Bilingualism: Language and Cognition

cambridge.org/bil

The word frequency effect in first- and second-language reading by Chinese and Dutch bilinguals

Longjiao Sui[1], Evy Woumans[2], Wouter Duyck[3,4] and Nicolas Dirix[3]

[1]School of Foreign Languages, Dalian Maritime University, Dalian, China; [2]Department of Translation, Interpreting and Communication, Ghent University, Ghent, Belgium; [3]Department of Experimental Psychology, Ghent University, Ghent, Belgium and [4]The Accreditation Organisation of the Netherlands and Flanders (NVAO), Den Haag, Netherlands

Research Article

Cite this article: Sui, L., Woumans, E., Duyck, W., & Dirix, N. (2025). The word frequency effect in first- and second-language reading by Chinese and Dutch bilinguals. *Bilingualism: Language and Cognition* **28**, 232–247. https://doi.org/10.1017/S136672892400035X

Received: 2 June 2023
Revised: 18 March 2024
Accepted: 28 March 2024
First published online: 3 May 2024

Keywords:
frequency effects; reading; bilingualism; eye-movements; cross-linguistic comparison

Corresponding author:
Longjiao Sui;
Email: longjiao.sui@dlmu.edu.cn, longjiao.sui@gmail.com

Abstract

High-frequency words are processed faster than low-frequency words, known as the word frequency effect (FE). Although the FE has been studied in various writing systems as well as in first- (L1) and second-language (L2) reading, existing theoretical hypotheses are mainly based on findings in alphabetic languages. To date, no study has investigated theoretical explanations of the FE such as the learning hypothesis, the lexical entrenchment hypothesis and the rank hypothesis apply to Chinese–English bilinguals. The present study, therefore, compared the FEs in Chinese– and Dutch–English bilinguals during natural paragraph reading in their L1 and L2, using eye-tracking measures. Chinese bilinguals exhibited a larger FE in L2 than in L1. They displayed smaller L1 FEs and much steeper L2 FE curves than Dutch bilinguals. These findings are not entirely consistent with the existing FE hypotheses, and the present study discusses theoretical accounts in light of the observed results.

1. Introduction

The word frequency effect (FE) refers to the phenomenon that words with a higher frequency of occurrence are processed faster than those that appear less often. It has been well-studied in monolinguals and bilinguals of alphabetic languages (e.g., English; Gollan et al., 2011) and is one of the strongest factors affecting word processing (Brysbaert et al., 2016). The effect is also evident in reading Chinese (e.g., Li et al., 2014), a writing system that systematically differs from that of alphabetic languages in terms of spelling and pronunciation. However, recent evidence shows that although there is an overall FE in Chinese paragraph reading, the effect decreases and eventually disappears as its character frequency, a language-specific factor, increases (Sui et al., submitted). Consequently, the magnitude of the Chinese FE in natural reading may differ from that in alphabetic languages. In addition to such between-group comparisons across languages, an interesting line of research has also compared FEs within readers, between first language (L1) and second language (L2), in research on alphabetical languages. Typically, the FE is found to be larger in L2 than in L1 (Cop et al., 2015; Duyck et al., 2008; Mor & Prior, 2022; Whitford & Joanisse, 2018; Whitford & Titone, 2012, 2017).

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

Besides the above cross-lingual complications, it is also important to consider the (experimental) context in which words appear. Word recognition in natural text is influenced by a wide range of contextual influences (e.g., syntactical or semantic expectations), and differs from reading isolated words (Dirix et al., 2019; Kuperman et al., 2013). And also the FE observed in isolated word reading (e.g., the lexical decision task) appears larger than that for words embedded in sentences (e.g., in eye-tracking research; Dirix et al., 2019). Apparently, studying the FEs of words in sentences, which closely resembles reading in

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

everyday life, is essential for understanding language processing, especially in Chinese reading. This is because words are important units in Chinese reading, and their boundaries are often not clearly defined, making word segmentation essential for reading sentences but not for isolated words. In the current study, we therefore investigate FEs by comparing two eye-tracking corpora, GECO (Ghent Eye-tracking COrpus; Cop et al., 2017a) and GECO-CN (Ghent Eye-tracking COrpus for Chinese–English bilinguals; Sui et al., 2023), which recorded eye-movement data for L1s and L2s in paragraph reading for Chinese– and Dutch–English bilinguals, respectively.

Eye-tracking is a popular method used to study the underlying processes involved in sentence reading by monitoring the eye movements of the reader while reading. This approach provides a range of eye-movement measures (Rayner, 2009), such as saccades (the action of rapidly moving eyes to a new point) and fixations (the duration of eyes fixating on a specific point). There are multiple fixation duration measures, including (a) first-fixation duration (FFD), the duration of the initial fixation on a word; (b) gaze duration (GD), the summed duration of fixation on a word in the first pass and (c) total-reading times (TRTs), the summed duration of all fixations and refixations on a word. The first two are generally viewed as early measures (reflecting the initial stages of word identification, such as lexical access), whereas the last one, incorporating second-pass time, is considered a late measure (reflecting later stages, such as verification and integration; e.g., Boston et al., 2008; Clifton et al., 2007). Skipping probability refers to whether the word is skipped during the reading, not just in the first pass.

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

1.1 L1 FE in alphabetic languages

FE, the difference in processing times between low-frequency (LF) and high-frequency (HF) words, has been studied extensively in L1 reading of alphabetic languages (e.g., Cop et al., 2015; Rayner & Raney, 1996; Whitford & Titone, 2017). The effect is one of the most potent phenomena (explaining over 30% of the variance in lexical decision mega-studies; Brysbaert et al., 2016; Ferrand et al., 2010; Keuleers et al., 2010b, 2012; Yap & Balota, 2009) and is robust in both monolingual and bilingual adults and children (Cop et al., 2015; Whitford & Joanisse, 2018). Numerous reading experiments have shown that when reading in the first or dominant language, alphabetic language readers spend more time fixating on LF words and are less likely to skip them than HF words (e.g., Cop et al., 2015; Duyck et al., 2008; Whitford & Joanisse, 2018; see Rayner, 2009, for a review). The FE appears to be modulated by the degree of language exposure: readers with more language exposure exhibit a smaller FE (e.g., Ashby et al., 2005; Cop et al., 2015; Whitford & Titone, 2012, 2017). Some studies have shown that L1 and L2 fixation durations decrease as L1 exposure increases, unaffected by L2 exposure (Cop et al., 2015). However, others found that L2 exposure affects FEs in young adults: as L2 exposure increases, the FE decreases for L2 and increases for L1 (Whitford & Titone, 2012, 2017). Furthermore, English monolinguals and alphabetic language bilinguals (e.g., Dutch–English) exhibit comparable FEs in L1 reading (or dominant language; Cop et al., 2015 in sentence reading; Diependaele et al., 2013 in lexical decision; but see Whitford & Joanisse, 2018, for a larger L1 FE in English–French children compared to English monolinguals).

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

1.2 L1 FE in Chinese writing systems

Chinese is a logographic language that is qualitatively distinct from alphabetic languages. Chinese characters are written in strokes, and are the components of words. In Chinese, there are about 5,000 commonly used characters and they can constitute more than 56,000 words. The most encountered word type is two-character words, while the commonly used word tokens are one-character words, i.e., the characters themselves. One- and two-character words account for the majority of commonly used Chinese words (97.2%; Li & Pollatsek, 2020). Obviously, Chinese words are, on average, much shorter than those of alphabetic languages. Another major difference between Chinese and alphabetic languages is that the words of the former are not visually separated in sentences, whereas the latter contains spaces between words. That is, a character might be a single-character word or form a word with its preceding or following character in a Chinese sentence. Since the word is an important processing unit in Chinese reading (for the discussion, see Li et al., 2014; Sui et al., 2023), word segmentation is challenging, but undoubtedly necessary for Chinese sentence reading.

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

Note that most studies investigating the FE on eye movements in Chinese reading only studied target words (primarily two-character

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

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

1.3 L2 frequency effect

An increasing number of studies have investigated whether an FE also occurs in L2 reading. Evidence has shown that unbalanced bilinguals usually have a larger FE in L2 than in L1 reading (e.g., Cop et al., 2015; Duyck et al., 2008; Mor & Prior, 2020; Whitford & Titone, 2012, 2017). When language exposure (often measured by its proxy vocabulary size; Brysbaert et al., 2017) is included as a predictor in the analyses, the difference between the FEs in L1 and L2 reading becomes negligibly small in the lexical decision tasks (Brysbaert et al., 2017) but not in eye-movement studies (Cop et al., 2015), where FEs remain larger in the L2 than in the L1. Cop et al. (2015) explained that the distinct results observed in different experiments may be due to the usage of disparate methods. The eye-movement measures are, not surprisingly, more complex and time-sensitive than the reaction times obtained in lexical decisions.

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

1.4 FE hypotheses

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

The FE can also be explained by the lexical entrenchment hypothesis (e.g., Diependaele et al., 2013), which highlights the strength of lexical representations in memory. Frequent exposure to a word leads to more entrenched representations, resulting in faster and more accurate processing compared with LF words. Given that unbalanced bilinguals are generally less exposed to their L2, the objective frequency of their L2 should be lower than that of their L1. Both theoretical hypotheses predict a larger FE in L2 than in L1, consistent with the existing findings (for detailed discussion, see Duyck et al., 2008). Interestingly, they also predict that once the L1 and/or L2 exposure is similar in balanced bilinguals, the L1 and/or L2 FEs of the two groups should be similar in size, regardless of their writing systems (e.g., Chinese and Dutch).

Another possible explanation, the rank hypothesis (Murray & Forster, 2004), later extended to bilinguals (Duyck et al., 2008), suggests that the lexicon is organized into frequency-ordered bins with sequential searching, starting with HF words. For bilinguals, the bins are either language-specific (i.e., L1 or L2) with specific scanning speeds (longer scanning speed in L2) or shared by all known languages, with processing time increasing non-linearly with decreasing word frequency (Duyck et al., 2008). Nevertheless, regardless of the lexicon type (i.e., language-specific or shared), bilinguals with similar word frequency rankings should have comparable FEs, irrespective of language dominance or writing systems. The lexicon type might also be assumed to interact with cross-lingual similarity, being shared only when L1 and L2 employ the same writing system (e.g., Dutch–English) but separate for those with different scripts (e.g., Chinese–English, for a discussion, see Duyck et al., 2008). In this case, different-script bilinguals might exhibit smaller L1 and L2 FEs than those that shared bins (due to the increased nonlinearly searching times). If they have longer scan speed in L2, as assumed above, their L2 FEs may be larger than their L1 FEs and be similar in size with the same-script bilinguals.

To summarize, all the above assumptions predict that language exposure moderates the word FE, regardless of writing systems or language dominance (L1 or L2). Bilinguals with comparable L1 exposure (or have their L1 proficiency included in the analysis) should exhibit similar L1 FEs (except for one of the extended rank hypothesis that posit an interaction between lexicon type and cross-lingual similarity as it also predicts a larger FE in different-script bilinguals). Balanced bilinguals with similar exposure to both languages should have similar FEs in their L1 and L2. Conversely, unbalanced bilinguals with less L2 exposure should have a larger FE in L2 reading, either due to the relatively more asymptotic learning in their LF words, relatively weak lexical representations or to the well-behind location of LF L2 words or longer L2 scanning speed in the frequency-ranked bins. In addition, various hypotheses generate different predictions regarding the L2 FEs for bilinguals with distinct L1 wiring systems. The learning, lexical entrenchment and frequency-ranked (which assume that bins are language-specific or shared among languages) hypotheses suggest that the L2 FE should not be

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

affected by the L1 writing systems. Instead, the frequency-ranked hypothesis, which assumes that bins are only shared by alphabetic languages, indicates that the L2 FE may vary with the L1 writing system and that different-script bilinguals should exhibit smaller L2 effects than those with the same script.

2 Current study

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

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

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

We will compare eye-movement data from two large corpora, GECO (Dutch–English bilinguals; Cop et al., 2017a) and GECO-CN (Chinese–English bilinguals; Sui et al., 2023), in which unbalanced bilinguals read different language versions of an entire novel in paragraphs. The corpora shared identical experimental procedures and used the same reading materials. In the experiments, readers read half of the novel in their L1 and the other half in L2. The novel has approximately 5,000 sentences and contains a wide range of word stimuli, and thus word frequencies, in each language. Logically, the linguistic properties these two datasets involve should be comparable and not interfere with the comparison between the bilingual groups (e.g., their frequency distributions do not seem to differ significantly, see Figure S.1 in the Supplementary materials). In addition, both corpora provide LexTALE scores (Lemhöfer & Broersma, 2012; Dutch and English; HSK [Chinese Proficiency Test, n.d.] score for Chinese), which reflect language proficiency by examining the vocabulary size, which we will use as a proxy of language exposure. Notably, there was no significant difference in the L2 LexTALE scores among the bilingual participants in both corpora (see Sui et al., 2023).

3. Method

3.1 Participants and materials

3.1.1 GECO

GECO (Cop et al., 2017a) is an eye-movement corpus where 19 Dutch–English bilinguals (average age: 21.2; SD = 2.2; undergraduate and master students; also see Table S.1 in the Supplementary materials) and 14 British English monolinguals read an entire novel (*The Mysterious Affair at Styles* by Agatha Christie) while their eye-movement behaviour was measured. The participants read the novel in four self-paced sessions that each contained a fixed number of chapters. After each chapter, multiple choice questions were presented to ensure participants were, as instructed, reading for comprehension. Dutch natives read half of the novel in Dutch and the other half in English, whereas monolinguals read the entire book in English. For the present study, only the bilingual data were used. For further information on the corpus, we refer the reader to Cop et al. (2017a, 2017b).

3.1.2 GECO-CN

GECO-CN is a dataset consisting of eye-movement data from 30 Chinese–English bilinguals (average age: 25.3; SD = 2.60; undergraduate, master and PhD students). It follows the identical experimental procedure and uses the same reading materials as the original GECO (Cop et al., 2017a). Participants read half of the novel in Chinese and the other half in English. They also complete a series of language proficiency tests in both languages (see Table S.1). For more details, we refer the reader to Sui et al. (2023).

3.2. Analysis

This study only investigated content words (for Chinese–English bilinguals, 511,157 data points in Chinese and 442,638 in English; for Dutch–English bilinguals, 275,458 data points in Dutch and 264,634 in English) excluding all cognates, as these orthographically and semantically overlapping equivalents may confound the investigation of the FE. The present work classified a word as a cognate if its Levenshtein distance between the two languages was greater than or equal to .7 (in orthography; 5.19% of words in Dutch and 7.29% in English; also see Da Silveira & van Leussen, 2015). Cognates were only present among the Dutch–English texts. Furthermore, the first and last words of a line and fixations of less than 100 ms were removed from the analysis, as the former could reflect the sentence wrap-up effect (e.g., Rayner et al., 1989; 10.31% in Chinese and 16.97% in English for Chinese–English bilinguals; 17.3% in Dutch and 16.8% in English for Dutch–English bilinguals), while the latter fixations

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

are considered too short to reflect word processing (e.g., Sereno & Rayner, 2003).

This experiment used R software (Version 494) to perform linear mixed-effects models (for fixation durations) and generalized linear mixed-effects models (for skipping probability) from the lme4 package (Version 1.1-26). We conducted separate analyses for L1 and L2 and for different reading time measures, and considered important psycholinguistic predictors as control variables. In each model, predictor variables included group (categorical, Chinese *vs.* Dutch bilinguals), word frequency (continuous), word length (continuous), proficiency of the relevant language, congruent with the model (continuous; if L1 FEs were investigated, it is L1 proficiency) and the sequential numbering of word repetition in sessions (continuous; see FFD in Table 1 for the full model). Additionally, we examined various eye-movement measures as dependent variables, including FFD, GD and TRT (e.g., Clifton et al., 2007) and skipping probability. The random effects were the participant and the word token. The predictors were all centred, whereas the dependent variables were Box-Cox transformed. Such transformation normalized the distribution without changing its functional relationship. In each reading time measure, fixation durations differing by more than 2.5 standard deviations (SDs) per individual and per language were discarded.

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

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

4. Results

4.1 Bilingual L1: Chinese versus Dutch

4.1.1 First-fixation duration

The FFD in L1 reading did not differ significantly between the Chinese and Dutch bilingual groups (see Table 1). Both bilingual groups showed an overall FE, significantly larger effect in Dutch compared to Chinese bilinguals, showing that fixations were shorter for HF words than for LF ones. However, the word length effect was not significant in either group. Frequency and word length interacted significantly in both groups. The FE became larger as word length increased and increased significantly more in Dutch than in Chinese bilinguals (see Figure 1A). Language proficiency did not influence word fixation durations or the FE in either group or did the repetition effect.

4.1.2 Gaze duration

Chinese and Dutch bilinguals showed no significant difference in GD (see Table 1) but in frequency and word length effects. Both groups exhibited frequency and word length effects, with shorter GDs for higher frequency or shorter words. However, Dutch bilinguals showed significantly steeper effects compared to Chinese bilinguals. The interaction between frequency and word length observed in Chinese bilinguals differed significantly from that in Dutch bilinguals. The FE increased with word length and was more pronounced in Dutch bilinguals (see Figure 1A). Similar to what was observed in FFD, the language proficiency of Chinese and Dutch bilinguals did not affect GD or the FE. Furthermore, neither Dutch nor Chinese bilinguals exhibited a word repetition effect.

4.1.3 Total reading time

Overall, there was no significant difference in TRTs between Chinese and Dutch bilinguals (see Table 1). The frequency and word length effects were significant in Chinese bilinguals, with significantly smaller frequency and significantly larger word length than the Dutch group. The interactions between frequency and word length was not evident in Chinese bilinguals, differing significantly from that observed in Dutch bilinguals (see Figure 1A). In Dutch bilinguals, word length exhibited a greater effect on LF words than on HF ones, showing a larger FE in long words. L1 proficiency did not appear to affect TRTs and FEs in either group. In contrast to the findings in FFD and GD, the repetition effect was significant in both groups, showing inhibitory effect in Chinese and facilitatory pattern in Dutch bilinguals.

4.1.4 Skipping probability

Chinese bilinguals demonstrated a significantly higher skipping probability than Dutch bilinguals (see Table 1). Their frequency and word length effects were significant, with higher frequency or shorter words being more likely to be skipped, and were significantly smaller in frequency and larger in word length compared to Dutch bilinguals. The interaction between frequency and word length did not yield significance in Chinese bilinguals (see Figure 1A), differing significantly from Dutch groups, where the FE was larger in short words. In addition, Chinese bilinguals exhibited an inhibitory repetition effect, with a low skipping probability for words that were repeated more often, significantly differing from the performance of Dutch bilinguals, who showed a facilitative effect, with higher word repetition related to higher skip rates.

4.2 Bilingual L2: English versus English

4.2.1 First-fixation duration

FFDs were not significantly different between Chinese and Dutch bilinguals, with a tendency for longer FFDs in the former group (see Table 2). Chinese bilinguals exhibited frequency and word length effects, which were significantly larger than those observed in Dutch bilinguals. The interaction between frequency and word length found in Chinese bilinguals significantly differed from that in Dutch bilinguals (see Figure 1B). As the word length reduced, the decrease in FFDs was greater for HF than for LF words in Chinese bilinguals, whereas it was greater for LF words in Dutch bilinguals. Language proficiency had no effects on Chinese bilinguals but interacted with frequency, not with FFD, in Dutch bilinguals (see Figure 2). The FE decreased as language

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

Table 1. Comparative analyses of fixation duration measures and skipping probabilities in first-language reading between Chinese– and Dutch–English bilinguals

	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
First-fixation duration	**(Intercept)**	**4.42**	**.01251**	**353.389**	**<.00025*** **	
	GroupDutch	−.006788	.02049	−.331	.74196	1.13
	Frequency	**−.002032**	**.00062**	**−3.278**	**.00104****	3.944
	Word length	−.01632	.007143	−2.285	.02231	4.397
	L1 proficiency	.000104	.003982	.026	.97934	4.112
	Repetition	.000012	.000007	1.684	.09217	2.087
	GroupDutch:Frequency	**−.005959**	**.000965**	**−6.174**	**<.00025*** **	4.464
	GroupDutch:Word length	.009347	.009995	.935	.34968	5.215
	GroupDutch:L1 proficiency	−.000767	.004675	−.164	.87036	3.949
	Frequency:Word length	**−.01321**	**.004298**	**−3.074**	**.00211****	4.285
	Frequency:L1 proficiency	−.000032	.000126	−.258	.79657	4.978
	GroupDutch:Repetition	−.00003	.000014	−2.06	.03942	2.044
	GroupDutch:Frequency:Word length	**−.01435**	**.005586**	**−2.568**	**.01022***	5.09
	GroupDutch:Frequency:L1 proficiency	.00035	.000143	2.455	.01409.	4.992
	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
Gaze duration	**(Intercept)**	**2.698**	**.004325**	**623.81**	**<.00025*** **	
	GroupDutch	−.001164	.007084	−.164	.8702	1.13
	Frequency	**−.0011**	**.000216**	**−5.091**	**<.00025*** **	3.938
	Word length	**.0249**	**.002492**	**9.991**	**<.00025*** **	4.371
	L1 proficiency	−.000102	.001377	−.074	.9413	4.112
	Repetition	.000004	.000002	1.738	.0821	2.089
	GroupDutch:Frequency	**−.002119**	**.000336**	**−6.31**	**<.00025*** **	4.448
	GroupDutch:Word length	**.009374**	**.003482**	**2.692**	**.0071***	5.165
	GroupDutch:L1 proficiency	−.000152	.001616	−.094	.9254	3.949
	Frequency:Word length	**−.007469**	**.001513**	**−4.938**	**<.00025*** **	4.21
	Frequency:L1 proficiency	.000077	.000043	1.796	.0724	4.969
	GroupDutch:Repetition	−.000012	.000005	−2.46	.0139.	2.045
	GroupDutch:Frequency:Word length	**−.008119**	**.001966**	**−4.131**	**<.00025*** **	4.983
	GroupDutch:Frequency:L1 proficiency	.000066	.000049	1.354	.1758	4.984
	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
Total reading time	**(Intercept)**	**2.236**	**.002739**	**816.247**	**<.00025*** **	
	GroupDutch	.001979	.004486	.441	.6613	1.13
	Frequency	**−.000761**	**.000147**	**−5.193**	**<.00025*** **	3.942
	Word length	**.03654**	**.001694**	**21.572**	**<.00025*** **	4.39
	L1 proficiency	−.000103	.000872	−.118	.90625	4.113
	Repetition	**.000007**	**.000002**	**4.064**	**<.00025*** **	2.091
	GroupDutch:Frequency	**−.001228**	**.000227**	**−5.404**	**<.00025*** **	4.449
	GroupDutch:Word length	**−.006887**	**.002362**	**−2.916**	**.00354***	5.185
	GroupDutch:L1 proficiency	−.000004	.001024	−.004	.99661	3.949
	Frequency:Word length	−.000182	.001034	−.176	.86064	4.235
	Frequency:L1 proficiency	.000027	.000029	.938	.34848	4.957
	GroupDutch:Repetition	**−.000019**	**.000003**	**−5.584**	**<.00025*** **	2.047

(*Continued*)

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

Table 1. (*Continued.*)

	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
	GroupDutch:Frequency:Word length	**−.01281**	**.001339**	**−9.567**	**<.00025*** **	4.999
	GroupDutch:Frequency:L1 proficiency	.000064	.000032	1.985	.04715	4.973
	Predictors	Estimate	Std. error	*z* value	Pr(>|*z*|)	VIF
Skipping probability	**(Intercept)**	**.2511**	**.07716**	**3.254**	**.00114****	
	GroupDutch	**−1.256**	**.126**	**−9.968**	**<.00025*** **	1.002
	Frequency	**.02367**	**.005184**	**4.565**	**<.00025*** **	2.937
	Word length	**−5.489**	**.06178**	**−88.837**	**<.00025*** **	2.756
	Repetition	**−.000447**	**.000055**	**−8.145**	**<.00025*** **	1.993
	GroupDutch:Frequency	**.09071**	**.009541**	**9.507**	**<.00025*** **	3.286
	GroupDutch:Word length	**−.539**	**.1045**	**−5.157**	**<.00025*** **	3.212
	Frequency:Word length	−.1011	.04226	−2.391	.01679.	2.182
	GroupDutch:Repetition	**.000853**	**.000123**	**6.952**	**<.00025*** **	1.945
	GroupDutch:Frequency:Word length	**−.164**	**.06386**	**−2.568**	**.01024***	2.293

Estimate, estimates; Std. error, standard errors; *t* value, *t*-values; Pr(>|*t*|), *p*-values (calculated using the lmerTest package); VIF, variance inflation factor; Bold values indicate $p < .0125$. *$p < .0125$, **$p < .0025$, ***$p < .00025$ (corrected significant level according to Von der Malsburg & Angele, 2017).

proficiency increased, with HF words being affected more than LF words. In addition, both groups spent more time reading frequently repeated words than infrequently repeated words, and the effect was similar between them.

4.2.2 Gaze duration

Different from the findings in FFD, Chinese bilinguals spent more time on GD than Dutch bilinguals (see Table 2). The observed frequency and word length effects in Chinese bilinguals were statistically larger than those in Dutch bilinguals (see Figure 1B). Both groups exhibited a significant interaction between frequency and word length, with Chinese bilinguals showing a statistically smaller decrease in the FE as word length reduced. Language proficiency showed no effect on fixation duration but did affect the FE in both groups. As language proficiency increases, Chinese bilinguals showed a significantly more pronounced decrease in FEs than Dutch bilinguals (see Figure 2), particularly in GD for LF words, whereas Dutch bilinguals showed a greater decrease in GD for HF than LF words. Additionally, Chinese and Dutch bilinguals showed repetition effects with comparable size, both spending more time reading frequently repeated words compared to infrequently repeated ones.

4.2.3 Total reading time

Chinese bilinguals spent more time on TRT than their Dutch counterparts (see Table 2). Frequency and word length in the two groups were negatively and positively correlated with TRTs, respectively, with smaller effect sizes in Dutch bilinguals. There was a significant interaction between frequency and word length in Chinese and Dutch bilinguals. The FE increased with word length and to a greater extent in Dutch bilinguals, especially in LF words (see Figure 1B). The effect of L2 proficiency on fixation duration was not evident in the two groups, as in FFD and GD (see Figure 2). However, it interacted with frequency in Chinese but not in Dutch bilinguals. Highly proficient readers exhibit a smaller FE, mainly manifested in the greater influence on the fixation duration of LF words. Both groups spent more time reading the more repeated words and to a similar extent.

4.2.4 Skipping probability

Chinese bilinguals skipped fewer words than Dutch bilinguals, different from findings in the L1 reading (see Table 2). Chinese bilinguals were affected by frequency and word length, not language proficiency, indicating a higher skipping probability for HF or short words. Their word length effect was statistically smaller than Dutch bilinguals but their frequency and language proficiency effects were comparable. Interactions between frequency and word length or between frequency and language proficiency were significant in Chinese bilinguals and differed significantly from those observed in Dutch bilinguals. As word length increased, the larger FE in short words decreased more in Dutch than in Chinese bilinguals (see Figure 1B). As language proficiency increased, the FE decreased in Chinese and increased in Dutch bilinguals, manifesting in higher skipping probability for LF and HF words, respectively (see Figure 2). The word repetition effect was significant and did not differ in the two groups. The more times a word is repeated, the lower the skipping probability.

4.3 Chinese–English bilingual: L1 versus L2

4.3.1 First-fixation duration

The FFD was significantly shorter in L1 than in L2 (see Table S.2 in the Supplementary materials). The FE was significant in L1, whereas the word length effect was not. Both were statistically smaller than those in L2. HF or shorter words were processed faster than LF or longer ones. The FE did not increase with word length in L1, but in L2, showing a significant difference between the two languages (see Figure S.1 in the Supplementary materials). Both languages exhibited a word repetition effect positively correlated with FFDs, with no significant difference between them. In addition, the results showed that FFD increased with language proficiency.

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

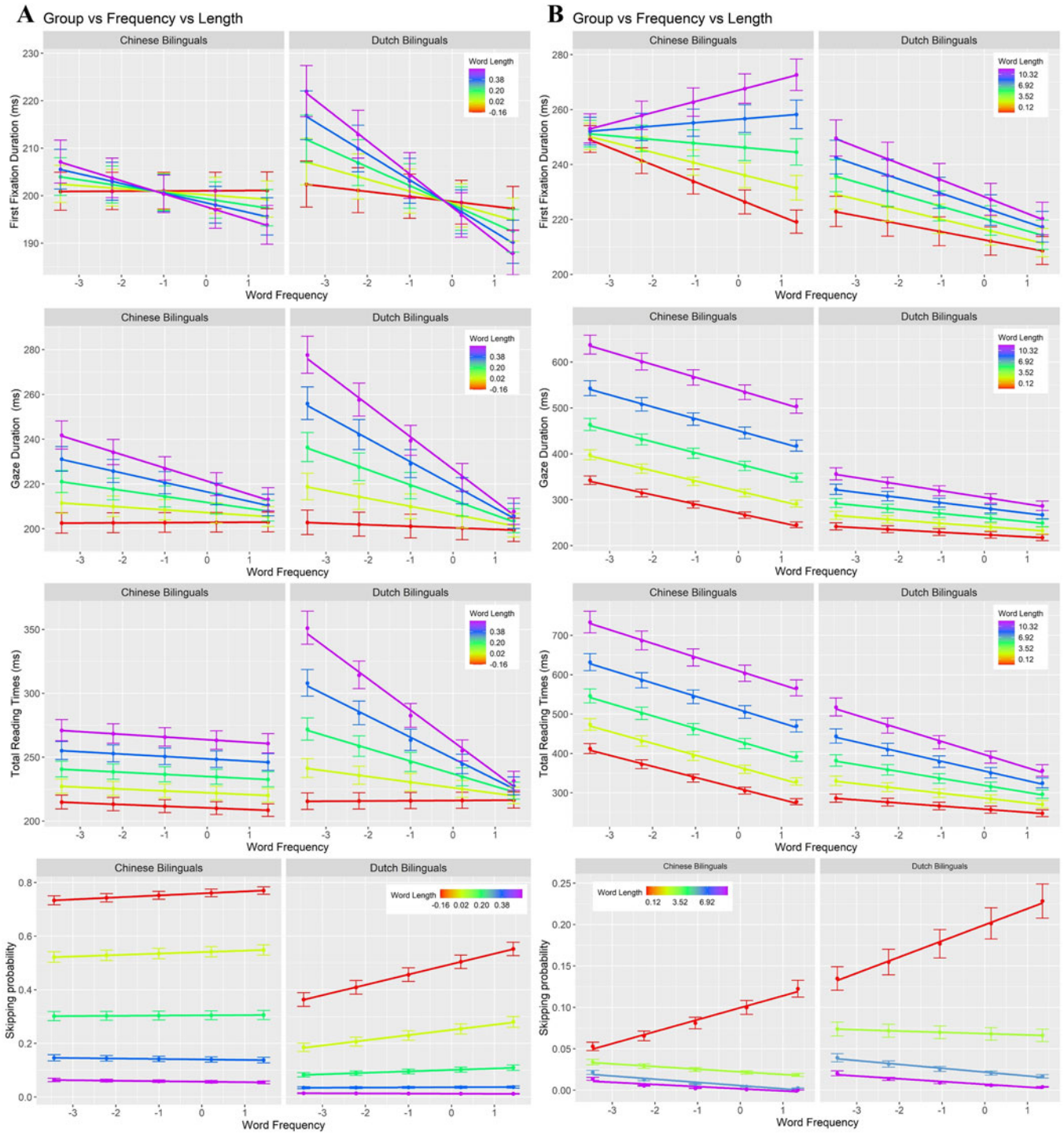


Figure 1. Three-way interaction plots between the group, word length and frequency in first-language (column A) and second-language (column B) reading.

4.3.2 Gaze duration

Readers spend less time reading in L1 than in L2 (see Table S.2). Frequency and word length effects were observed in both languages, with statistically larger effects in L2. GDs were negatively correlated with frequency and positively correlated with word length. Frequency and word length interacted in both languages. FE increased with word length and increased significantly more in L2 than in L1 (see Figure S.2 in the Supplementary materials). Language proficiency was negatively correlated with GD, showing that highly proficient readers had shorter GDs than low-proficient ones. Language proficiency also interacted with frequency, with FE decreasing as proficiency increased. In addition, the repetition effect was significant, with no major differences between the two languages. The higher the number of word occurrences, the longer the GDs.

4.3.3 Total reading time

The TRTs were significantly longer in L2 than in L1 (see Table S.2). Frequency and word length effects were significant

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

Table 2. Comparative analyses of fixation duration measures and skipping probabilities in second-language reading between Chinese– and Dutch–English bilinguals

	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
First-fixation duration	**(Intercept)**	**6.267**	**.02529**	**247.844**	**<.00025*** **	
	GroupDutch	−.1014	.04061	−2.498	.01621.	1.00
	Frequency	**−.02875**	**.001148**	**−25.044**	**<.00025*** **	3.78
	Word length	**.01566**	**.000618**	**25.328**	**<.00025*** **	3.68
	L2 proficiency	−.002214	.00213	−1.039	.30413	1.71
	Repetition	**.000057**	**.000019**	**2.948**	**.0032***	2.36
	GroupDutch:Frequency	**.008806**	**.001807**	**4.873**	**<.00025*** **	3.48
	GroupDutch:Word length	**−.008814**	**.001012**	**−8.707**	**<.00025*** **	3.62
	GroupDutch:L2 proficiency	.001527	.003311	.461	.64681	1.71
	Frequency:Word length	**.004098**	**.00034**	**12.049**	**<.00025*** **	2.34
	Frequency:L2 proficiency	−.000005	.000053	−.085	.93257	1.58
	GroupDutch:Repetition	.000012	.000032	.383	.70178	2.28
	GroupDutch:Frequency:Word length	**−.005291**	**.000558**	**−9.486**	**<.00025*** **	2.34
	GroupDutch:Frequency:L2 proficiency	**−.000229**	**.000088**	**−2.598**	**.00937***	1.58
	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
Gaze duration	**(Intercept)**	**3.805**	**.01096**	**347.163**	**<.00025*** **	
	GroupDutch	**−.09668**	**.0176**	**−5.493**	**<.00025*** **	1.00
	Frequency	**−.02813**	**.000503**	**−55.904**	**<.00025*** **	3.71
	Word length	**.02072**	**.000276**	**74.981**	**<.00025*** **	3.72
	L2 proficiency	−.000942	.000923	−1.02	.313	1.71
	Repetition	**.000027**	**.000007**	**3.936**	**<.00025*** **	1.53
	GroupDutch:Frequency	**.01736**	**.000743**	**23.363**	**<.00025*** **	3.21
	GroupDutch:Word length	**−.01102**	**.000431**	**−25.547**	**<.00025*** **	3.55
	GroupDutch:L2 proficiency	.000494	.001435	.344	.732	1.71
	Frequency:Word length	**.000783**	**.00015**	**5.206**	**<.00025*** **	2.23
	Frequency:L2 Proficiency	**.000148**	**.000022**	**6.834**	**<.00025*** **	1.59
	GroupDutch:Frequency:Word length	**−.001457**	**.000229**	**−6.375**	**<.00025*** **	2.04
	GroupDutch:Frequency:L2 proficiency	**−.000271**	**.000035**	**−7.649**	**<.00025*** **	1.59
	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
Total reading time	**(Intercept)**	**3.675**	**.01068**	**344.146**	**<.00025*** **	
	GroupDutch	**−.07784**	**.01715**	**−4.539**	**<.00025*** **	1.00
	Frequency	**−.02886**	**.000541**	**−53.307**	**<.00025*** **	3.78
	Word length	**.01756**	**.000297**	**59.081**	**<.00025*** **	3.79
	L2 proficiency	−.000741	.000899	−.823	.4146	1.71
	Repetition	**.00002**	**.000007**	**2.759**	**.0058***	1.53
	GroupDutch:Frequency	**.01552**	**.00079**	**19.647**	**<.00025*** **	3.27
	GroupDutch:Word length	**−.005946**	**.000458**	**−12.983**	**<.00025*** **	3.61
	GroupDutch:L2 proficiency	−.000279	.001398	−.199	.8429	1.71
	Frequency:Word length	**.000929**	**.000161**	**5.767**	**<.00025*** **	2.25
	Frequency:L2 proficiency	**.000187**	**.000022**	**8.496**	**<.00025*** **	1.59

(*Continued*)

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

Table 2. (*Continued.*)

	Predictors	Estimate	Std. error	*t* value	Pr(>|*t*|)	VIF
	GroupDutch:Frequency:Word length	**−.002101**	**.000243**	**−8.66**	**<.00025*** **	2.06
	GroupDutch:Frequency:L2 proficiency	**−.000181**	**.000036**	**−4.999**	**<.00025*** **	1.59
	Predictors	Estimate	Std. error	*z* value	Pr(>|*z*|)	VIF
Skipping probability	**(Intercept)**	**−2.565**	**.09327**	**−27.498**	**<.00025*** **	
	GroupDutch	**.9059**	**.1496**	**6.057**	**<.00025*** **	1.00
	Frequency	**.1218**	**.009138**	**13.327**	**<.00025*** **	4.77
	Word length	**−.4628**	**.005604**	**−82.588**	**<.00025*** **	4.48
	L2 proficiency	.01801	.00784	2.297	.0216	1.71
	Repetition	−.0001921	.00007779	−2.469	.0135.	1.47
	GroupDutch:Frequency	−.02233	.01213	−1.841	.0656	4.24
	GroupDutch:Word length	**.105**	**.007585**	**13.849**	**<.00025*** **	4.24
	GroupDutch:L2 proficiency	−.009839	.01218	−.808	.4192	1.71
	Frequency:Word length	**−.09443**	**.002817**	**−33.521**	**<.00025*** **	2.15
	Frequency:L2 proficiency	**−.002984**	**.0004415**	**−6.757**	**<.00025*** **	2.25
	GroupDutch:Frequency:Word length	**.04805**	**.003937**	**12.206**	**<.00025*** **	2.04
	GroupDutch:Frequency:L2 proficiency	**.007032**	**.000596**	**11.806**	**<.00025*** **	2.22

Estimate, estimates; Std. error, standard errors; *t* value, *t*-values; Pr(>|*t*|), *p*-values (calculated using the lmerTest package); VIF, variance inflation factor; Bold values indicate $p < .0125$. *$p < .0125$, **$p < .0025$, ***$p < .00025$ (corrected significant level according to Von der Malsburg & Angele, 2017).

in L1, with longer TRTs for LF or long words. These effects were statistically smaller in L1 than in L2. The interaction between frequency and word length was not significant in L1 and differs statistically from that in L2 (see Figure S.2). FE in L2 increased with word length. Language proficiency affects GD and FE, with highly proficient readers having shorter GDs and smaller

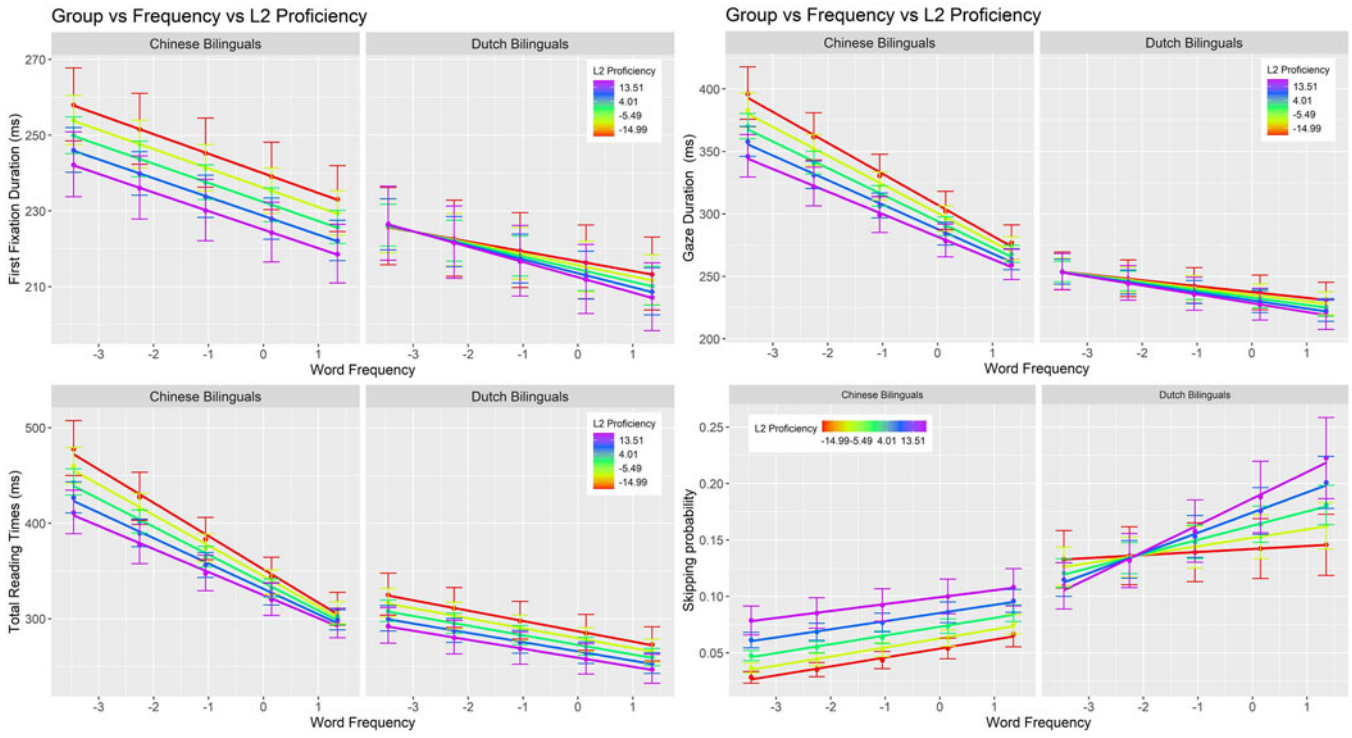


Figure 2. Three-way interaction plots between the group, language proficiency and frequency in second-language reading.

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

FE. The repetition effect was significant and positively correlated with TRTs, with no differences between languages.

4.3.4 Skipping probability

The skipping probability in L1 was higher than that in L2 (see Table S.2). The frequency and word length effects were significant in L1, with statistically smaller frequency and larger word length effects than those in L2. The higher the frequency or the shorter the word, the higher the probability of skipping it. The frequency and word length did not interact in L1, significantly different from that in L2 (see Figure S.2). The results showed that the reverse FE increased with decreasing word length. The language proficiency effect was significant. Readers with higher proficiency have a higher skipping probability. It also interacted with FE, showing that reverse FE decreased with increasing language proficiency. The repetition effect was significant, with no difference between L1 and L2. The more the repetitions, the lower the probability of skipping.

5. Discussion

This work compared FEs between bilinguals with the same and different scripts in L1 and L2 reading and between L1 and L2 in Chinese bilinguals. Our three objectives were to examine (a) whether the L1 FEs are similar in size across writing systems, (b) whether L2 FEs differ across readers with distinct L1 writing systems and (c) whether the L2 FE is larger than that of L1 in Chinese–English bilinguals. Language proficiency, known to affect the FE, was also taken into account to ensure that any potential differences in FEs between the two groups were not due to variations in language proficiency. Below, we will discuss the comparative results of FEs in L1 and L2 reading between the Chinese and Dutch groups. Within-group comparisons of Chinese–English bilinguals (L1 *vs.* L2) will be discussed briefly in the first subsection to avoid repetition. Following that, we will relate these empirical observations to the predictions of the different theoretical accounts of the FE.

5.1. Bilingual L1: Chinese versus Dutch

In contrast to previous studies reporting longer fixation durations for Chinese readers in single-sentence reading (Liversedge et al., 2016; Rayner et al., 2005; but see Sui et al., 2023), this work shows that they read their L1 as quickly as alphabetic language readers and has a much higher skipping probability. That is, Chinese readers have a much higher reading speed than Dutch bilinguals for texts of comparable length in L1. Divergent findings from previous studies may be due to the different nature of the reading material (e.g., single sentences *vs.* paragraphs; controlled sentences *vs.* natural sentences; for a discussion, see Sui et al., 2023). Here, we used a very natural form of reading, with meaningful, contextualized materials (a book).

Chinese words, although generally much shorter than alphabetic ones, show a reliable length effect in GD, TRT and skipping probabilities. Fixation duration increased with word length, deviating from the U-shaped pattern found in previous research using a lexical decision task (Ferrand et al., 2010; Tsang et al., 2017). The discrepancy may stem from methodological differences, as the limited shared variance between lexical decision task and eye-tracking data (Dirix et al., 2019; Kuperman & Van Dyke, 2013) shows that lexical decision differs substantially from the natural reading process. Word length interacts with FEs in all reading measures for Dutch bilinguals but only in FFD and GD for Chinese bilinguals. In Dutch reading, the lexical access for LF words is prolonged more as word length decreases than HF words, and the same logic applies to later word processing stages. In Chinese reading, however, the prolongation effect is limited to the earlier stages of word recognition.

The L1 FEs of alphabet readers appear to be influenced by language proficiency rather than language quantity (monolinguals or bilinguals; Cop et al., 2015; Diependaele et al., 2013) or language (Dutch or French; Diependaele et al., 2013). Based on this logic, readers with similar proficiency levels should exhibit comparable L1 FEs. Yet, Dutch bilinguals exhibited larger FEs than Chinese bilinguals in all reading measures and skipping probabilities in this study (see Figures 1A and 3), inconsistent with certain discussed FE hypotheses, which will be explored further below. In addition, the FEs in the L1 were significantly smaller than those in the L2 across all reading time measures and the skip probability in Chinese–English bilinguals, congruent with previous findings (e.g., Cop et al., 2015; Mor & Prior, 2020). HF words had shorter fixation times or higher skip probability.

Interestingly, when examining a single-reading measure, the impact of language proficiency on FEs was significant in FFD and TRT for Dutch bilinguals and marginally significant in GD for Dutch and Chinese bilinguals. Proficient readers spent less time on LF words than those with lower language proficiency, consistent with Kuperman and Van Dyke's (2013) explanation of larger FE in those with less language exposure due to having less exposure to LF words. Yet, when investigating multiple eye-tracking measures, the interaction became insignificant for Dutch and Chinese bilinguals after adjusting the significant level to avoid an increase in false-positive probability (Von der Malsburg & Angele, 2017). Nevertheless, language proficiency appears to affect FEs for Dutch bilinguals to some extent, although not statistically powerful enough.

The smaller FE for Chinese readers in skipping probabilities could be due to a ceiling effect, as their skip rate reaches a surprisingly high .6. The relatively smaller FE in the time measures may have several possible explanations: firstly, certain language-specific factors, namely, character complexity, may affect FE. However, previous research did not find an interaction between character complexity and word frequency, arguing against the assumption (Sui et al., submitted). Secondly, the number of words in languages may affect the FE. If Chinese has significantly fewer words than Dutch, Chinese words are likely to occur more often, resulting in a reduced FE. However, the number of commonly used words in Chinese and Dutch (about 56,000 words in Chinese and 54,319 in Dutch; Brysbaert et al., 2019; Li & Su, 2022) and the frequency distribution of the analysed data (see Figure S.1) do not differ significantly, collectively arguing against this possibility.

Thirdly, the FE on Chinese reading may be more limited than on Dutch. Given that word frequency interacts with the frequency of its constituent characters in Chinese reading (for a discussion, see Sui et al., submitted), its impact on word recognition could be attenuated by character frequency. If, as previous work suggested, HF characters may have a greater inhibitory effect on HF words but a greater facilitative effect on LF words, it can explain the small FEs and the limited effect of language proficiency on FE in Chinese reading. However, the interaction of character and word frequencies has primarily been explored in two-character words, leaving unclear its applicability to other word lengths, especially single-character words with high collinearity between

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

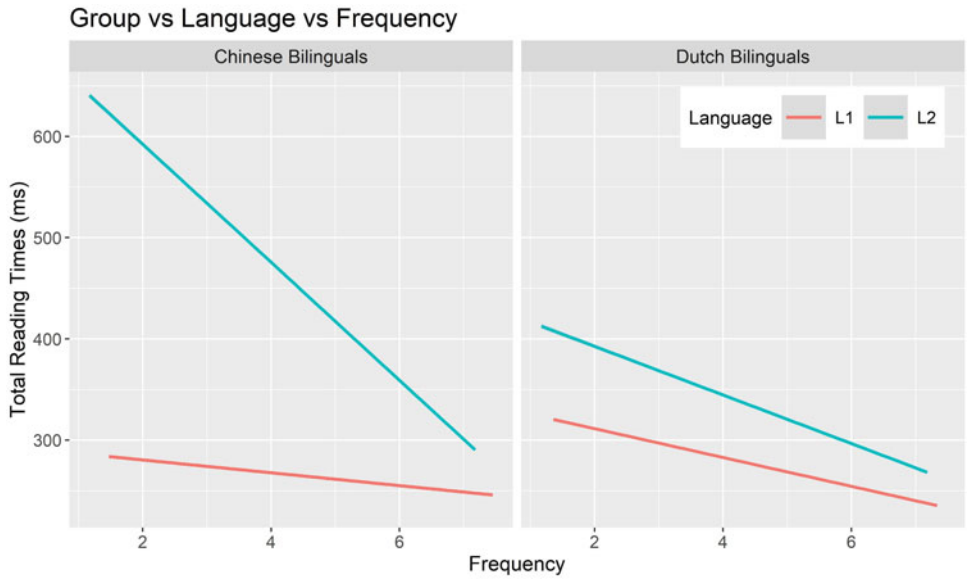


Figure 3. Graph of raw TRTs for word FEs in L1 and L2 reading for Chinese and Dutch bilinguals. Grey shadows are confidence intervals.

them and multi-character words with varying numbers of characters and character frequencies. Further research is needed to investigate whether the character frequency influence is responsible for the smaller word FE in Chinese with solid bases and well-designed.

Fourthly, the shorter word length constrains the degree of variation in FE. Since FE decreases with word length, which positively correlates with visual complexity, and the fact that Chinese words are generally much shorter than alphabetic language words, it is not surprising that their FEs are smaller and less affected by language proficiency, given the limited variations in FE. However, Chinese characters are composed of strokes, and the visual complexity of a short Chinese word may not necessarily be lower than that of a long alphabetic language word (i.e., number of letters). Hence, whether the effect of word length on FEs is similar in Chinese and Dutch and whether it can explain the smaller FEs in Chinese reading require further verification.

5.2. Bilingual L2: English versus English

In L2 reading, Chinese bilinguals exhibit longer fixation durations and larger FEs across reading time measures than Dutch bilinguals, differing from findings in L1. Yet, their skipping probabilities are lower than Dutch bilinguals, and their smaller FE observed in it could be explained by a floor effect. As Figures 2 and 3 illustrate, Chinese bilinguals read somewhat slower than their Dutch counterparts, even for HF words. It implies that even with similar L2 proficiency and read the same material in the same L2, bilinguals whose languages are from different writing systems are less efficient at visual word processing than those from one writing system.

One possible explanation for the findings could be the relatively limited exposure of Chinese bilinguals to the alphabetic writing system (i.e., letters and their specific combinations in orthographic structures such as bigrams or trigrams). Indeed, LexTALE scores indicated comparable English proficiency for Dutch and Chinese bilinguals (see Sui et al., 2023), and so should their exposure to the L2. Yet, the two languages of Dutch–English bilinguals use the same Latin alphabet and share some underlying orthographic structures, exhibiting more similarities in writing than those of Chinese–English bilinguals. Thus, in this particular context, the reported FE differences between the Dutch and Chinese groups might be explained by the exposure to *alphabetic languages* but not to English (i.e., L2). The facilitation effect may be greater on LF than on HF words, as the latter may already be approaching the ceiling effect. This possibility could explain longer English reading times overall, even for HF words, and larger FEs in Chinese compared to Dutch bilinguals in both early and late measures.

Another possible explanation is cross-lingual lexical interactions. The languages of bilinguals are well-known to co-activate even in unilingual reading (Brysbaert & Duyck, 2010; Dijkstra & Van Heuven, 2002; Schwartz & Kroll, 2006). Word recognition in a target language is influenced by non-target language words, explaining cognate and cross-language neighbourhood effects, etc. (e.g., Cop et al., 2017b; Dirix et al., 2017; Whitford & Joanisse, 2021). Previous studies have shown that the greater the within- or/and cross-language neighbourhood density, the smaller the L2 FE of the LF words (Dirix et al., 2017; Whitford & Titone, 2019). Chinese characters, however, differ fundamentally from the Latin alphabet, resulting in a much more limited cross-linguistic effect than those of the same writing system. Thus, they should have slower reading speeds and a larger FE than those with the same script, compatible with what we found.

One may argue that the larger FE observed in Chinese bilinguals could be due to a specific language proficiency test or lower language proficiency. Indeed, despite the comparable LexTALE scores between the two groups, Chinese bilinguals scored lower than Dutch bilinguals in the WRAT4 and Lexical decision task and should, therefore, show larger FEs. To investigate this possibility, we employed the same procedures and models, substituting LexTALE scores for WRAT4 and Lexical decision task ones, respectively (see Table S.3 in the Supplementary materials). Evidence shows that L2 FEs for Chinese–English bilinguals remain significantly larger than those for Dutch–English bilinguals, arguing against the possibilities.

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

Another argument is that different language exposure environments explain the varied L2 FEs between groups. Chinese–English bilinguals in the study may have a greater exposure to academic English. They may know HF words but need to comprehend LF words through context, resulting in a larger FE and longer fixation duration in L2. If so, word frequency would not affect fixation durations beyond a certain frequency level. Additionally, the effect of language proficiency on FE should disappear since it primarily affects LF words. Yet, evidence shows a linear negative correlation between FE and fixation duration, with slower growth at lower frequencies and interactions between language proficiency and FE, arguing against this hypothesis.

The interactions between frequency and language proficiency were observed in GD and TRT in Chinese bilinguals and at the early processing stage in Dutch (i.e., FFD and GD), with significant differences between the groups. The absence of interaction in the very early measure of Chinese bilinguals may be due to the multiple fixations they adopted (Chinese bilinguals fixate more on LF long words than on HF or short words; also see Figure S.3 in the Supplementary materials) or the fact that language proficiency does not affect the earliest stages of word recognition in Chinese bilinguals, such as the sub-lexical orthographic stage. The disproportionate effect of language proficiency on word processing was greater for LF words in Chinese bilinguals, compatible with previous findings (e.g., Mor & Prior, 2020; Whitford & Titone, 2012).

Unexpectedly, Dutch bilinguals showed different patterns. Language proficiency had a greater impact on the processing of HF words, with negative correlations in reading measures and a positive correlation in skipping probability. Considering trends in L1 and the observed patterns in Chinese bilinguals in L2 are congruent with previous findings (Whitford & Titone, 2012 in French–English bilinguals; Mor & Prior, 2020 in Hebrew–English bilinguals), we speculate that it is due to the diverse language environment to which bilinguals are exposed. Some highly proficient bilinguals may encounter LF L2 words as frequently as less-proficient ones but have more exposure to HF words. Then, their language proficiency should particularly affect the recognition of HF words, consistent with what we found.

Failing to find the effect of language proficiency on LF words in Dutch bilinguals may also be due to the influence of L1 proficiency, as obtained previously (Whitford & Titone, 2019). The L1, generally with greater exposure, should have a stronger impact on the L2 frequency than vice versa. Consequently, Dutch–English bilinguals may exhibit different results in their two languages but not Chinese bilinguals, as proficiency in distinct languages is unlikely to influence each other (Mor & Prior, 2020). To examine this possibility, we performed an additional analysis, employing the same analysis procedure and models to investigate the effect of L1 proficiency on L2 frequency. Given the high correlation between L1 and L2 proficiencies in Dutch bilinguals ($r = .69$), we include only L1 proficiency instead of both in the analysis. Results show that L1 proficiency has a greater effect on LF words in GD ($\beta = .000156$, SE = .000030, t-value = 5.234; also see Figure S.4 in the Supplementary materials), congruent with Cop et al. (2015). In this case, the L1 proficiency of Dutch bilinguals should increase, but not reduce the difference in fixation duration of LF words between proficient and less-proficient bilinguals, arguing against this possibility.

Chinese bilinguals exhibited larger word length effects than Dutch bilinguals, possibly due to not being accustomed to reading long words, as English words are usually longer than Chinese and shorter than Dutch words. Word length affects FE more in Dutch than in Chinese bilinguals. Frequency curves shifted upwards with word length, more with word length in Chinese bilinguals and more in the LF ranges in Dutch bilinguals. Note that at the earliest stages of word recognition (i.e., FFD), the FE of Chinese bilinguals decreases with word length increase, especially for HF words. We speculate that it is due to refixations and conducted further analysis using the same analysis procedure, taking fixation counts as the dependent variable. FE interacted with word length in the Chinese group ($\beta = -.0009618$, SE = .0002217, t-value = −4.338), significantly different from that in the Dutch group ($\beta = -.001433$, SE = .000352, t-value = −4.07). Chinese bilinguals refixated LF long words more frequently than short or HF words or Dutch readers (see Figure S.3), explaining the absence of the word length effect in LF words in FFD but in GD and TRT.

5.3. Theoretical discussion

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

These models do not differentiate in their assumptions regarding distinct writing systems. They all predict that the FE becomes smaller with increasing language exposure. That is, bilinguals with greater language exposure should have smaller FEs than those with limited exposure and similar FEs for those with similar exposure or proficiency, regardless of language writing systems and dominance (L1 or L2). In the present study, Chinese bilinguals reported much smaller L1 FEs than Dutch ones in all reading time measures. Their L1 FEs were unaffected by language exposure, while Dutch bilinguals exhibited trends. Interestingly, Chinese bilinguals showed larger L2 FEs despite similar L2 proficiency levels (assessed by English LexTALE), i.e., similar exposure to the L2, between them and Dutch–English bilinguals. These findings cannot be explained by the aforementioned FE hypotheses, challenging their applicability to logographic writing systems.

Another extended frequency-ranked hypothesis suggests language-specific bins for non-alphabetic and alphabetic language pairs and shared bins for alphabetic language pairs. That is, if Dutch–English bilinguals have much less L2 exposure than L1 (with all L2 words ranking behind L2 words), their L1 FEs may be comparable with Chinese–English bilinguals; otherwise, they should exhibit larger L1 FE, consistent with the findings obtained. Yet, this hypothesis fails to explain why the FE of Chinese words is unaffected by language exposure. More importantly, it predicts Dutch bilinguals to have a larger L2 FE than Chinese bilinguals, contradicting the current findings that Dutch showed much smaller, rather than larger, FEs in their L2 than Chinese bilinguals.

Alternative theoretical accounts for the present FE findings are the BIA+ model (Dijkstra & van Heuven, 2002) and the "weaker links" hypothesis (Gollan et al., 2005). These models suggest that exposure influences the activation speed and strength of links between word forms and lexical representations, respectively. Consequently, bilinguals, who often have less exposure to their L2 than L1, should exhibit larger L2 FEs. Yet, they seem, for

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

the moment, failed to explain the larger L2 FEs in Chinese than in Dutch bilinguals despite having similar L2 proficiency. Another explanation for FE is the language-competition hypothesis (Diependaele et al., 2013; also see Dijkstra & Van Heuven, 2002), which suggests that co-activated representations compete for selection across languages. In this case, the FE in L2 may be larger than in L1 as the interference from dominant L1 representations is likely greater than vice versa. Additionally, Chinese–English bilinguals should exhibit smaller FEs than same-alphabet bilinguals due to less cross-language competition between different writing systems. This hypothesis can explain the L1 FE found in this study but fails to explain its L2 findings as well as the differences in FEs within monolinguals and bilinguals of alphabetic languages. For instance, FEs in the same L2 do not vary with the orthographic similarity between the first- and second-alphabetic languages (e.g., Dutch–English *vs.* German–English; Diependaele et al., 2013). Clearly, all the existing hypotheses on word FEs fail to account for the current findings.

5.4. The word frequency hypotheses: implementation and limitations

So far, existing FE hypotheses have generally been considered universal across languages. Indeed, the effect has been taken as evidence of the similarity between Chinese and alphabetic writing systems in the underlying processes it involves (e.g., Li et al., 2014; but see Sui et al., submitted). Yet, this study shows that due to differences in L1 writing systems, the FEs vary between L1s and even between the same L2s, suggesting potential variations in the underlying processes involved in FE between Chinese– and Dutch–English bilinguals.

One key point in explaining the present findings is the word component effect (e.g., characters or letters), as the constituents of a word inevitably affect word recognition and FEs. With this extended assumption, the learning and lexical entrenchment accounts, but not the rank hypothesis, could explain the findings. That is, word frequency may affect activation thresholds or baselines, or the entrenchment of lexical representations. In Chinese, character frequency moderates the word FE. HF characters may facilitate the recognition of LF words and cause interference with HF words (Sui et al., submitted). Since Chinese characters often appear as single-character words with word frequency, they may only provide additional activation rather than having a similar influence as words.

In contrast, languages within the same writing system apparently share word components (e.g., bigrams, trigrams), which affect word processing (e.g., Kuperman et al., 2008; New & Grainger, 2011). These components may have similar facilitative effects alike to word frequency or provide extra activation for accessing target words. They should affect both L1 and L2 FEs, being more prominent on LF words and limited on HF words that are already close to the threshold. Thus, since Chinese characters have not only facilitative but also cause interference effect, the L1 word FE in Chinese reading is expected to be smaller than in Dutch. Additionally, Chinese–English bilinguals, lacking morphologically language-shared components, are expected to exhibit longer overall fixation durations and larger FEs in L2 than Dutch–English bilinguals, who have greater exposure to language-shared word component, explaining the variation in L2 FEs with L1 writing systems.

Until now, there are no (computation) models of reading that simulate reading across different script languages, as well as the interactions between those languages. Dominant models of bilingualism like BIA+ (Dijkstra & van Heuven, 2002) have almost exclusively only been validated for same-script bilingualism. At present, it is unclear to what extent BIA+'s assumption about cross-language similarity holds for different-script bilingualism. At least, our observed differences between languages suggest that processing differs across such different languages in various aspects. Future research could examine this with more diverse language pairs.

In addition, word frequency typically explains about 30–40% of variances in reaction times in Chinese (e.g., Tsang et al., 2017; Tse et al., 2017) and different alphabetic scripts (e.g., Ferrand et al., 2010, 2018; Keuleers et al., 2010a) in isolated word recognition, inconsistent with the smaller FE observed in the current study for Chinese than for Dutch reading. The discrepancy could be explained by the use of different methodologies (also see Dirix et al., 2019; Kuperman et al., 2013). In addition to word-level variables, the only factors that affect reading times in isolated word recognition, word recognition in context is also affected by top-down factors such as contextual content. Consequently, the FE in the former condition should be smaller than in the latter (Dirix et al., 2019). Future research should take the use of different research methods into account, depending on the research purpose. In addition, future studies could also examine the applications of the available effects observed in context-free conditions across paradigms.

6. Conclusion

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

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S136672892400035X.

Data availability statement. Materials and data are from GECO (Dutch–English bilinguals; Cop et al., 2017a) and GECO-CN (Chinese–English bilinguals; Sui et al., 2023), while the analysis code for this study is available at: https://osf.io/bg7nd/.

Acknowledgements. We would like to thank Professor Titus von der Malsburg for his valuable suggestions in correcting the significance level appropriately and two anonymous reviewers for their valuable input on earlier versions of this paper.

Competing interests. None.

References

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam sentence corpus. *Journal of Eye Movement Research*, *2*(1), 1–12. doi: 10.16910/jemr.2.1.1

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

Brysbaert, M., Keuleers, E., & Mandera, P. (2019). Recognition times for 54 thousand Dutch words: Data from the Dutch crowdsourcing project. *Psychologica Belgica*, *59*(1), 281–300. https://doi.org/10.5334/pb.491

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., Dufau, S., Mathôt, S., & Grainger, J. (2018). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods*, *50*(3), 1285–1307. https://doi.org/10.3758/s13428-017-0943-1

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition 33*(7), 1220–1234. https://doi.org/10.3758/bf03193224

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

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

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

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

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

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

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

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

Li, X., 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., & 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* (2nd ed.). The Commercial Press.

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

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

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

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

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

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

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press

Mor, B., & Prior, A. (2022). Frequency and predictability effects in first and second language of different script bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(9), 1363–1383. https://doi.org/10.1037/xlm0000927

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

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

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

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

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

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

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

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

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

Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2023). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese–English bilinguals. *Behavior Research Methods*, *55*, 2743–2763.

Sui, L., Woumans, E., Duyck, W., & Dirix, N. (submitted) Word frequency effects differ across writing systems: The word and character frequency effect in Chinese reading. *Language and Cognition*.

Tsang, Y.-K., Huang, J., Lui, M., Xue, M., Chan, F., Wang, S., & Chen, H.-C. (2017). MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods*, *50*, 1763–1777. doi: 10.3758/s13428-017-0944-0

Tse, C.-S., Yap, M. J., Chan, Y.-L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese lexicon project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*, *49*, 1503–1519.

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

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133.

Whitford, V., & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A focus on word frequency effects. *Journal of Experimental Child Psychology*, *173*, 318–337.

Whitford, V., & Joanisse, M. F. (2021). Eye movement measures of within-language and cross-language activation during reading in monolingual and bilingual children and adults: A focus on neighborhood density effects. *Frontiers in Psychology*, *12*, 674007. https://doi.org/10.3389/fpsyg.2021.674007

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

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

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

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

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

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

https://doi.org/10.1017/S136672892400035X Published online by Cambridge University Press