



GECO-CN: Ghent Eye-tracking CORpus of sentence reading for Chinese-English bilinguals

Longjiao Sui¹ · Nicolas Dirix¹ · Evy Woumans² · Wouter Duyck^{1,3}

Accepted: 10 July 2022 / Published online: 27 July 2022
© The Psychonomic Society, Inc. 2022

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 for use by researchers interested in (bilingual) reading.

Keywords Bilingualism · Reading · Eye-tracking · Corpus

Reading is an essential skill for distilling content from written text. Psycholinguistic researchers have striven 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 the 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 for explaining 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 et al., 2007a; 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; Liveredge et al., 2014; Zola, 1984; Balota et al., 1985; Rayner & Well, 1996; Rayner et al., 2005, 2006). 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

✉ Longjiao Sui
longjiao.sui@ugent.be; longjiao.sui@gmail.com

¹ Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

² Department of Translation, Interpreting and Communication, Ghent University, Ghent, Belgium

³ The Accreditation Organisation of the Netherlands and Flanders (NVAO), Den Haag, Netherlands

who 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 whether there are any potential adverse impacts. Understanding these questions is 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 on 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. The Chinese character, however, is a type of string formed by a number of strokes. 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 the use of a limited number of characters (approximately 5000 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 frequently 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 the twentieth century, 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 is 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 were faster if the target character could compose a word with the adjacent legible character than when it could not (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 contain 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 versus longer words, with longer fixations and fewer skips for longer words (Zang et al., 2018). Additionally, Rayner et al. (2007a) claim that eye-movement control in reading appears to be fundamentally similar between Chinese and English readers (see also 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 lead 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 language 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 et al., 2007b) and 7.5 letters in English reading (Rayner et al., 2007b).

To sum up this section, 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-scale 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 also influence 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 the work of Li et al. (2014) failed to find. 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-scale 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 provides the possibility for 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 fixation 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 us to examine existing hypotheses and models, investigate multiple main effects and interactions of factors involved in reading, and evaluate 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 et al. (2013), who had 43 participants read 205 British English sentences with a total of 1931 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 et al. (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 2689 tokens and 1197 unique word types in total.

Corpora not including English include the German-language Potsdam Sentence Corpus (Kliegl et al., 2006) with data for 1138 tokens read by 222 readers, the Dutch Eye-Movements Online Internet Corpus (DEMONIC; Kuperman et al., 2010) containing data for 55 participants reading 1746 tokens, the Russian Sentence Corpus (RSC, Laurinavichyute et al., 2019) with 96 Russian monolinguals and reading a total of 1362 tokens, and the Beijing Sentence Corpus (Pan et al., 2021) with eye-tracking data for 60 Chinese native participants reading sentences from newspapers, totalling in 936 types and 1685 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 on second-language (L2) reading. The first work to introduce a corpus of bilingual reading is the Ghent Eye-Tracking Corpus (GECO, Cop et al., 2017a), 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 on 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, and 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 spoke 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-scale 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 for words across various language contexts, in this case the specific texts, as Dirix et al. (2019) have shown when comparing the two databases (GECO, Cop et al., 2017a; 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 et al., 2017a) 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 et al., 2015a). 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 et al., 2015b). 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 to be 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 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 et al., 2017b). 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 et al., 2017a), 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 a 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 word boundaries due to the lack of visual clues 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 significant challenge. However, it must be confronted because it is an essential part of the actual performance of Chinese readers, which is the first reason that it is important to have an eye-tracking corpus that simulates real-world reading for non-Western languages as well. 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 is semantically interconnected. 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 et al. (2013) and Dirix et al. (2019).

Unlike 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 should take other coordination processes into account (Kuperman 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 5000 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 for investigation and comparison of diverse aspects involved in reading, and the ability to examine 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 et al., 2007a) and Chinese reading model (CRM), Li & Pollatsek, 2020) in predicting eye movements when reading Chinese in paragraphs using data on 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 to investigate 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 for the construction and development of both Chinese and universal reading models.

Furthermore, this eye-tracking corpus allows one to examine 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 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 enables 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 et al., 2017a), 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.¹ The remaining 30 participants (8 male) with an average age of 25.3 years (range: 20–29; SD = 2.60) were master's 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 (Language Experience and Proficiency Questionnaire, Marian et al., 2007) to investigate their language background; the HSK test (Chinese Proficiency Test, n.d., level 6) to explore Chinese proficiency; and three tasks to objectively assess English proficiency (in accordance with the first GECO, Cop et al., 2017a, 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 un-speeded lexical decision task; WRAT4 (Wide Range Achievement Test 4, Wilkinson & Robertson, 2006), a spelling task; and a classic speeded lexical decision task (see details in Table 1).

Based on the LexTALE classification, 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.

Materials

Following the research of Cop et al. (2017a), 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 et al., 2017a).

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 Chapters 1–7 in L1 (Chinese) and Chapters 8–13 in L2 (English), while the other 15 participants read in

Footnote 1 (continued)

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.

¹ 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

Table 1 Descriptive statistics of subject information

	Chinese bilinguals		Dutch bilinguals		<i>t</i> Value L1–L2 (Chinese-English)	<i>t</i> Value	
	L1 (Chinese) M [SD]	L2 (English) M [SD]	L1 (Dutch) M [SD]	L2 (English) M [SD]		L1–L1 (Chinese-Dutch)	L2–L2 (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%.

Table 2 Descriptive statistics of the Chinese and English versions of the novel *The Mysterious Affair at Styles*

	Chinese			English		
Number of words	59,403			56,841 ^a		
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]

^aNote that there are some differences with GECO (Cop et al., 2017a) in the reported number of words and nouns. In contrast to Cop et al. (2017a), 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)

*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 for the two languages is the same.

reverse language order. Chinese text was displayed in simplified form. The Chinese version of the novel has 59,403 words with 5053 unique types, and the English version has 56,841 words with 5363 unique types. More detailed information of the novel is presented in Table 2.

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 et al., 2017a). 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 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 on 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 chapters 1–4 during the first session, chapters 5–7 during the second session, chapters 8–10 during the third session, and chapters 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 read in silence and to try not to move their heads while reading, 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 nine-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 regularly approximately every 10–15 minutes. 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 space bar 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 0.5°; 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 the same as 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 since 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 means 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,² 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 the *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 the *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

² 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).

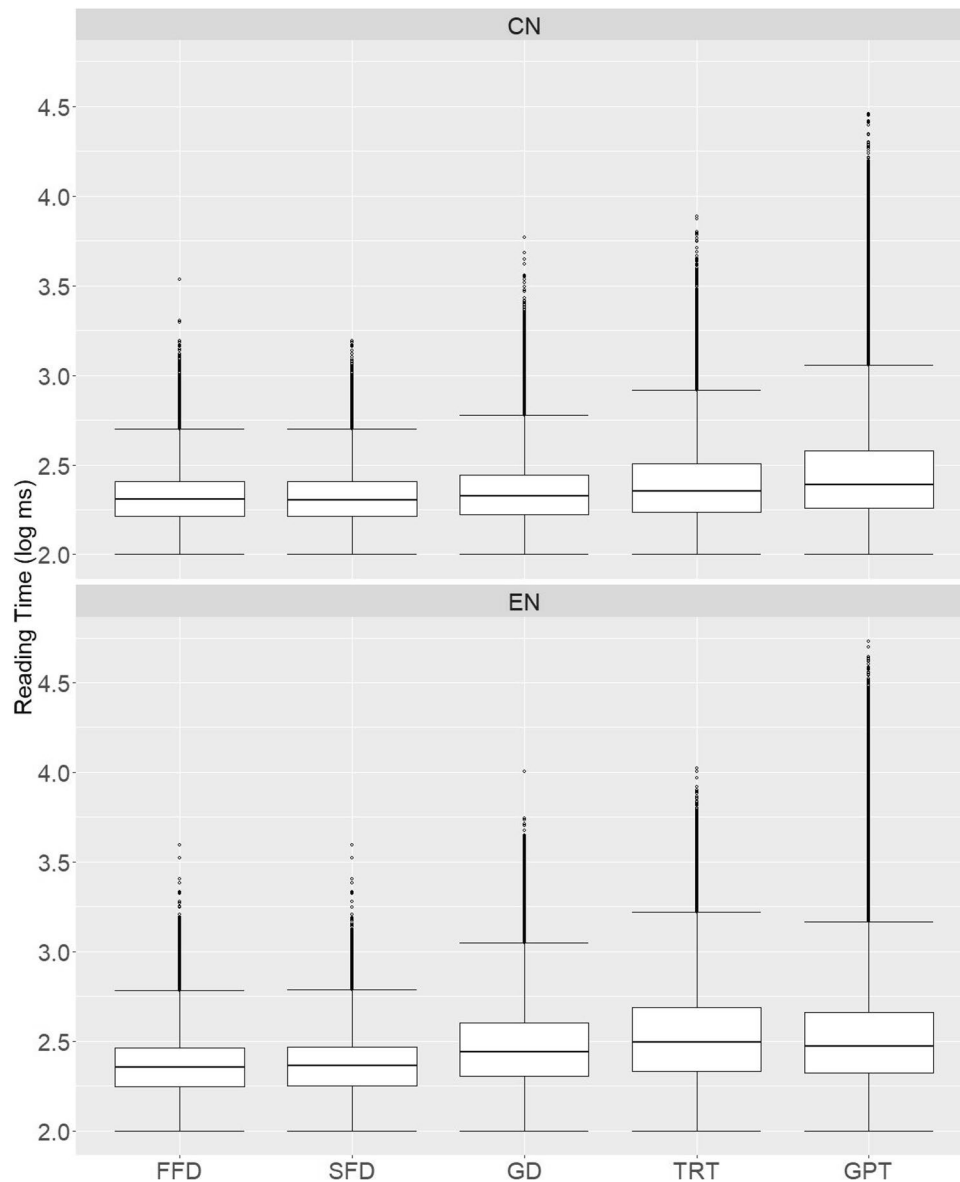


Fig. 1 Boxplots of reading times. Boxplots present the reading times of the first and second language of Chinese-English bilinguals. The reading times were log-transformed (y-axis, in milliseconds). The

upper plot presents the reading times of the L1 (Chinese), while the lower plot indicates the reading times of the second language (English).

(left-side) words from the time 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 Fig. 1. This paper also makes all data freely available online (access link: 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 otherwise noted.

Reading time 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 Fig. 1). A large number of positive outliers were found,

which is consistent with GECO (Cop et al., 2017a). 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 et al., 2017a). 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),³ 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 (see also 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 most participants in this work were perhaps more proficient readers than those in previous studies. Most likely, a combination of the factors mentioned here contributed to the high Chinese reading rate in this corpus.

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 et al., 2015a; 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. 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 et al. (2017a). The statistical values of the Lilliefors and skewness tests are presented in the corresponding panels of Fig. 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 was true for all measures. The average first fixation duration was 24.18 ms longer in L2 than in L1, and the difference in average total reading time reached 113.01 ms. This is consistent with previous work showing 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 et al., 2015a). 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 durations in L1 were highly correlated with fixation durations in L2 (>.80 for all measures). This suggests that the reading speed in L1 is predictive of the reading speed in L2.

The 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 with Dutch bilinguals and English monolinguals (Cop et al., 2017a), 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 the duration of fixation on a Chinese word by Chinese readers was comparable to that in previous studies on reading paragraphs (Liversedge et al., 2016; Feng et al., 2009) but somewhat faster than in studies 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

³ 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).

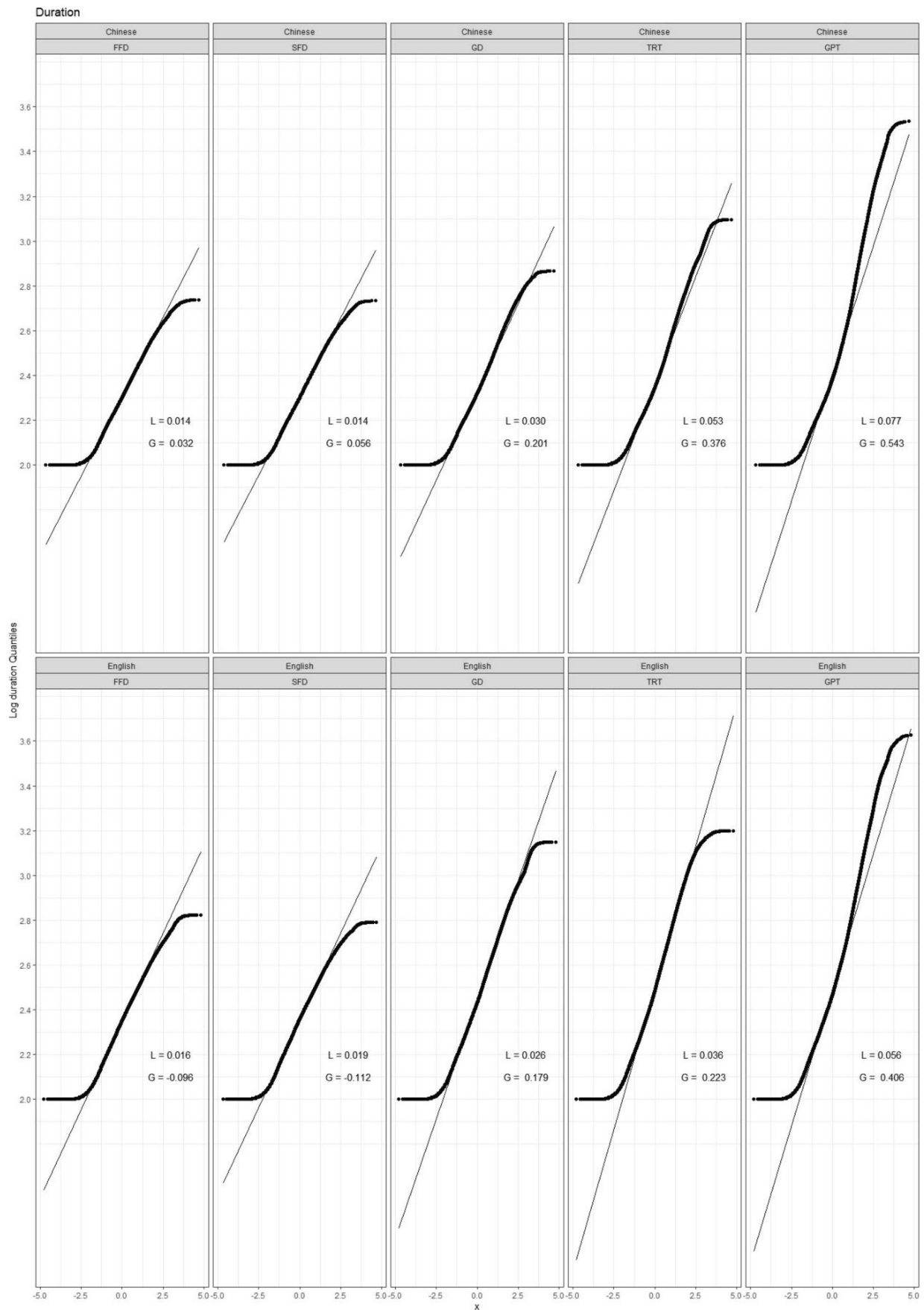


Fig. 2 Quantile–quantile plots of reading times. Quantile–quantile plots present the reading times of the five measures in each language condition (Chinese in the upper figures and English in the lower figures). Reading times were trimmed and log-transformed. The statistical values of the Lilliefors 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.

when processing a word because, for instance, words could be more predictable in the former condition.

Interindividual consistency in reading times

In order to test the reliability of the dataset, split-half correlations with a Spearman–Brown correction were conducted using the R 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 in the reading times was quite high and similar to those displayed in GECO (Cop et al., 2017a), 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.

The unbalanced consistency in reading times is likely an effect of lower 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).

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 et al., 2015b; 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

Table 3 Descriptive statistics of five reading time measures

	Chinese bilinguals								Dutch bilinguals		
	Bilingual L1			Bilingual L2			<i>t</i> Value	Correlation	Bilingual L2		
	(Chinese)			(English)			(L1–L2)	(L1–L2)	(English)		
	M	SD	Range	M	SD	Range	<i>t</i> [DF]	<i>r</i>	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 et al., 2017a). 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.

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

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 Fig. 5 presents the effect of word length on skipping probability.

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 et al., 2017a), 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 higher, 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 et al., 2017a) 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 (偵探 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.

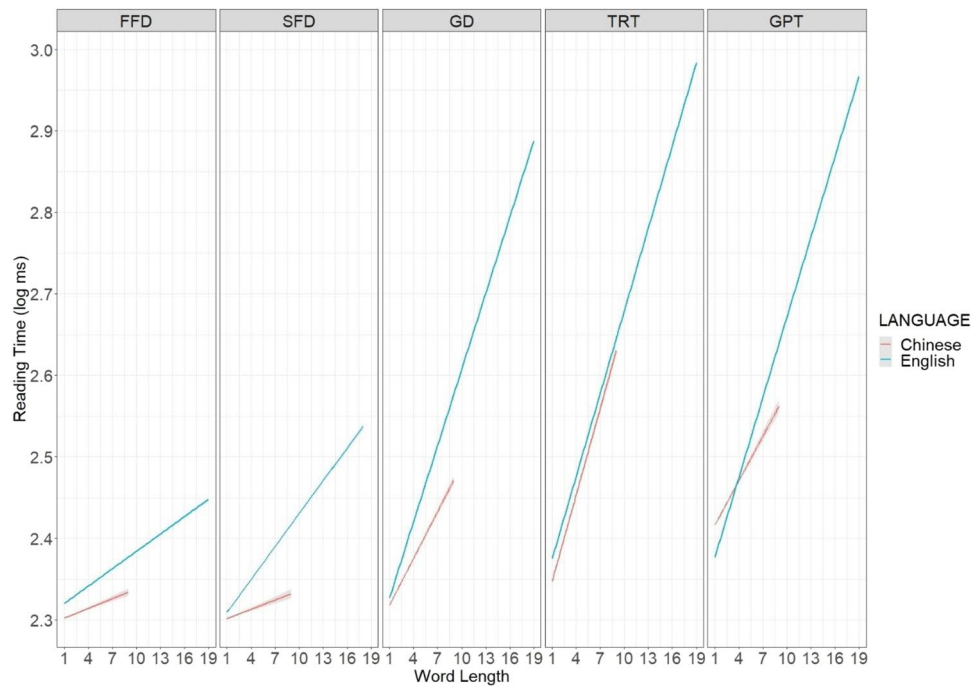


Fig. 3 Plots of the word length effect on reading times. It shows the word length effect on the reading times of the five measures when reading in the first (Chinese, red line) and second (English, blue line) languages. The grey shadow is the confidence interval.

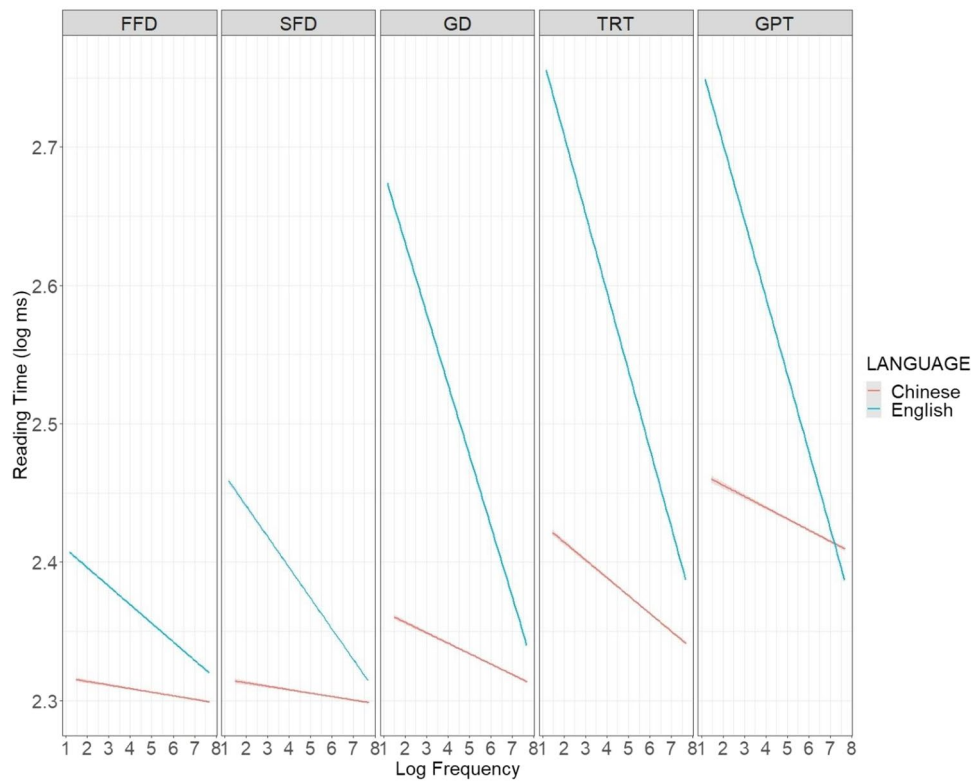


Fig. 4 Plots of the word frequency effect on reading times. It shows the effect of word frequency on the five measurements of reading times in the first (Chinese, red line) and second (English, blue line) languages reading. The grey shadow is the confidence interval.

Table 5 Skipping probabilities

	Bilingual L1			Bilingual L2		
	(Chinese)			(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.

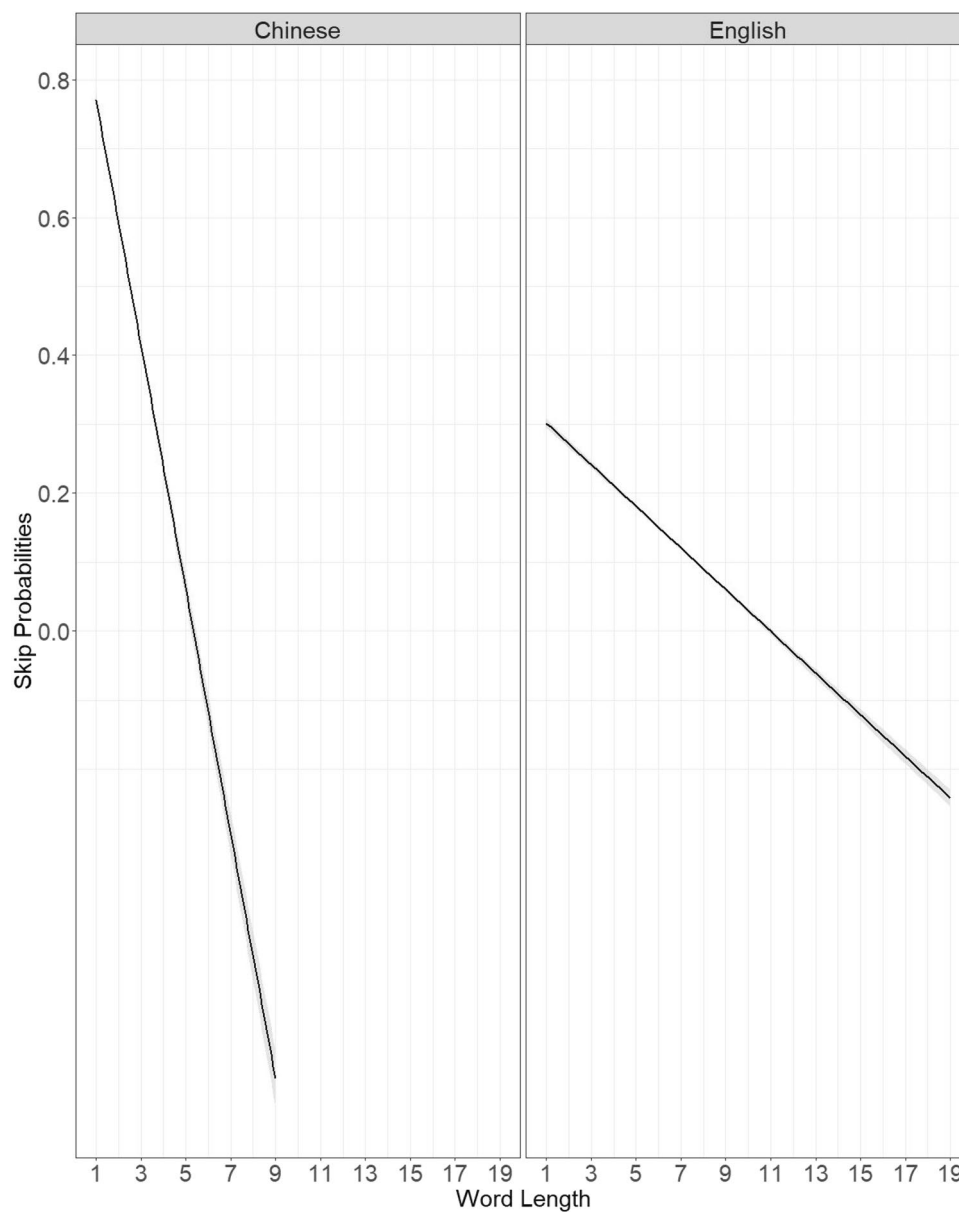


Fig. 5 Plots of the word length effect on the skipping probability. It shows the effect of word length on the skipping probability when reading in the first (Chinese, left side) and second (English, right side) languages. The grey shadow is the confidence interval.

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 et al., 2017a), 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 the 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 us to examine the diversity and generalisation of these models. Following on the success of the first GECO⁴ (Cop et al., 2017a), 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 et al., 2017a; Liversedge et al., 2016). Unlike the original GECO, where the reading time consistency between the two languages was similar, this work showed somewhat lower reading time consistency in L1, albeit still very high. The difference between the two language groups may be due to the disparity 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 et al., 2015a, 2017a). In addition, Chinese bilinguals also exhibited larger frequency effects in L2, similar to what was observed in Dutch-English bilinguals (Cop et al., 2015b).

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 corpora. 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 et al., 2013).

Another notable finding from this work is the influence of L1 on L2 processing. This work, along with previous GECO (Cop et al., 2017a) 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 equally 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

⁴ 152 Google Scholar citations in the first five years of publication, measured 19 May 2022.

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 (a 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 enormous 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 a 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 et al., 2007a; CRM, Li & Pollatsek, 2020) related to reading should be able to explain the abundance of phenomena encountered in natural reading.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-022-01931-3>.

Author Note We would like to thank Prof. Chu Zexiang and Prof. Su Xinchun for their valuable suggestions on word segmentation and two anonymous reviewers for their valuable input on earlier versions of this paper.

Open Practices Statement

GECO-CN's data and materials are made available at the project's OSF page—see “Results and discussion” section above and Appendix A for details of supplementary materials.

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. <https://doi.org/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., & 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.
- 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). Praeger.
- Chen, H. C., Vaid, J., & Wu, J. T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language & Cognitive Processes*, 24(7-8), 967–982.
- Chinese Proficiency Test. (n.d.) *HSK Level VI*. <http://www.chinesetest.cn/gosign.do?id=1&lid=0#>
- 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., Drieghe, D., & Duyck, W. (2015a). 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. (2015b). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, 22(5), 1216–1234.
- 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.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109, 193–210. <https://doi.org/10.1016/j.cognition.2008.07.008>
- Dictionary Editorial Office of Linguistic Research Institute, Chinese Academy of Social Sciences. (2016). *Modern Chinese Dictionary* (7th ed.). 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., 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.
- 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.
- 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 uns-paced 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 ed.). The Commercial Press.
- 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., Rotsztein, 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*.
- 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).
- 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. <https://doi.org/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*. The Commercial Press.
- 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., & Su, X. (2022). *Lexicon of common words in contemporary Chinese Second Edition*. The Commercial Press.
- Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology*, 58(4), 525–552.
- 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., 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., 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). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199324576.013.14>

- 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.
- Liu, Y. H. (2019). *Practical Modern Chinese Grammar Third Edition*. The Commercial Press.
- 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. <https://doi.org/10.1371/journal.pone.0055440>
- 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.
- 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.
- 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.
- Pynthe, 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., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, 27(6), 537–544.
- 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., 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., Li, X., & Pollatsek, A. (2007a). Extending the E-Z reader model of eye movement control to Chinese readers. *Cognitive Science*, 31(6), 1021–1033.
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007b). 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 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. <https://doi.org/10.1016/j.bandl.2011.03.001>
- Wilkinson, G. S., & Robertson, G. (2006). *Wide Range Achievement Test 4 professional manual*. Psychological Assessment Resources.
- Woolley, G. (2011). Reading comprehension. In *Reading comprehension* (pp. 15–34). Springer.
- 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). 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.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.