation effect may be greater on LF than on HF words, as the latter may already be approaching the ceiling effect. This possibility may explain why Chinese bilinguals have longer English reading times overall, even for HF words, and why their FEs were larger than those of Dutch bilinguals in both early and late measures.

Another possible explanation could be cross-lingual lexical interactions. It is well known that the languages of bilinguals are activated even in unilingual sentence reading (Dijkstra & Van Heuven, 2002; Schwartz & Kroll, 2006; Brysbaert & Duyck, 2010). Word recognition in a target language is influenced by words in non-target languages, with cognate and cross-language neighborhood effects across languages, etc. (e.g., Cop, Dirix et al., 2017b; Dirix et al., 2017). Chinese characters, however, are fundamentally different from the Latin alphabet. The cross-language effects between them are very likely to be much more limited than those of the same writing system. Previous studies have shown that the greater within- or/and cross-language neighborhood density, the smaller the L2 FE of the LF words (Dirix et al., 2017; Whitford & Titone, 2019). Given there are no cross-language neighbors between Chinese and English, one can infer that Chinese-English bilinguals should have slower reading speeds and a larger FE than those with the same script, compatible with what we found.

The interaction between frequency and language proficiency was significantly different between the two groups. Significant interactions were found in GD and TRT in Chinese bilinguals and at the early processing stage in Dutch (i.e., FFD and GD). The absence of interaction in the very early measure of Chinese bilinguals may be related to the multiple fixations they adopted, which we will discuss below. It is, of course, also possible that language proficiency does not affect the earliest stages of word recognition in Chinese bilinguals, such as the sub-lexical orthographic stage. The disproportionate effect of language proficiency on word processing was greater for LF words in Chinese bilinguals, compatible with previous findings (e.g., Whitford & Titone, 2012; Mor & Prior, 2020).

However, Dutch bilinguals showed a different pattern, with language proficiency having a greater impact on the processing of HF words, negatively correlated in the reading measures and positively correlated in the skipping probability. This finding was indeed unexpected. Considering that the patterns we found in L1, and Chinese bilinguals are congruent with previous findings (Whitford & Titone, 2012 in French-English bilinguals; Mor & Prior, 2020 in Hebrew–English bilinguals), we speculate that it is due to the diverse language environment to which bilinguals are exposed. Some bilinguals with high language proficiency may be exposed to LF L2 words as frequently as those with low language proficiency but are exposed to HF words more frequently. Thus, language proficiency particularly affects the recognition of HF words in those bilinguals, consistent with the results of this study.

One may argue that the reason for not finding the language proficiency effect in LF words in Dutch bilinguals is due to the influence of L1 proficiency, as obtained previously (Whitford & Titone, 2019). Since the L1 should have greater language exposure, its impact on the L2 frequency should be stronger than vice versa. Thus, different results were found in the L1 and L2 of Dutch bilinguals. Whereas Chinese bilinguals should be less likely to be greatly influenced by their L1 proficiency, as proficiency in distinct languages is unlikely to influence each other (Mor & Prior, 2020). To examine this possibility, we performed an additional analysis on Dutch bilinguals. Given that their L1 and L2 proficiency are highly correlated ($r = 0.69$), we only analyzed the effect of the L1 proficiency on the L2 frequency rather than taking both language proficiencies in the analysis. We followed the same analysis procedure, except for replacing the L2 proficiency with the L1 proficiency. The results show that L1 proficiency has a greater effect on LF words than on HF words in GD ($\beta = 0.000156$, SE = 0.000030, t-value = 5.234; also see Fig.A.4), congruent with Cop et al. (2015). Yet such effect was not present in FFD ($\beta = -0.000014$, SE= 0.000085, t-value = -0.167). In this case, the L1 proficiency of Dutch bilinguals does not reduce but increases the difference in fixation duration of LF words between high- and low-language proficiency bilinguals, arguing against this possibility.

The word length effect was larger for Chinese bilinguals than for Dutch across eye-tracking measures. One explanation could be that Chinese bilinguals are not as accustomed to reading long words as Dutch bilinguals, as the average word length in English is usually longer than in Chinese and shorter than in Dutch. As a result, Chinese bilinguals are more susceptible to word length when reading a L2. Although the frequency and word length had greater effects on Chinese bilinguals, the FE was more strongly influenced by word length in Dutch bilinguals, presumably because of more variation of this measure in Dutch. Frequency curves shifted upwards with word length, more with word length in Chinese bilinguals, and more in the LF ranges in Dutch bilinguals. In the FFD of Chinese bilinguals, however, the interaction between frequency and word length showed an interesting pattern. The FE decreases with increasing word length, which is manifested by an increase in the fixation duration of HF words.

Given that this pattern occurs only at very early stages of word recognition, we speculate that the number of word fixations may affect the interaction. Further analyses were conducted using the same analysis procedure, with fixation counts as the dependent variable. The results showed an interaction between frequency and word length in the Chinese group (β = -0.0009618, SE = 0.0002217, t-value = -4.338), which was significantly different from that in the Dutch group (β = -0.001433, SE = 0.000352, t-value = -4.07). Chinese bilinguals re-fixated LF long words more frequently than short or HF words or Dutch readers (see Fig.A.2 in APPENDIX C). That is, when Chinese readers encounter a difficult word, such as a LF word, they will process it through multiple re-fixations rather than a single fixation. Therefore, the word length effect is not evident in LF words at the earliest processing stages (i.e., FFD) but in GD and TRT.

Theoretical discussion

We discussed three theoretical accounts of the mechanism explain FEs in bilingual reading, all formulated for alphabetic language reading. The learning hypothesis indicates that FEs reveal learning that becomes progressively smaller with increased word occurrences. The lexical entrenchment hypothesis states that lexical representations strengthen as word occurrences increase. The frequency-ranked hypothesis suggests that words are frequency-ordered in bins and that serial searching begins with the highest-frequency word. Its extension for bilinguals assumes that bins are either language-specific, with potential different search speeds, or shared across languages.

These models do not differentiate in their assumptions regarding distinct writing systems. They all predict that the FE becomes smaller with increasing language exposure.

That is, bilinguals with higher exposure to the language should exhibit smaller FEs than those with limited language exposure. In addition, groups with similar language exposure or proficiency should exhibit comparable FEs, regardless of their language writing systems and language dominance (L1 or L2). In the present study, the effect of language exposure on L1 FE was found in Dutch bilinguals, but not in Chinese bilinguals. Additionally, Chinese bilinguals reported much smaller L1 FEs than Dutch bilinguals in all reading time measures. However, although Chinese- and Dutch-English bilinguals had similar L2 proficiency levels (assessed using the same language proficiency test), i.e., similar exposure to the L2, their L2 FEs were quite different. These findings could not be explained by the above FE hypotheses, thereby challenging their applicability for logographic writing systems.

Another extended frequency-ranked hypothesis states that the bins are language-specific for non- and alphabetic language pairs but shared for alphabetic language pairs. Thus, one might expect that Chinese- and Dutch-English bilinguals might show comparable FEs in their L1 when Dutch bilinguals have much lower exposure to their L2 than to their L1. However, if their exposure to the L2 is not entirely lower than their L1, then their L1 FE should be greater than that of Chinese bilinguals, consistent with the findings obtained. However, this hypothesis fails to explain why the FE of Chinese words is not affected by language exposure. More importantly, it predicts that Dutch bilinguals should have a larger FE in their L2 than Chinese bilinguals. However, this prediction contradicts the results of the present study. The Dutch bilinguals in this study showed much smaller, rather than larger, FEs in their L2 than Chinese bilinguals.

An alternative theoretical account for the present findings of FE is the language-competition hypothesis (Diependaele et al., 2013; also see Dijkstra & Van Heuven, 2002), which suggests that co-activated representations compete for selection across languages. In this case, the FE in L2 may be larger than in L1 as the interference from dominant L1 representations is likely greater than vice versa. Additionally, Chinese-English bilinguals should exhibit smaller FEs than same-alphabet bilinguals due to less cross-language competition between languages with different writing systems. This hypothesis explains the FE observed in the L1 in this study. However, it fails to explain the findings in the L2 and the differences in FEs within monolinguals and bilinguals of alphabetic languages. For example, the FEs in the same L2 do not vary with the orthographic similarity between the first and second alphabetic languages (Diependaele et al., 2013). Clearly, all the existing hypotheses on word FEs fail to account for the findings obtained in the present study.

The Word Frequency Hypotheses: Implementation and Limitations

So far, the existing hypotheses on the FE have generally been considered to be language universal. Indeed, the effect has been viewed as evidence of the similarity between Chinese and alphabetic writing systems in underlying processes (e.g., Li et al., 2014; but see Sui et al., submitted). Yet, this study shows that the FEs differ in L1 across writing systems and vary in L2 due to differences in L1 writing systems, so different underlying mechanism in visual word processing may be at play for Chinese- and Dutch-English bilinguals. A key point in explaining the present findings is thus to consider the influences of the word components (e.g., characters or letters), which inevitably affect word recognition and FEs as they are the constituents of a word.

Firstly, as previously discussed, character frequency moderates the Chinese word FE, as HF characters may facilitate the recognition of LF words and cause interference with HF words (Sui et al., submitted). Thus, the effect of Chinese character frequency can explain the smaller word FE in reading Chinese as a first language than in Dutch. Secondly, languages within the same writing system apparently share some word components (e.g., bigrams, trigrams), which can affect word processing (e.g., Kuperman et al., 2008; New & Grainger, 2011). Thus, the possible explanation for the L2 FEs varying with the L1 writing systems is that the same-script bilinguals could have more exposure to language-shared word components than those with different scripts and thus exhibit a smaller FE.

Considering the effect of word components, the learning and lexical entrenchment account, but not the rank hypothesis, could explain the present findings. That is, word frequency affects activation thresholds or baselines, as the learning hypothesis indicates, or affects the entrenchment of lexical representations, as the lexical entrenchment hypothesis suggests. The language-shared alphabetic components may have effects similar to those described above for word frequency or provide extra activation for accessing target words. This facilitative effect should affect the FE of both the L1 and L2 and is expected to be prominent on LF words but limited on HF words that are already close to the threshold.

In contrast, Chinese characters usually act as single-character words with word frequency. Hence, they may only provide additional activation rather than having a similar influence as words. Since Chinese characters not only facilitate the recognition of LF words but also cause interference with HF words, the word FE of Chinese as the L1 is expected to be smaller than that of alphabetic languages. Furthermore, given the limited morphological similarity between Chinese and English, and thus the absence of language-shared components, Chinese-English bilinguals should have longer overall fixation durations and larger FEs in L2 than Dutch-English bilinguals. Clearly, with the extended assumptions of the word components effect, the learning and lexical entrenchment hypothesis can explain the current findings.

Conclusion

This work examined the word FEs of Chinese and Dutch bilinguals in L1 and L2 reading. It showed that even after considering language proficiency, Chinese bilinguals still have much smaller and larger FEs than Dutch bilinguals in the L1 and L2 reading, respectively. These results further confirm, that the underlying processes of word FEs are indeed different in Chinese and alphabetic languages. Furthermore, this indicates that the L1 writing system affects L2 reading but that some phenomena are constant. The results of this study fill an important gap of empirical evidence on bilingual natural reading of alphabetic and non-alphabetic languages.

Reference

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye Movements of Highly Skilled and Average Readers: Differential Effects of Frequency and Predictability. *The Quarterly Journal of Experimental Psychology Section A*, 58(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*, 133(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, 20(3), 530-548. doi:10.1017/S1366728916000353

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual memory after 15 years of service? *Bilingualism: Language & Cognition*, 13(3), 359-371.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 441–458. https://doi.org/10.1037/xhp0000159

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5, e10729. https://doi.org/10.1371/journal.pone.0010729

Chinese Proficiency Test. (n.d.) HSK Level VI. http://www.chinesetest.cn/gosign.do?id=1&lid=0#

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Elsevier. https://doi.org/10.1016/B978-008044980-7/50017-3

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017a). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, 49(2), 602-615.

Cop, U., Dirix, N., Van assche, E., Drieghe, D., & Duyck, W. (2017b). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, 20(4), 747-769. doi:10.1017/S1366728916000213

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, 22(5), 1216–1234.

Cui, L., Wang, J., Zhang, Y., Cong, F., Zhang, W., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology*, 74(4), 610-633.

Cui, L., Yan, G., Bai, X., Hyönä, J., Wang, S., & Liversedge, S. P. (2013). Processing of compound-word characters in reading Chinese: An eye-movement-contingent display change study. *Quarterly Journal of Experimental Psychology*, 66(3), 527-547.

Da Silveira, A. P., & van Leussen, J. W. (2015). Generating a bilingual lexical corpus using interlanguage normalized Levenshtein distances. In *ICPhS*.

De Groot, A. M. B. (2010). Comprehension Processes: Word Recognition and Sentence Processing. In A. M. B. De Groot (Eds.), *Language and Cognition in Bilinguals and Multilinguals* (pp. 155–219). New York, NY: Psychology Press.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197.

Dirix, N., & Duyck, W. (2017). The first-and second-language age of acquisition effect in first-and second-language book reading. *Journal of Memory and Language*, 97, 103-120.

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior Research Methods*, 51(6), 2800–2816.

Dirix, N., Cop, U., Drieghe, D., & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *Journal of experimental psychology. Learning, memory, and cognition*, 43(6), 887–915. https://doi.org/10.1037/xlm0000352

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. (2008). The frequency effect in second-language visual word recognition: evidence against a language-selective rank hypothesis. *Psychonomic Bulletin and Review*, 15(4), 850-855.

Ferrand L, New B, Brysbaert M, Keuleers E, Bonin P, Méot A, Augustinova M, Pallier C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42, 488–496. http://dx.doi.org/10.3758/BRM.42.2.488

Keuleers, E., Brysbaert, M. & New, B. (2010b). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, 42(3), 643-650.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010a). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, 1, 174. http://dx.doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44, 287–304. http://dx.doi.org/10.3758/s13428-011-0118-4

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23, 1089–1132. https://doi.org/10.1080/01690960802193688.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, 66(3), 563–580.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 802–823.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325–343.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, 127(6), 1139–1162.

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), 895–913.

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, 147, 1-20.

Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition*, 22(3-4), 441–457. https://doi.org/10.1080/13506285.2014.889260

Liu, Y., Yu, L., Fu, L., Li, W., Duan, Z., & Reichle, E. D. (2019). The effects of parafoveal word frequency and segmentation on saccade targeting during Chinese reading. *Psychonomic bulletin & review*, 26(4), 1367-1376.

Ma, G., Li, X., & Rayner, K. (2015). Readers extract character frequency information from nonfixated-target word at long pre-target fixations during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 41(5), 1409-1419.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Lawrence Erlbaum Associates, Inc.

Mor, B., & Prior, A. (2020). Individual differences in L2 frequency effects in different script bilinguals. *International Journal of Bilingualism*, 24(4), 672–690. https://doi.org/10.1177/1367006919876356

Morton, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Murray, W. S., & Forster, K. I. (2004). Serial Mechanisms in Lexical Access: The Rank Hypothesis. Psychological Review, 111(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

New, B., & Grainger, J. (2011). On letter frequency effects. *Acta Psychologica*, 138(2), 322-328.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, 3(2), 245-248.

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., Sereno, S. C.,Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, 4(special issue), 21–49.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of memory and language*, 55(2), 197-212.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11), 489–493.

Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods*, 1-21.

Sui, L., Woumans, E., Duyck, W., Dirix, N., (submitted) Word Frequency Effects Differ Across Writing Systems: The Word and Character Frequency Effect in Chinese Reading. *Language Learning*.

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176-1190.

Whitford, V., & Titone, D. (2017). The effects of word frequency and word predictability during first- and second-language paragraph reading in bilingual older and younger adults. *Psychology and Aging*, 32(2), 158–177. https://doi.org/10.1037/pag0000151

Whitford, V., & Titone, D. (2019). Lexical entrenchment and cross-language activation: Two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Language and Cognition*, 22(1), 58-77. doi:10.1017/S1366728917000554

Whitford, V., & Titone, D.A. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, 19, 73-80. 10.3758/s13423-011-0179-5.

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.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60, 502–529. http://dx .doi.org/10.1016/j.jml.2009.02.001

Zang, C., Zhang, M., Bai, X., Yan, G., Paterson, K. B., & Liversedge, S. P. (2016). Effects of word frequency and visual complexity on eye movements of young and older Chinese readers. *Quarterly Journal of Experimental Psychology*, 69(7), 1409–1425. https://doi.org/10.1080/17470218.2015.1083594

CHAPTER 5. GENERAL DISCUSSION

The objective of the present dissertation was to advance our understanding of the linguistic and cognitive processes involved in reading from the perspective of Chinese-English bilinguals. It focused on exploring the reading process of Chinese as a native language, as well as the effect of the first-language writing system on second-language processes. The investigation of the reading performance of Chinese-English bilinguals undoubtedly has important theoretical implications. Currently, most available theories and hypotheses are based on findings from alphabetic languages and can explain the data of those monolinguals and bilinguals well (e.g., frequency effect, e.g., Duyck et al., 2008). It is imperative to examine their applicability to non-alphabetic readers, such as Chinese-English bilinguals. By doing so, existing theories and hypotheses related to reading and bilingualism (e.g., the rank hypothesis of frequency effect, Murray & Forster, 2004), especially those aimed at explaining universal behavior across languages or bilinguals, can be examined, evaluated, and refined. In addition, it allows for gaining a more comprehensive and accurate understanding of the universality and diversity of language processes.

An essential prerequisite for exploring the performance of Chinese-English bilinguals in natural reading in their two languages is to have a high-quality eye-movement database with greater statistical power. The corpus needs to use natural, unmodified reading materials in both Chinese and English. Ideally, it would also allow comparisons between bilinguals with different first-language writing systems. Therefore, this dissertation introduced the first-ever Chinese-English bilingual eye-tracking corpus in Chapter 2, in which participants read the entire novel in paragraphs in their first and second languages. Secondly, since Chinese characters, the language-specific factors, may affect word recognition and its effects, the underlying processes in Chinese may differ somewhat from those in alphabetic languages, even for the robust word frequency effect. Thus, this dissertation examined the frequency effects of words and characters in Chinese sentence reading by analyzing their frequencies as continuous variables in Chapter 3.

Finally, the available hypotheses of word frequency effect are based primarily on studies of alphabetic languages (e.g., the lexical entrenchment hypothesis, Diependaele et al., 2013). Different hypotheses differed in their postulations about word frequency effects for bilinguals with the same and disparate writing systems (e.g., Chinese- and Dutch-English bilinguals). To verify these hypotheses, this dissertation compared word frequency effects in the first and second languages between Chinese- and Dutch-English bilinguals in Chapter 4. In the following sections, we will restate the main findings from this project, followed by discussions of the limitations and potential avenues for further research.

Recapitulation of the main findings

A corpus investigation of the overall reading performance of Chinese-English bilinguals

Chapter 2 presented an eye-tracking corpus of Chinese-English bilinguals reading in both languages and revealed several noteworthy results. Firstly, Chinese readers exhibited similar fixation durations on words as Dutch bilinguals and English monolinguals when reading in their native language. This finding contradicts previous results that Chinese readers have longer fixation duration compared to alphabetical language readers (e.g., Feng et al., 2009). They also demonstrated that bilinguals did not read slower than monolinguals in their native language due to bilingualism (also see Cop et al., 2015). Secondly, Chinese readers showed a significantly higher rate of skipping words than Dutch-English bilinguals and previous research on Chinese reading, which is different from the previous findings (e.g., Yan et al., 2006). However, it should be noted that there were no significant differences in the comprehension scores between Chinese and Dutch bilinguals. It indicates that the higher skipped rate of Chinese readers may be attributed to language differences rather than compensating for comprehension. Overall, these novel findings suggest that Chinese readers could read efficiently and rapidly even without inter-word spaces and with the presents of word ambiguity. Furthermore, they reinforce the indication that performance obtained in natural reading may differ from those found in controlled studies.

Consistent with what was found in alphabetic languages (e.g., Cop et al., 2017), Chinese-English bilinguals took longer to read in their weaker second language than in their first language. In addition, the frequency and word length effects, regarded as the crucial predictors of reading behavior in alphabetic languages, were also observed in Chinese reading. It supports the assumption of words as the primary unit of sentence reading in Chinese (also see Li et al., 2014). Furthermore, Chinese-English bilinguals have been found to recognize high-frequency or shorter words more quickly in their second language (also see Wei et al., 2013). Together, these findings indicate that despite the vast differences between Chinese and alphabetic languages and the expected different influences from/on the same second language, there may be similarities in the processes involved in reading between bilinguals with different first-language writing systems. These results are theoretically important for exploring the universality of word recognition processes. They also have practical implications for improving reading efficiency across populations.

Word and Character frequency effects in Chinese reading

Chapter 3 investigated the word and character frequency effects by accessing them as continuous variables on Chinese text reading and found several important and novel results. First, the study confirmed the existence of a word frequency effect in Chinese reading, whereby high-frequency words are processed faster than low-frequency words in general. It implies that word frequency is a universal predictor that influences word recognition across writing systems and languages. Secondly, the present study observed a character frequency effect, with the first character demonstrating a facilitative frequency effect and the second character showing an overall inhibited frequency effect. Such findings suggest that although words are the primary processing unit in Chinese reading, the properties of characters can still influence word recognition. Moreover, it suggests that Chinese word processing differs from that of alphabetic languages due to the inclusion of such language-specific factors.

The second noteworthy finding of this study is the observation of the second character (frequency) effect, which had not been previously reported in the literature (e.g., Cui et al., 2021; Yan et al., 2006). Previous experiments may have failed to obtain this effect due to the use of controlled reading materials, such as sentences with the same frame but varying in target words only (e.g., Cui et al., 2021) or the examination of only the first character frequency (e.g., Yu et al., 2021), as mentioned in Chapter 3. Hence, the current study highlights the importance of utilizing natural reading materials to avoid the potential effects of Chinese word segmentation on word recognition.

Finally, the study uncovered a novel finding of the interaction between word frequency effect and character frequency and vice versa. The results showed that as character frequency increased, the facilitative word frequency effect gradually diminished and even showed an inhibitory trend. Similarly, as word frequency increased, the facilitative character frequency effect disappeared and showed an inhibitory effect. These findings can explain the inconsistent character frequency effects in previous studies, some of which observed facilitative (Yan et al., 2006) or inhibitory effects (e.g., Cui et al., 2021) while others failed to find them (e.g., Li et al., 2014). Furthermore, the present study highlights the limitations of categorizing continuous variables into binary classification (also see Balota et al., 2004), indicating that it may not capture variations in effects, such as the Chinese word and character frequency effects.

Comparisons of frequency effects between Chinese- and Dutch-English bilinguals

Chapter 4 compared the frequency effects in first and second-language reading between Chinese-English and Dutch-English bilinguals to examine the frequency effect hypotheses (i.e., the learning, lexical entrenchment, frequency-ranked and language-competition hypotheses, Morton, 1970; Diependaele et al., 2013; Murray & Forster, 2004). The results showed that Chinese bilinguals spent similar amounts of time reading in their first language as Dutch-English bilinguals but had a higher rate of skipping. However, they took longer to read and had lower skip rates than Dutch-English bilinguals in second-language reading. The word length effect was found in both languages of Chinese and Dutch bilinguals, indicating faster processing of short words than long words. Notably, the word length had an impact on frequency effects, with a more pronounced increase observed in Dutch bilinguals compared to Chinese bilinguals in both their first and second language readings as word length increased. Furthermore, Chinese-English bilinguals exhibited smaller frequency effects in first-language reading than Dutch-English bilinguals but larger effects in second-language reading. In addition, Chinese-English bilinguals showed larger frequency effects in second-language reading than in first-language. Furthermore, the frequency effect was influenced by language exposure in alphabetic languages but not in Chinese. The effect decreased with increased language exposure in Dutch-English bilinguals and the second language of Chinese-English bilinguals.

These results, however, are not entirely consistent with the predictions of existing frequency effect hypotheses based on alphabetic language data. With increasing occurrences of a word, it may be that the amount of learning becomes progressively smaller, as the learning hypothesis (e.g., Morton, 1970) suggests, or that the lexical representations are strengthened, as the lexical entrenchment hypothesis (also see Diependaele et al., 2013) indicates, or that it ranks higher in the bins and is therefore searched for more quickly, as the frequency-ranked hypothesis (Murray & Forster, 2004) states. The frequency-ranked hypothesis with extended assumptions to account for the performance of bilinguals suggests that bins are either language-specific with different scan speeds or shared across languages with the same speed (Duyck et al., 2008). According to these hypotheses, word frequency effects are not influenced by language writing systems (e.g., Chinese or Dutch) or dominance (e.g., the first or the second language) but by language exposure. That is, frequency effects for unbalanced bilinguals should be larger in their relatively less proficient second language than in their native languages and should be similar in size between Chinese- and Dutch-English bilinguals in their first and second languages.

Another extension of the frequency-ranked hypothesis suggests that bilinguals with different writing systems may have separate bins, while bilinguals with alphabetic languages have shared bins (Duyck et al., 2008). If this is the case, then the frequency effect for Chinese bilinguals should be similar or smaller than for Dutch bilinguals in first-language reading and smaller in second-language reading. Frequency effects were indeed greater in the second language than in the first language, both in the Chinese-English bilinguals in this study and in the alphabetic language bilinguals in previous studies (e.g., Cop et al., 2015). However, the effects were smaller in the first language and larger in the second language for Chinese-English bilinguals than for Dutch-English bilinguals, inconsistent with the frequency effect hypothesis discussed above. Apparently, these findings indicated the limitation of existing frequency effect hypotheses in explaining the performance of non-alphabetic languages unless with further assumptions. Additionally, the comparison between Chinese and Dutch bilinguals in this work suggested that frequency effects in the same second language (i.e., English) appear to be influenced by the writing system of the first language.

Directions for Future Research

Given the theoretical and practical implications of the research findings presented in this dissertation, as well as utilizing the eye-movement corpus of Chinese-English bilinguals presented in Chapter 2, there are several main areas for future research. First, validating the phenomena observed in experimental or controlled studies through natural reading (e.g., neighborhood effect, also see Chapter 3) and explore unknown variables or processes involved in natural reading. In everyday life, we often come across readings of passages. However, such readings involve various factors that affect word recognition (e.g., contextual context), resulting in different phenomena than those observed in experimental or controlled studies (also see Kuperman et al., 2013; Dirix et al., 2019). It should be particularly the case in Chinese reading, where word segmentation is required and affect word recognition. Second, explore the similarities and differences in language processing across writing systems (e.g., Chinese- and Dutch-English). Due to the significant differences between non-alphabetic and alphabetic languages (e.g., morphology), the underlying processes involved in their reading may be different (e.g., involve language-specific factors, like the number of strokes).

Here, we list several specific research directions. The first direction for future research is to expand the GECO family (i.e., GECO and GECO-CN, Cop et al., 2017, Sui et al., 2022). In related research, the Multilingual Eye-movement Corpus (MECO, Siegelman et al., 2022; Kuperman et al., 2022) includes English monolinguals and numerous bilinguals with diverse native languages (mostly European languages) and English as their second language. The corpus has 12 texts (one-screen length) each in first and second languages, with 1487-2412 tokens in the first-language materials and 1,653 in the second-language. However, the GECO family of exploring novel reading (with approximately 59,000 and

55,000 tokens in first- and second-language reading materials, respectively) currently only has data for Dutch- and Chinese-English bilinguals as well as English monolinguals.

Although they allow for comparisons between highly distinct writing systems (e.g., Chinese and Dutch), they cannot be employed to investigate the similarities and differences within the same writing system (e.g., Dutch and French). In addition, most existing eye-tracking databases for bilinguals (e.g., MECO projects, Siegelman et al., 2022) focus on those with English as their second language (but see the Bilingual Russian Sentence Corpus, Parshina, 2020, which had the second language (Russian) data only). Therefore, the use of these databases is limited in exploring the generalization of bilingual language processing, such as language processing in first and second languages, and the mutual influence between them. Future research can expand the GECO family by introducing more eye-tracking databases of bilingual speakers with diverse native (e.g., English) and second languages (e.g., Chinese).

The second direction for future research is to gain a deeper understanding of the Chinese reading process by differentiating the influence of the second language on the first language (i.e., Chinese). It is well-known that bilinguals' two languages are co-activated and influence each other during processing (e.g., Dijkstra & Van Heuven, 2002; Van Assche et al., 2009). That is, the non-target language interferes with the recognition of Chinese words by Chinese-English bilinguals (also see Wu et al., 2013). Therefore, studying the reading performance of Chinese monolinguals is optimal for investigating Chinese reading. However, the proportion of Chinese youth who speak only one language is quite limited. It is because English is one of the compulsory courses in schools and universities in China. As a result, most educated Chinese adults have English as their second language, thus making it very difficult to recruit young monolinguals. It is important to note that most previous studies on Chinese word recognition have recruited young adults (university students, e.g., Yan et al., 2006; Ma et al., 2014) who should possess knowledge of a second language as participants. Accordingly, the findings of these studies incorporate second language influences and may lead to a biased understanding of Chinese word recognition. Thus, exploring the second-language influence on first-language reading can enhance our understanding of Chinese word recognition and bilingual processing, ultimately advancing the fields of Chinese reading and bilingualism.

The third direction for future research is to explore the predictability effect in a narrative context within and across groups. The predictability of a word, depending on the preceding context, affects the speed of word recognition (e.g., Rayner et al., 2005, 2006).

Words with higher predictability tend to be fixated on shortly and are more likely to be skipped (also see Rayner et al., 2005). However, Chinese reading involves word segmentation, which can impact word recognition. One potential research direction is to explore the predictability effect on Chinese word segmentation. As reading in everyday life often involves context, previous information may facilitate word segmentation speed (e.g., possibly by increasing the activation of a word candidate).

Another research direction is to investigate predictability effects. The reading material in GECO-CN (Sui et al., 2022) is a coherent novel containing numerous sentences where certain words, such as objects (e.g., strychnine) and locations, are frequently repeated. Therefore, the predictability of some words may vary across different sentences. It allows us to examine the predictability of target words in sentence reading, which could be a research direction. In addition, the recognition of a word may be affected by its frequency of appearance (e.g., the repetition effect of words in the whole novel or sections). However, how the word repetition effect affects reading efficiency remains unknown (e.g., affect fixation duration or skipping probabilities) and can be another research direction. In addition, the predictability effect in first-language reading may differ between Chinese and Dutch bilinguals due to the word segmentation in Chinese reading, which is another research direction. Furthermore, existing research suggests that readers do not anticipate upcoming words in second-language as they do in first-language reading (e.g., Martin et al., 2013, but see Whitford & Titone, 2017). Future research could also investigate whether the predictability effect in a second language varies based on the first-language writing system.

The fourth direction for future research is to explore the within-linguistic neighborhood effect. The neighborhood effect refers to the influence of words that differ from the target word by only one letter (e.g., <u>ca</u>t and <u>ca</u>p) on the target word recognition (Dirix et al., 2017). The effect has been observed in both alphabetic languages (e.g., Whitford & Titone, 2019) and Chinese (Huang et al., 2006; Yao et al., 2021). However, an interesting question emerges as to whether the neighborhood effect in Chinese reading is fundamentally similar to that in alphabetical languages. As mentioned earlier, Chinese words are composed of characters, while characters are formed by a certain number of strokes. In addition, Chinese words are typically short, with one or two-character words being the most common word types (also see Li & Su, 2022; Sui et al., 2022). Hence, most Chinese studies have investigated the orthographic neighborhood effect in two-character words, and words that differ from the target word by only one character are considered orthographic neighbors (e.g.,

花园(garden) and 花店 (flower store), Yao et al., 2021). However, it is important to note that Chinese characters are far more complex than the letters in alphabetical languages (e.g., Chinese characters have pronunciations and meanings). Therefore, it remains unclear whether the nature of the orthographic neighborhood effects in Chinese and alphabetic languages is analogous. Future research can explore this issue for a comprehensive understanding of the universality of language processes.

Conclusions

In summary, the present dissertation showed that the word recognition of Chinese-English bilinguals differs from those with alphabetic languages. This dissertation provided a freely accessible database (GECO-CN) for exploring the performance of Chinese-English bilinguals and presented novel findings that the Chinese word frequency effect is affected by character frequency and vice versa. It also showed that Chinese-English bilinguals exhibited smaller word frequency effects in the first-language reading and larger effects in the second-language compared to Dutch-English bilinguals. In addition to examining existing hypotheses on word frequency effects, this research also sheds light on new avenues and cautionary for future research that many available hypotheses may fail to explain the performance of non-alphabetic languages and that similar patterns observed in alphabetic and non-alphabetic languages may result from different underlying processes.

References

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*, 133(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, *49*(2), 602–615. https://doi.org/10.3758/s13428-016-0734-0

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review, 22(5), 1216–1234.*

Cui, L., Wang, J., Zhang, Y., Cong, F., Zhang, W., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology*, 74(4), 610-633.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly*

Journal of Experimental Psychology, 66(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197.

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior Research Methods*, 51(6), 2800–2816.

Dirix, N., Cop, U., Drieghe, D., & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(6), 887.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. (2008). The frequency effect in second-language visual word recognition: evidence against a language-selective rank hypothesis. *Psychonomic Bulletin and Review*, 15(4), 850-855.

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.

Huang, H. W., Lee, C. Y., Tsai, J. L., Lee, C. L., Hung, D. L., & Tzeng, O. J. L. (2006). Orthographic neighborhood effects in reading Chinese two-character words. Neuroreport, 17(10), 1061-1065.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, 66(3), 563–580.

Kuperman, V., Siegelman, N., Schroeder, S., Alexeeva, A., Acartürk, C., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Fonseca, S., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C., Kalaitzi, A., Lõo, K., Marelli, M., … Usual, K. A. (2022). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus (MECO). *Studies in Second Language Acquisition*, 1–35.

Li, X., & Su, X. (2022). *Lexicon of common words in contemporary Chinese Second Edition*. The Commercial Press.

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), 895–913.

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. Journal of Experimental Psychology: Human Perception and Performance, 40(3), 1046–1059. https://doi.org/10.1037/a0035389.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

Morton, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Murray, W. S., & Forster, K. I. (2004). Serial Mechanisms in Lexical Access: The Rank Hypothesis. Psychological Review, 111(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

Parshina, O. (2020). Bilingual reading fluency and prediction: Heritage language versus second language (Doctoral dissertation, City University of New York).

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., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465.

Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., Ahn, H., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Fonseca, S., Dirix, D., Duyck, W., Fella, A., Frost, R., Gattei, C., Kalaitzi, A., Kwon, N., Lõo, K., … Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research* Methods, 54. 1-21. 10.3758/s13428-021-01772-6.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. Psychological science, 20(8), 923-927.

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.

Whitford, V., & Titone, D. (2017). The effects of word frequency and word predictability during first- and second-language paragraph reading in bilingual older and younger adults. *Psychology and Aging*, *32*(2), 158–177. https://doi.org/10.1037/pag0000151

Whitford, V., & Titone, D. (2019). Lexical entrenchment and cross-language activation: Two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Language and Cognition*, 22(1), 58-77. doi:10.1017/S1366728917000554

Wu, Y. J., Cristino, F., Leek, C., & Thierry, G. (2013). Non-selective lexical access in bilinguals is spontaneous and independent of input monitoring: Evidence from eye tracking. Cognition, 129(2), 418-425.

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.

Yao, P., Staub, A., & Li, X. (2021). Predictability eliminates neighborhood effects during Chinese sentence reading. Psychonomic Bulletin & Review, 1-10.

Yu, L., Liu, Y., & Reichle, E. D. (2021). A corpus-based versus experimental examination of word- and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of Experimental Psychology: General*, 150(8), 1612–1641. https://doi.org/10.1037/xge0001014

CHAPTER 6. ENGLISH SUMMARY

Reading is undoubtedly an important means of accessing information. Numerous studies have investigated reading and revealed that word recognition involves complex processes, particularly for bilinguals who speak two languages. Thus, the present research examined word recognition processes from the perspective of Chinese-English bilinguals, whose two languages have distinct morphologies. Most proposed theories and hypotheses on reading and bilingualism are based on studies of alphabetic languages. Therefore, the added value of this dissertation lies in the examination of the applicability of said theories to non-alphabetic language readers or bilinguals, as it is crucial in the development of universal reading theories.

This dissertation introduced the first natural reading eye-tracking corpus of Chinese-English bilinguals in Chapter 2, in which 30 native Chinese adults with English as a second language read an entire novel in both languages. The dissertation also provided their overall reading performance in both languages. In Chapter 3, this dissertation investigated the frequency effects of Chinese words and characters in natural reading by using data from GECO-CN with word and character frequencies as continuous variables. It also explored whether language-specific factors (i.e., character frequency) affect the word frequency effect in Chinese reading. In Chapter 4, this work compared the word frequency effect between bilinguals with different first-language writing systems but with English as their second language. The study discussed whether existing frequency effect theories could account for the findings observed in first- and second-language reading.

The current findings allow for the drawing of three important conclusions. Firstly, despite the significant differences between Chinese and alphabetic languages and the potential influence of word ambiguity due to the lack of clear word boundaries in Chinese, Chinese readers exhibited similar fixation durations to their Dutch counterparts and much higher skipping rates in native language reading. These findings are inconsistent with previous research (e.g., Feng et al., 2009; Yan et al., 2006), suggesting that performance in natural reading may differ from that in controlled and experimental studies (also see Dirix et al., 2019). The frequency and word length effects, considered key predictors of alphabetic language reading behavior (e.g., Brysbaert et al., 2018), were also observed in first and second-language reading by Chinese-English bilinguals. Additionally, Chinese-English bilinguals showed longer reading times for their weaker second language. Together, these findings suggest the reading processes of bilinguals with different first-language writing systems have some similarities.

Secondly, the current findings further showed that there are similarities and differences between the Chinese and alphabetic reading processes. The word frequency effect was observed in Chinese reading, with high-frequency words being processed faster than low-frequency words, similar to that found in alphabetic languages (e.g., Cop et al., 2015; Li et al., 2014). However, the Chinese word frequency effect was also affected by character frequency, the language-specific factors, suggesting that there are fundamental differences from the underlying processes in alphabetic languages. In addition, the study observed a character frequency effect, showing that the facilitative character frequency effect became flat and reversed as word frequency increased. The findings suggest that even though Chinese words are the primary processing units (e.g., Li et al., 2014), the properties of characters can still affect word recognition (e.g., Yan et al., 2006). It also highlights differences in underlying processes between Chinese and English reading and further emphasizes the necessity of natural reading materials to explore everyday reading performance.

Finally, the present research revealed that existing frequency effect hypotheses have limitations in explaining the performance of non-alphabetic languages. Chinese-English bilinguals exhibited smaller word frequency effects in first-language reading and larger effects in the second language than Dutch-English bilinguals. Furthermore, frequency effects were influenced by language proficiency in alphabetic languages but not in Chinese reading. These findings challenge existing word frequency effect theories in alphabetical languages (e.g., the learning hypothesis, Morton, 1970; the lexical entrenchment hypothesis, Diependaele et al., 2013; the frequency-ranked hypothesis, Murray & Forster, 2004), which state that Chinese-English bilinguals should have similar or smaller word frequency effects in the first and second languages than Dutch-English bilinguals. Apparently, these frequency effect hypotheses cannot explain the performance of Chinese-English and Dutch-English bilinguals in this study unless further assumptions are made.

In conclusion, this dissertation enhances our understanding of the reading processes of Chinese-English bilinguals. In addition, it suggests various research directions for future exploration, such as verifying the phenomena observed in experimental or controlled studies through natural reading, and exploring the similarities and differences in language processing across different writing systems. Furthermore, future research could expand the GECO family by introducing more eye-tracking databases of bilinguals with various native and second languages. They could also explore the influence of the second language on the first language, as well as the predictability effects in narrative contexts and the impact of predictability on Chinese word segmentation.

Reference

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45-50.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic bulletin & review*, 22, 1216-1234.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior Research Methods, 51*(6), 2800–2816.

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.

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), 895–913.

Morton, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Murray, W. S., & Forster, K. I. (2004). Serial Mechanisms in Lexical Access: The Rank Hypothesis. Psychological Review, 111(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

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.

CHAPTER 7. NEDERLANDSTALIGE SAMENVATTING

Lezen is ongetwijfeld een belangrijk middel om toegang te krijgen tot informatie. Talrijke studies hebben leesgedrag onderzocht en aangetoond dat woordherkenning complexe processen impliceert, in het bijzonder voor tweetaligen. Daarom werd in dit onderzoek het woordherkenningsproces onderzocht vanuit het perspectief van Chinees-Engelse tweetaligen, wier twee talen verschillende morfologieën hebben. Demeeste voorgestelde theorieën en hypothesen over lezen en tweetaligheid gebaseerd zijn op studies van alfabetische talen. Daarom is het voor de ontwikkeling van universele leestheorieën van cruciaal belang dat hun toepasbaarheid op lezers van niet-alfabetische talen of tweetaligen wordt onderzocht, wat dan ook de toegevoegde waarde is van het huidige proefschrift.

Dit proefschrift introduceerde het eerste natuurlijke leescorpus van Chinees-Engelse tweetaligen in hoofdstuk 2, waarin 30 Chinese volwassenen met Engels als tweede taal een hele roman in beide talen lazen. Het proefschrift rapporteerde ook hun algemene leesprestaties in beide talen. In hoofdstuk 3 onderzocht dit proefschrift de frequentie-effecten van Chinese woorden en tekens bij natuurlijk lezen door gebruik te maken van gegevens van GECO-CN met woord- en tekenfrequenties als continue variabelen. Ook werd onderzocht of taalspecifieke factoren (d.w.z. tekenfrequentie) van invloed zijn op het woordfrequentie-effect bij Chinees lezen. In hoofdstuk 4 werd het woordfrequentie-effect vergeleken tussen tweetaligen met verschillende schrijfsystemen in hun eerste taal, maar met Engels als tweede taal. De studie besprak of bestaande frequentie-effecttheorieën de bevindingen die werden waargenomen bij eerste- en tweedetaallezen konden verklaren.

Uit de huidige bevindingen kunnen drie belangrijke conclusies worden getrokken. Ten eerste, ondanks de aanzienlijke verschillen tussen Chinees en alfabetische talen en de mogelijke invloed van woordambiguïteit door het ontbreken van duidelijke woordgrenzen in het Chinees, vertoonden Chinese lezers een vergelijkbare fixatieduur als hun Nederlandse tegenhangers en een veel hoger percentage overslagen bij het lezen in de moedertaal. Deze bevindingen zijn niet consistent met eerder onderzoek (e.g., Feng et al., 2009; Yan et al., 2006), wat suggereert dat de prestaties bij natuurlijk lezen kunnen verschillen van die in gecontroleerde en experimentele studies (also see Dirix et al., 2019). De frequentie- en woordlengte-effecten, die als belangrijke voorspellers van alfabetisch leesgedrag worden beschouwd (e.g., Brysbaert et al., 2018), werden ook waargenomen bij het lezen in de eerste en tweede taal door Chinees-Engelse tweetaligen. Bovendien vertoonden Chinees-Engelse tweetaligen langere leestijden voor hun zwakkere tweede taal. Samen suggereren deze bevindingen dat de leesprocessen van tweetaligen met verschillende schrijfsystemen in de eerste taal enige overeenkomsten vertonen.

Ten tweede toonden de huidige bevindingen verder aan dat er overeenkomsten en verschillen zijn tussen de Chinese en alfabetische leesprocessen. Het woordfrequentie-effect werd waargenomen bij Chinees lezen, waarbij hoogfrequente woorden sneller werden verwerkt dan laagfrequente woorden, vergelijkbaar met wat in alfabetische talen werd gevonden (e.g., Cop et al., 2015; Li et al., 2014). Het Chinese woordfrequentie-effect werd echter ook beïnvloed door tekenfrequentie, de taalspecifieke factoren, wat suggereert dat er fundamentele verschillen zijn met de onderliggende processen in alfabetische talen. Bovendien werd in de studie een tekenfrequentie-effect waargenomen, waaruit bleek dat het faciliterende tekenfrequentie-effect vlak werd en omkeerde naarmate de woordfrequentie toenam. De bevindingen suggereren dat, ook al zijn Chinese woorden de primaire verwerkingseenheden (e.g., Li et al., 2014), de eigenschappen van tekens nog steeds de woordherkenning kunnen beïnvloeden (e.g., Yan et al., 2006). Het belicht ook de verschillen in onderliggende processen tussen Chinees en Engels lezen en benadrukt verder de noodzaak van natuurlijk leesmateriaal om de dagelijkse leesprestaties te onderzoeken.

Ten slotte bleek uit het huidige onderzoek dat bestaande hypothesen over frequentie-effecten beperkingen hebben bij het verklaren van de prestaties van niet-alfabetische talen. Chinees-Engelse tweetaligen vertoonden kleinere woordfrequentie-effecten in de eerste taal en grotere effecten in de tweede taal dan Nederlands-Engelse tweetaligen. Bovendien werden frequentie-effecten beïnvloed door taalvaardigheid bij alfabetische talen, maar niet bij Chinees lezen. Deze bevindingen betwisten de bestaande theorieën over woordfrequentie-effecten in alfabetische talen (e.g., the learning hypothesis, Morton, 1970; the lexical entrenchment hypothesis, Diependaele et al., 2013; the frequency-ranked hypothesis, Murray & Forster, 2004), die stellen dat Chinees-Engelse tweetaligen vergelijkbare of kleinere woordfrequentie-effecten zouden moeten hebben in de eerste en tweede taal dan Nederlands-Engelse tweetaligen. Blijkbaar kunnen deze frequentie-effecthypothesen de prestaties van Chinees-Engelse en Nederlands-Engelse tweetaligen in deze studie niet verklaren, tenzij verdere veronderstellingen worden gemaakt.

Concluderend kan worden gesteld dat dit proefschrift ons begrip van de leesprocessen van Chinees-Engelse tweetaligen vergroot. Daarnaast suggereert het verschillende onderzoeksrichtingen voor toekomstig onderzoek, zoals het verifiëren van de in experimentele of gecontroleerde studies waargenomen verschijnselen door middel van natuurlijk lezen, en het onderzoeken van de overeenkomsten en verschillen in taalverwerking tussen verschillende schrijfsystemen. Verder zou toekomstig onderzoek de GECO-familie kunnen uitbreiden door meer eye-tracking databases van tweetaligen met verschillende moeder- en tweede talen te introduceren. Zij zouden ook de invloed van de tweede taal op de eerste taal kunnen onderzoeken, evenals de voorspelbaarheidseffecten in narratieve contexten en het effect van voorspelbaarheid op Chinese woordsegmentatie.

Reference

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45-50.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic bulletin & review*, 22, 1216-1234.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior Research Methods, 51*(6), 2800–2816.

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.

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), 895–913.

Morton, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Murray, W. S., & Forster, K. I. (2004). Serial Mechanisms in Lexical Access: The Rank Hypothesis. Psychological Review, 111(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

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.

APPENDIX A

Online Materials

The following five files are available online (https://osf.io/pmvhd/?view_only=77def2827a514254957cc846e14826cf), including two Chinese and English novel materials files (file 1 and 2), two eye-tracking data files of reading the first and second languages (file 3 and 4), and one file of participant language proficiency (file 5).

1) ChineseMaterial
2) EnglishMaterial
3) L1ReadingData
4) L2ReadingData
5) SubjectInformation

All files include a sheet named 'LEGEND' (in the lower-left corner after opening the file; see Image 1). This sheet consists of two columns. The first column lists all the Column Names displayed in the file; The second columns list the descriptions of the corresponding Column names.

Image 1

ChineseMaterial & EnglishMaterial

Description of the 'ChineseMaterial.xlsx' and 'EnglishMaterial.xlsx' files. The sheet and column names are in the first column, and the corresponding description of the contents of that sheet or column is presented in the second column.

Sheet Name	Description
ALL	Each word is presented on a separate line.
NOUNS	Each noun of the novel is presented on a separate line.
SENTENCE	Each sentence in the novel is presented on a separate line.

Column Name	Description
WORD_ID	It is the identification number of the word. WORD_ID is in an A-B-C format. The first column A refers to the part of the novel (1, 2, 3, or 4), the second column B indicates the trial number in that part, and the last column C shows the word number in that trial.
SENTENCE_ID	It is the identification number of the sentence. SENTENCE_ID is in an A-B format. The first number A refers to the part of the novel (1, 2, 3, or 4), and the second number B indicates the sentence number in that part.
CHRON_ID	It is the chronological identification number of the current word.
WORD	It is the word presented in the current interest area.
PART_OF_SPEECH	It is the syntactic function of the current word in the sentence context.
CONTENT_WORD	It indicates whether the current word is a content word ("1") or a function word ("0"). If the word is a noun, verb, adjective or adverb, etc., classify it as a content word.
WORD_LENGTH	It is the number of characters of the current word.
IA_AREA	It is the size of the interest area around the current word in pixels.
IA_TOP	It is the top position of the current interest area in pixels.
IA_BOTTOM	It is the bottom position of the current interest area in pixels.
IA_LEFT	It is the left position of the current interest area in pixels.
IA_RIGHT	It is the right position of the current interest area in pixels.
SENTENCE	It is the sentence to which the current SENTENCE_ID refers.
NUMBER_WORDS_ SENTENCE	It is the number of words in the current sentence.

L1ReadingData & L2ReadingData

Description of the ‘L1ReadingData.xlsx’ and ‘L2ReadingData.xlsx’ files. The sheet and column names are in the first column, and the corresponding description of the contents of that sheet or column is presented in the second column.

Sheet Name	Description
DATA	Each interest area is presented on a separate line.
NAME_VARIABLE	**DESCRIPTION**
PP_ID	The identification number of the participant.
GROUP	The group to which the participants belong ("Bilingual" or "Monolingual").
LANGUAGE_RANK	The language of the participants (first language "L1" or second language "L2").
LANGUAGE	The language of the current part of the novel ("Chinese" or "English").
PART	The number of the part of the novel.
TRIAL	The number of the trial.
TRIAL_FIXATION_COUNT	The total number of fixations in the current trial.
TRIAL_DWELL_TIME	The sum of the duration of all fixations in the current trial.
IA_ID	The chronological identification number of the word in the current trial.
WORD_ID	The identification number of the word. The first digit refers to the novel part (1, 2, 3, or 4), the second digit to the trial number, and the last digit to the word number in the trial.
IA_LABEL	The word that displayed in the current interest area.
IA_AVERAGE_FIX_PUPIL_SIZE	The average pupil size of all fixations in the current word.
IA_FIXATION_COUNT	The total number of fixations on the current word.
IA_FIXATION_%	The percentage of the number of fixations on the current word in the trial.
IA_RUN_COUNT	The number of times the current word has been entered and left (runs).
IA_FIRST_RUN_START_TIME	The start time of the first run of fixations on the current word.
IA_FIRST_RUN_END_TIME	The end time of the first run of fixations on the current word.
IA_FIRST_RUN_FIXATION_COUNT	The number of all fixations in a trial falling in the first run of the word.
IA_FIRST_RUN_FIXATION_%	The percentage of all fixations in a trial falling in the first run of the word.
IA_SELECTIVE_REGRESSION_PATH_DURATION	The total fixation durations in the first run on the current word.
IA_SECOND_RUN_START_TIME	The start time of the second run of fixations on the current word.
IA_SECOND_RUN_END_TIME	The end time of the second run of fixations on the current word.
IA_SECOND_RUN_FIXATION_COUNT	The number of all fixations in a trial falling in the second run of the word.
IA_SECOND_RUN_FIXATION_%	The percentage of all fixations in a trial falling in the second run of the word.
IA_SECOND_RUN_DWELL_TIME	The total fixation durations in the second run on the current word.
IA_THIRD_RUN_START_	The start time of the third run of fixations on the current word.

TIME	
IA_THIRD_RUN_END_TIME	The end time of the third run of fixations on the current word.
IA_THIRD_RUN_FIXATION_ %	The percentage of all fixations in a trial falling in the third run of the word.
IA_THIRD_RUN_DWELL_ TIME	The total fixation durations in the third run on the current word.
IA_FIRST_FIXATION_ DURATION	The duration of the first fixation of the current word.
IA_FIRST_FIXATION_INDEX	The ordinal sequence of the first fixation on the current word.
IA_FIRST_FIXATION_RUN_ INDEX	The number of runs of fixations since the first fixation on a word. The current run is included in the tally.
IA_FIRST_FIXATION_TIME	The start time of the first fixation of the current word.
IA_FIRST_FIXATION_VISITE D_IA_COUNT	The number of different words visited before the first fixation of the current word.
IA_FIRST_FIXATION_X	The X position of the first fixation of the current word.
IA_FIRST_FIXATION_Y	The Y position of the first fixation of the current word.
IA_FIRST_FIX_PROGRESSIVE	Indicate whether there is a visit to the following words before the first fixation of the current word. If there is, it is "1". Otherwise it is "0".
IA_SECOND_FIXATION_ DURATION	The duration of the second fixation of the current word, regardless of the run.
IA_SECOND_FIXATION_RUN	The run index of the second fixation of the current word.
IA_SECOND_FIXATION_TIM E	The time of the second fixation of the current word, regardless of run.
IA_SECOND_FIXATION_X	The X position of the second fixation of the current word.
IA_SECOND_FIXATION_Y	The Y position of the second fixation of the current word.
IA_THIRD_FIXATION_ DURATION	The duration of the third fixation of the current word, regardless of the run.
IA_THIRD_FIXATION_RUN	The run index of the third fixation of the current word.
IA_THIRD_FIXATION_TIME	The time of the third fixation of the current word, regardless of run.
IA_THIRD_FIXATION_X	The X position of the third fixation of the current word.
IA_THIRD_FIXATION_Y	The Y position of the third fixation of the current word.
IA_LAST_FIXATION_ DURATION	The duration of the last fixation of the current word, regardless of the run.
IA_LAST_FIXATION_TIME	The time of the last fixation of the current word, regardless of run.
IA_LAST_FIXATION_X	The X position of the last fixation of the current word.
IA_LAST_FIXATION_Y	The Y position of the last fixation of the current word.
IA_REGRESSION_PATH_ DURATION (GO_PAST_TIME)	The sum of all fixation duration since the current word is first fixated until the eyes enter a word with a higher IA_ID.
IA_DWELL_TIME	The sum of the duration of all fixations on the current word.
IA_DWELL_TIME_%	The percentage of trial time spent on the current word.
IA_SPILLOVER	The duration of the first fixation on the next word after leaving the current word in the first pass.
IA_SKIP	If there is no fixation on the word in the first reading, the word is considered to be skipped ("1"). Otherwise ("0").

SubjectInformation

Description of the 'SubjectInformation' file. The sheet and column names are in the first column, and the corresponding description of the contents of that sheet or column is presented in the second column.

Sheet Name	Description
DATA	The file includes information on the language proficiency score, age, and gender of the participants
Column Name	**Description**
PP_ID	It is the identification number of the participant.
AGE	It is the age of the participant in years.
SEX	It is the gender of the participant ("F"=Female, "M"=Male).
EXPOSURE_CN(%)	It is the percentage of exposure to the Chinese language environment currently on average.
EXPOSURE_EN(%)	It is the percentage of exposure to the English language environment currently on average.
LEXTALE_EN	It is the score of the English LexTALE (Lexical Test for Advanced learners of English; Lemhöfer & Broersma, 2012)
SPELLING_EN(%)	It is the percentage of the score on the English spelling test (WRAT4; Dell, Harrold, & Dell, 2008)
LEX_DEC_ACC_EN(%)	It is the percentage of the score on the English lexical decision task, corrected for false positives.
HSK_CN(%)	It is the percentage of the score on HSK (Hanyu Shuiping Kaoshi, level VI, http://www.chinesetest.cn/gosign.do?id=1&lid=0)
COMPR_CN(%)	It is the percentage of the score on multiple-choice questions for the Chinese chapters of the novel.
COMPR_EN(%)	It is the percentage of the score on multiple-choice questions for the English chapters of the novel.

APPENDIX B

Method

Participants, Materials, and procedure

The GECO-CN (Sui et al., 2022) is an eye-tracking corpus in which Chinese-English bilinguals read an entire novel (The Mysterious Affair at Styles or 斯泰尔斯庄园奇案 in Chinese, by Agatha Christie, 1920). A group of 30 native Chinese speakers with an average age of 25.3 years (SD = 2.60) read half of the book in their first language (2 sessions) and the other half in their second language (2 sessions). For more details on the experimental procedure, materials, and participants of the database, we refer the reader to Sui et al., 2022. As the current work aims to document the FEs of Chinese word and character and the potential interaction between them, we only used eye movement data from Chinese reading, more specifically, of the two-character content words in the corpus. The Chinese version of the novel contains 59,403 words, 36331 content words, and 4,835 unique content word types.

Analysis

Fixation durations shorter than 100 ms were not considered to reflect meaningful word processing and were therefore removed from the analysis (e.g., Sereno & Rayner, 2003). Names (e.g., Mary) or words that appear at the beginning or end of a line were excluded from the analysis to avoid potential contamination and wrap-up effects (e.g., Rayner et al., 1989). This left us with 212,135 data points. Reading times that differed by more than 2.5 standard deviations from the subject mean in each measure were excluded from the analysis.

All analyses reported here were performed using R (Version 2021.09.1-372, developed by R Core Team). We ran linear mixed-effects models (LMMs) using the lme4 package (Version 1.1–12). Since this experiment used natural materials without any artificial controls, factors affecting word processing, besides character and word frequency, were also taken into account to avoid them obscuring the results. The model includes word frequency and numbers of word repetitions in Chinese sessions as word-level predictors, frequency and complexity of each character being character-level predictors, and Chinese proficiency of participants.

All predictors are continuous variables and were centred. The word and character frequencies used in this work are Zipf frequencies, a standardized frequency measure independent of corpus size, in SUBTLEX-CH (Cai & Brysbaert, 2010). We obtained the number of strokes of each character from the *Modern Chinese Dictionary 7th Edition* (2016). Random effects of the model were participants and word tokens. The dependent variables were eye movement measures, including FFD, SFD, GD, and TRT. The dependent variable was Box-Cox transformed to normalize the distribution. Such transformation does not change the functional relationship between the dependent and predictor variables.

The model in each reading time measure starts with the same full model. The full model includes interactions between word frequency and C1 and C2 frequency, as well as complexity and language proficiency. Furthermore, the interaction between C1 and C2 frequency and the main effect of word repetition (as a covariate) were included. We first strive to maximize the random factor structure (Barr et al., 2013) and compare the incremental model with the full model using restricted maximum likelihood (REML). If the model fails to converge or is not significantly different from the full model, the previous model will be used. Ultimately, the random factor structure in all reading time measures is unchanged.

We then discover the optimal model by performing stepwise selection to remove non-significant predictor terms. The maximum likelihood is used to select the optimal model, while the REML is used for the final optimal model. In addition, to estimating multicollinearity of coefficients in regression models, the Variance Inflation Factor (VIF) was calculated for each model, using *car* package. A VIF greater than 10 is considered a problem with severe multicollinearity (Fox & Weisberg, 2010; cited from Dirix & Duyck, 2017), whereas a VIF greater than 5 is considered a moderate influence. The largest VIF in the analyses reported below is 1.958, indicating there are no multicollinearity issues (see Table 1).

Table A2 *Tests for Linear Regression*

		Word Frequency			Second Character Frequency	
		Facilitative	**Inhibitory**		**Facilitative**	**Inhibitory**
First Fixation Duration	**Second Character Frequency**	6.1933	7.5754	**Word Frequency**	2.2461	4.2348
	Chisq	3.8458	2.6837	**Chisq**	3.8474	3.8658
	Pr(>Chisq)	0.0499	0.1014	**Pr(>Chisq)**	0.0498	0.0493

		Word Frequency			Second Character Frequency	
		Facilitative	**Inhibitory**		**Facilitative**	**Inhibitory**
Single Fixation Duration	**Second Character Frequency**	6.4607	7.5754	**Word Frequency**	1.4744	4.2554
	Chisq	3.8535	0.6643	**Chisq**	3.6883	3.8439
	Pr(>Chisq)	0.0496	0.4150	**Pr(>Chisq)**	0.0548	0.0499

		Word Frequency			Second Character Frequency	
		Facilitative	**Inhibitory**		**Facilitative**	**Inhibitory**
Gaze Duration	**Second Character Frequency**	6.2837	7.4507	**Word Frequency**	2.7251	4.0753
	Chisq	3.8595	3.8665	**Chisq**	3.8445	3.8448
	Pr(>Chisq)	0.0495	0.0493	**Pr(>Chisq)**	0.0499	0.0499

		Word Frequency			Second Character Frequency	
		Facilitative	**Inhibitory**		**Facilitative**	**Inhibitory**
Total Reading Time	**Second Character Frequency**	6.2387	7.3107	**Word Frequency**	2.5883	4.0234
	Chisq	3.8604	3.8477	**Chisq**	3.8610	3.8570
	Pr(>Chisq)	0.0494	0.0498	**Pr(>Chisq)**	0.0494	0.0495

*p < .05.

APPENDIX C

Chinese-English Bilingual: L1 vs L2

First Fixation Duration

The FFD was significantly shorter in L1 than in L2 (see Table A.1). The main effect of the frequency in L1 was significant, while word length was not, and both were statistically smaller than those in L2. HF or shorter words are processed faster than LF or longer words. The language proficiency effect was significant in L1. FFD increased with L1 language proficiency. The FE increased with word length, and the effect increased more in L2 than in L1 (see Fig.A.1). Word repetition effects were observed in L1 and L2, positively correlated with FFD. The effect was not significantly different between the two languages.

Gaze Duration

Readers spend less time reading in L1 than in L2 (see Table A.1). Frequency, word length, and language proficiency had significant main effects in both languages, with larger effects in L2. GDs were negatively correlated with frequency and positively correlated with word length in L1 and L2. They were positively and negatively correlated with proficiency in L1. Frequency and word length exhibited significant interactions, with FE increasing with word length and the increased effect being larger in L2 than in L1 (see Fig.A.1). Frequency also interacted with language proficiency, with FE decreased as proficiency increased. Note that GDs of highly proficient readers were shorter than those of low proficient. The repetition effect was significant, with no major differences between the two languages. The higher the number of word occurrences, the longer the GDs.

Total Reading Time

The TRTs in L2 were longer than those in L1 (see Table A.1). Significant frequency and word length effects were found in L1 and were statistically smaller than in L2. The interaction between frequency and word length was insignificant in L1, statistically different from that in L2 (see Fig.A.1). FE increased with word length in L2. The repetition effect was significant, showing a positive correlation with TRT and no differences between languages.

Skipping probability

The skipping probability was higher in L1 than in L2 (see Table A.1). The main effects of frequency and word length were significant in L1, with statistically smaller frequency and larger word length effects than those in L2. The higher the frequency, or the shorter the word, the higher the probability of skipping it. The proficiency effect was significant. Readers with higher proficiency have a higher skipping probability. The interactions between frequency and word length and between frequency and language proficiency in L1 were insignificant, different from those in L2 (see Fig.A.1). The results showed that the reverse FE increased with decreasing word length and decreased with increasing L2 proficiency. Repetition effects were significant, with no difference between L1 and L2. The more repetitions, the lower the probability of skipping.

Table A.1
Comparative Analyses of Fixation Duration Measures and Skipping Probabilities between First and Second Language Reading in Chinese-English Bilinguals

	Predictors	Estimate	Std. Error	*t value*	*Pr(>|t|)*		VIF
First Fixation Duration	**(Intercept)**	**5.459000**	**0.019210**	**284.149**	**< 0.001**	***	
	LanguageEN	**0.118200**	**0.002103**	**56.216**	**< 0.001**	***	3.78
	Frequency	**-0.003281**	**0.000979**	**-3.353**	**0.001**	***	4.83
	Word Length	-0.020070	0.011530	-1.741	0.082	.	5.37
	Proficiency	**-0.001055**	**0.000083**	**-12.651**	**< 0.001**	***	3.36
	Repetition	**0.000022**	**0.000011**	**1.98**	**0.048**	*	2.48
	LanguageEN:Frequency	**-0.020640**	**0.001341**	**-15.392**	**< 0.001**	***	5.34
	LanguageEN:Word Length	**0.269900**	**0.014950**	**18.055**	**< 0.001**	***	6.19
	Frequency:Word Length	**-0.016420**	**0.006914**	**-2.375**	**0.018**	*	4.58
	LanguageEN:Repetition	0.000028	0.000019	1.499	0.134		2.43
	LanguageEN:Frequency:Word Length	**0.080200**	**0.008580**	**9.348**	**< 0.001**	***	5.75

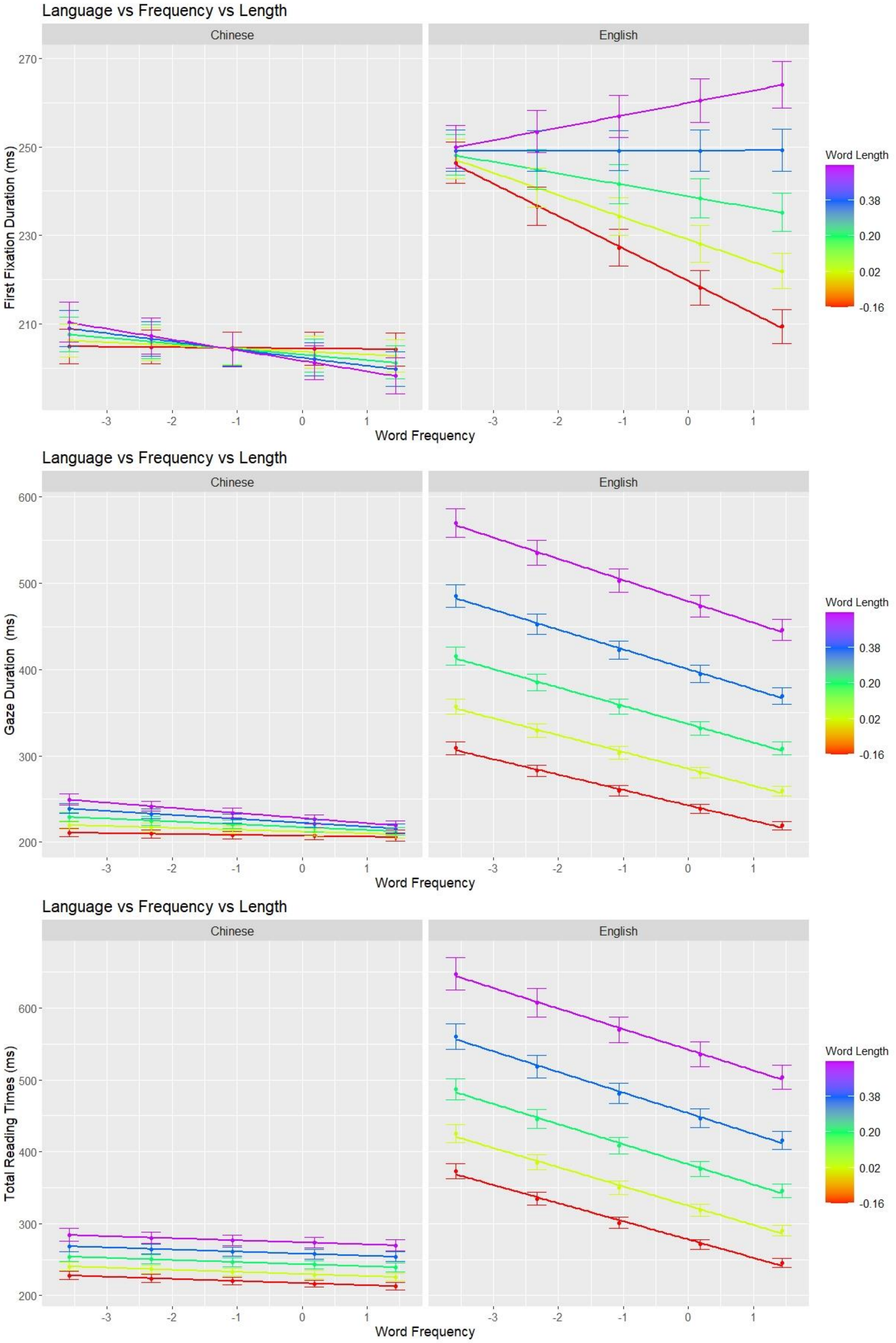
	Predictors	Estimate	Std. Error	t value	Pr(>|t|)		VIF
Gaze Duration	**(Intercept)**	**3.198000**	**0.006873**	**465.367**	**< 0.001**	***	
	LanguageEN	**0.086080**	**0.000757**	**113.779**	**< 0.001**	***	3.49
	Frequency	**-0.002697**	**0.000380**	***-7.098***	**< 0.001**	***	4.85
	Word Length	**0.041980**	**0.004338**	**9.678**	**< 0.001**	***	4.98
	Proficiency	**-0.000561**	**0.000030**	**-18.96**	**< 0.001**	***	3.13
	Repetition	**0.000013**	**0.000003**	**3.825**	**< 0.001**	***	1.63
	LanguageEN:Frequency	**-0.016180**	**0.000557**	**-29.06**	**< 0.001**	***	5.88
	LanguageEN:Word Length	**0.235300**	**0.005656**	**41.61**	**< 0.001**	***	5.61
	Frequency:Word Length	**-0.008898**	**0.002589**	**-3.436**	**0.001**	***	4.01
	Frequency:Proficiency	**0.000095**	**0.000015**	**6.394**	**< 0.001**	***	1.77
	LanguageEN:Frequency:Word Length	**0.019960**	**0.003113**	**6.41**	**< 0.001**	***	4.65

	Predictors	Estimate	Std. Error	t value	Pr(>|t|)		VIF
Total Reading Time	**(Intercept)**	**2.954000**	**0.006083**	**485.58**	**< 0.001**	***	
	LanguageEN	**0.080070**	**0.000699**	**114.535**	**< 0.001**	***	3.38
	Frequency	**-0.002853**	**0.000357**	**-7.982000**	**< 0.001**	***	4.73
	Word Length	**0.080990**	**0.004093**	***19.79***	**< 0.001**	***	4.87
	Proficiency	**-0.000490**	**0.000027**	**-18.182**	**< 0.001**	***	3.00
	Repetition	**0.000018**	**0.000003**	**5.64**	**< 0.001**	***	1.63
	LanguageEN:Frequency	**-0.014310**	**0.000524**	**-27.309**	**< 0.001**	***	5.65
	LanguageEN:Word Length	**0.128300**	**0.005379**	**23.86**	**< 0.001**	***	5.52
	Frequency:Word Length	0.001353	0.002458	0.55	0.58		3.96
	Frequency:Proficiency	**0.000111**	**0.000013**	**8.22**	**< 0.001**	***	1.70
	LanguageEN:Frequency:Word Length	**0.012070**	**0.002965**	**4.071**	**< 0.001**	***	4.62

	Predictors	Estimate	Std. Error	z value	Pr(>|z|)		VIF
Skipping probability	(Intercept)	0.127500	0.072980	1.746	0.081	.	
	LanguageEN	**-2.370000**	**0.014380**	**-164.84**	**< 0.001**	***	3.66
	Frequency	**0.038830**	**0.006200**	**6.264**	**< 0.001**	***	3.83
	Word Length	**-5.517000**	**0.062480**	**-88.295**	**< 0.001**	***	2.42
	Proficiency	**0.014550**	**0.000564**	**25.801**	**< 0.001**	***	3.05
	Repetition	**-0.000384**	**0.000050**	**-7.702**	**< 0.001**	***	1.60
	LanguageEN:Frequency	**0.081230**	**0.012300**	**6.604**	**< 0.001**	***	4.44
	LanguageEN:Word Length	**-3.439000**	**0.125400**	**-27.416**	**< 0.001**	***	3.49
	Frequency:Word Length	-0.077410	0.042200	-1.834	0.067	.	1.82
	Frequency:Proficiency	**-0.002312**	**0.000395**	**-5.857**	**< 0.001**	***	2.39
	LanguageEN:Frequency:Word Length	**-1.70E+00**	**0.06638**	**-25.545**	**< 0.001**	***	**2.2**

Estimate = Estimates; SE = standard errors; *t* = t-values; ***p*** = *p*-values (calculated using lmerTest package); VIF = variance inflation factor.

Figure A.1. Three-way Interaction Plots between the Language (L1 and L2), Word Length, and Frequency in Chinese-English Bilinguals.



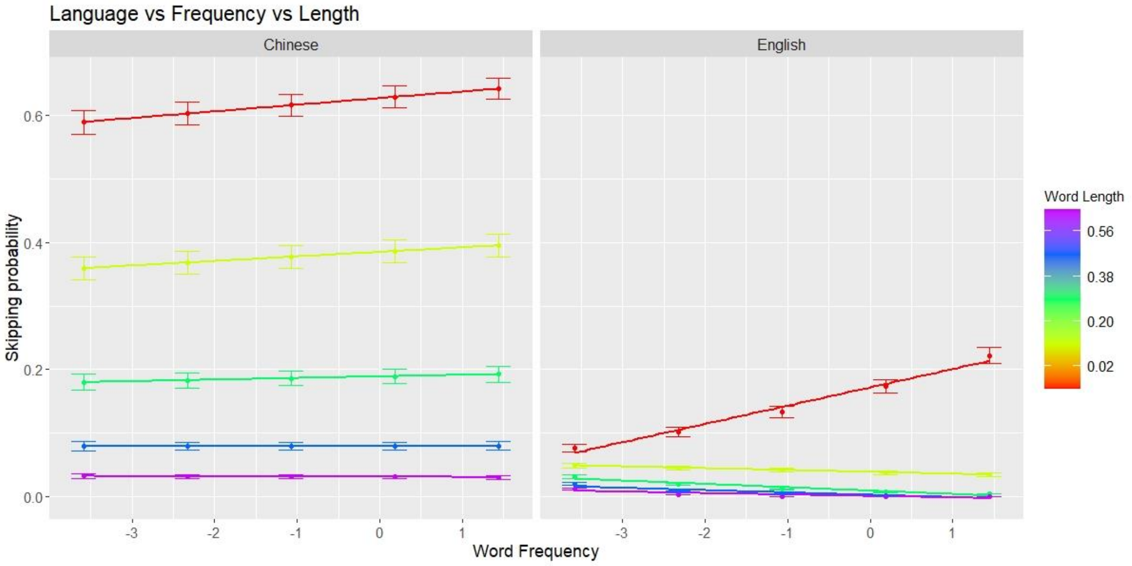


Figure A.2 The effects of Group, Word Length, and Frequency on Fixation Count (Numbers) in Second Language Reading.

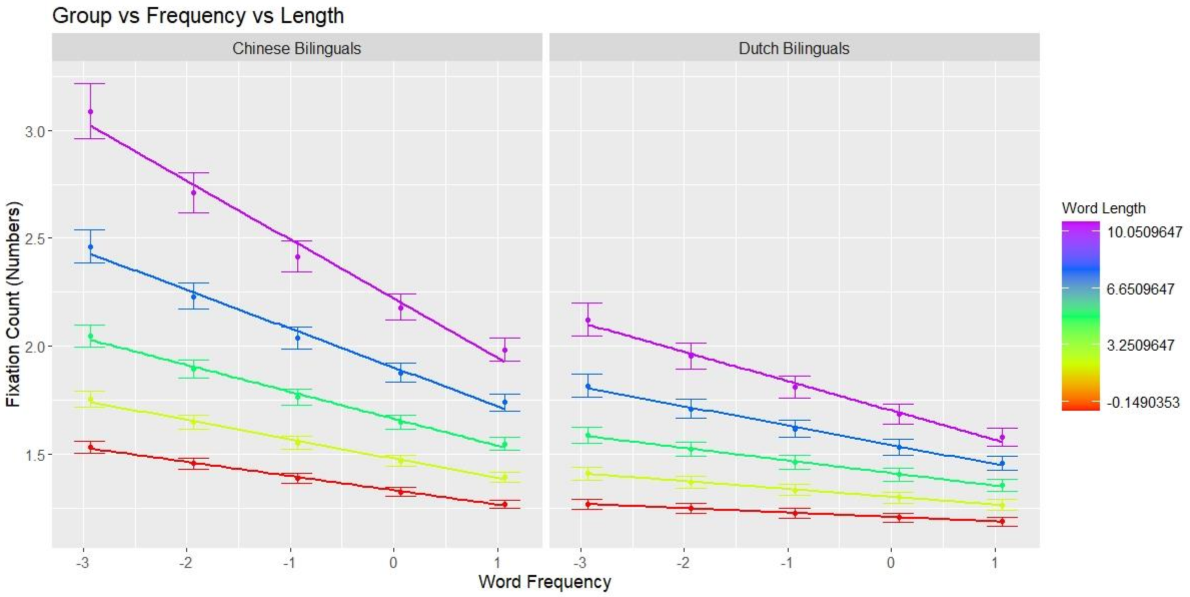


Figure A.3 Histograms of the word frequency distribution of the Chinese and Dutch data analyzed in this work.

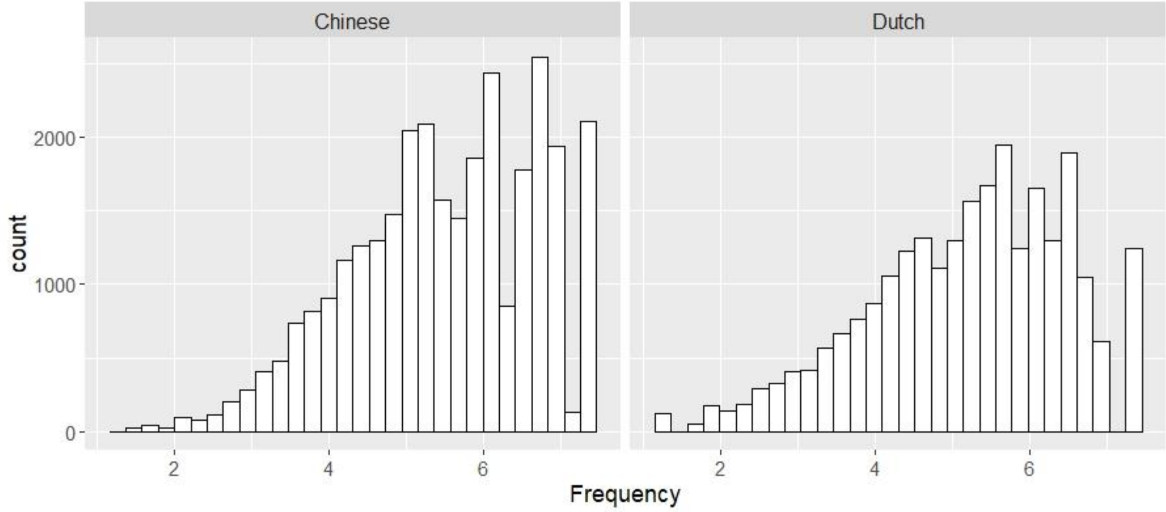
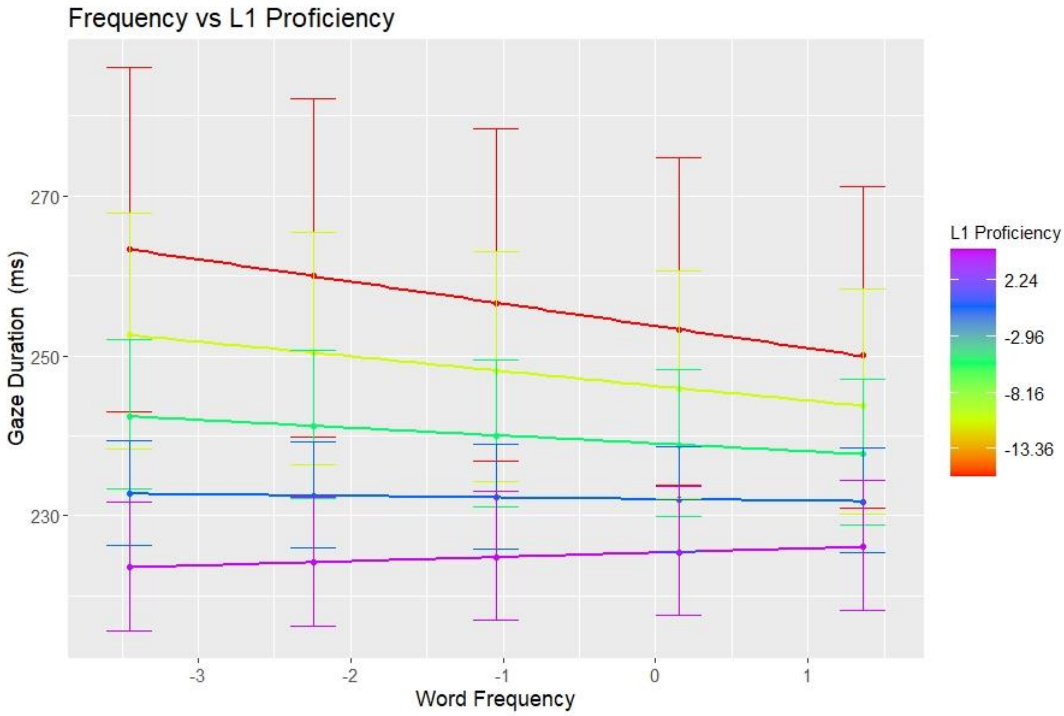


Figure A.4 The effects of First language proficiency and Frequency on Gaze Duration (millisecond) in Second Language Reading.



DATA STORAGE FACT SHEETS

DATA STORAGE FACT SHEET CHAPTER 2

Author: Longjiao Sui
Date: 12-05-2023

1. Contact details
==
1a. Main researcher

- name: Longjiao Sui
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: longjiao.sui@ugent.be; longjiao.sui@gmail.com

1b. Responsible Staff Member (ZAP)

- name: Nicolas Dirix
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: nicolas.dirix@ugent.be

- name: Wouter Duyck
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: wouter.duyck@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==
* Reference of the publication in which the datasets are reported:
a) Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods*, 1-21.

b) Sui, L. (2023). An Eye-tracking Corpus study of Chinese-English bilingual reading (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:
All datasets reported in Chapter 2 of the doctoral dissertation.

3. Information about the files that have been stored
==
3a. Raw data
----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [X] research group file server
  - [X] other (specify): researcher external hard drive AND computers in the Eye-Tracking Lab

* Who has direct access to the raw data (i.e., without intervention of another person)?
  - [X] main researcher
  - [X] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

3b. Other files
----
* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to reported results. Specify: ...
  - [X] file(s) containing processed data. Specify: Excel data files
  - [ ] file(s) containing analyses. Specify: ...
  - [ ] files(s) containing information about informed consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [X] research group file server
  - [X] other (specify): researcher external hard drive AND OSF

* Who has direct access to these other files (i.e., without intervention of another person)?
  - [X] main researcher
  - [X] responsible ZAP
  - [X] all members of the research group
  - [X] all members of UGent
  - [X] other (specify): The Excel data files are publicly available with the publication of Chapter 2 and can be accessed via the OSF link provided in the article.

4. Reproduction
===
* Have the results been reproduced independently?: [ ] YES / [X] NO
* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:

DATA STORAGE FACT SHEET CHAPTER 3

Author: Longjiao Sui
Date: 12-05-2023

1. Contact details
==
1a. Main researcher

- name: Longjiao Sui
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: longjiao.sui@ugent.be; longjiao.sui@gmail.com

1b. Responsible Staff Member (ZAP)

- name: Nicolas Dirix
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: nicolas.dirix@ugent.be

- name: Wouter Duyck
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: wouter.duyck@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==
* Reference of the publication in which the datasets are reported:
a) Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods*, 1-21.

b) Sui, L. (2023). An Eye-tracking Corpus study of Chinese-English bilingual reading (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:
All datasets reported in Chapter 3 of the doctoral dissertation.

3. Information about the files that have been stored
==
3a. Raw data
----
* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [X] research group file server
  - [X] other (specify): researcher external hard drive AND computers in the Eye-Tracking Lab

* Who has direct access to the raw data (i.e., without intervention of another person)?
  - [X] main researcher
  - [X] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

3b. Other files
----
* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to reported results. Specify: ...
  - [X] file(s) containing processed data. Specify: Excel data files
  - [ ] file(s) containing analyses. Specify: ...
  - [ ] files(s) containing information about informed consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [X] research group file server
  - [X] other (specify): researcher external hard drive AND OSF

* Who has direct access to these other files (i.e., without intervention of another person)?
  - [X] main researcher
  - [X] responsible ZAP
  - [X] all members of the research group
  - [X] all members of UGent
  - [X] other (specify): The Excel data files are publicly available with the publication of Chapter 2 and can be accessed via the OSF link provided in the article.

4. Reproduction
===
* Have the results been reproduced independently?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:

DATA STORAGE FACT SHEET CHAPTER 4

Author: Longjiao Sui
Date: 12-05-2023

1. Contact details
===
1a. Main researcher

- name: Longjiao Sui
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: longjiao.sui@ugent.be; longjiao.sui@gmail.com

1b. Responsible Staff Member (ZAP)

- name: Nicolas Dirix
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: nicolas.dirix@ugent.be

- name: Wouter Duyck
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: wouter.duyck@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
===
* Reference of the publication in which the datasets are reported:
a) Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods*, 1-21.

b) Sui, L. (2023). An Eye-tracking Corpus study of Chinese-English bilingual reading (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:
All datasets reported in Chapter 4 of the doctoral dissertation.

3. Information about the files that have been stored
===
3a. Raw data
----
* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [X] research group file server
  - [X] other (specify): researcher external hard drive AND computers in the Eye-Tracking Lab

* Who has direct access to the raw data (i.e., without intervention of another person)?
  - [X] main researcher
  - [X] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

3b. Other files
----
* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to reported results. Specify: ...
  - [X] file(s) containing processed data. Specify: Excel data files
  - [ ] file(s) containing analyses. Specify: ...
  - [ ] files(s) containing information about informed consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [X] research group file server
  - [X] other (specify): researcher external hard drive AND OSF

* Who has direct access to these other files (i.e., without intervention of another person)?
  - [X] main researcher
  - [X] responsible ZAP
  - [X] all members of the research group
  - [X] all members of UGent
  - [X] other (specify): The Excel data files are publicly available with the publication of Chapter 2 and can be accessed via the OSF link provided in the article.

4. Reproduction
===
* Have the results been reproduced independently?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail: