https://doi.org/10.29140/9781763711686-3

Can Native Sentence Reading Processing Mechanisms Predict Second Language Reading Behavior? Computational Modeling of GECO-CN Eye-Tracking Data in Chinese-English Bilinguals

Fang Yifei
College of Chinese Language and Culture, Nankai University, China

Abstract

This study investigates whether cognitive parameters extracted from first-language (L1, Chinese) reading can directly predict second-language (L2, English) reading behaviors and whether such parameters reflect interlanguage development. As a methodological innovation, we adopt a learner-corpus-based computational approach that integrates eye-tracking data from the GECO-CN corpus, which includes 30 Chinese-English bilinguals reading a full novel, with individualized E-Z Reader model fitting. By applying L1-fitted parameters to L2 texts for zero-parameter transfer prediction, we found that L1 processing routines do not transfer directly to L2 reading; the transfer model yielded significantly higher prediction errors than a native-English baseline. Crucially, online processing dynamics (specifically changes in word-frequency sensitivity and saccadic preparation) were stronger predictors of L2 reading comprehension accuracy than offline vocabulary size (LexTALE), with the former showing significant positive and negative correlations, respectively. This integration of learner corpora with computational modeling reveals a dissociation between lexical competence and procedural reading skills, suggesting that static vocabulary knowledge alone cannot account for real-time reading efficiency. The findings

E-mail: fangyifei@mail.nankai.edu.cn

In Y. Wang, J. Matthews, & B. Zou (Eds.), *Proceedings of the 2nd International Applied Linguistics Research Conference, 2026* (pp. 17–25). Castledown.

offer empirical insights for L2 reading instruction, demonstrating that computational modeling of naturalistic eye-tracking data can uncover online processing parameters that are more predictive of reading comprehension than static vocabulary measures alone, thus highlighting the need to attend to procedural aspects of reading beyond lexical accumulation.

Keywords: learner corpus; eye-tracking; computational modeling; second language reading; Chinese-English bilinguals

Introduction

Reading involves complex interactions between lexical processing, prediction, and eye-movement control. During natural reading, readers continuously allocate visual attention and adjust fixation patterns according to lexical and contextual information. Eye-tracking provides direct measurements of these online processes, while computational models allow researchers to infer latent cognitive mechanisms underlying observed eye movements.

Cross-linguistic differences between Chinese and English provide an important context for investigating bilingual reading. Chinese uses a character-based writing system without explicit interword spaces, whereas English relies on alphabetic representations with clear word boundaries. These differences influence visual processing, lexical segmentation, and eye-movement strategies. Previous studies have shown that although reading mechanisms share certain universal characteristics, such as frequency effects and predictability effects, language-specific orthographic properties also shape reading behavior (Li & Pollatsek, 2020; Li et al., 2022).

Research on second-language (L2) reading has traditionally focused on behavioral outcomes, such as comprehension scores or vocabulary measures, but these measures provide limited information about real-time cognitive processes. Eye-tracking and computational modeling offer a more direct method for examining how bilingual readers adapt their processing strategies across languages. Models such as E-Z Reader have successfully decomposed reading behavior into psychologically interpretable parameters, including lexical access efficiency and saccadic control (Reichle et al., 2003).

However, whether reading mechanisms developed in a native language can predict L2 reading behavior remains unclear. Chinese-English bilinguals provide a particularly informative case because their two languages differ substantially in orthographic structure. The present study uses naturalistic eye-tracking data from the GECO-CN corpus and individualized computational modeling to address two questions: (1) Can cognitive parameters extracted from L1 Chinese reading predict L2 English reading behavior without re-fitting? and

(2) Which processing parameters change during cross-linguistic adaptation, and are these changes related to reading comprehension? We hypothesize that although some general reading mechanisms may transfer across languages, bilingual readers must adapt specific processing strategies to accommodate L2 reading demands.

Method

Materials

Data were obtained from the GECO-CN corpus (Sui et al., 2022), an open-access bilingual eye-tracking corpus containing natural reading data from Chinese-English bilinguals. The corpus includes 30 native Chinese speakers who were late learners of English and completed naturalistic reading of the novel The Mysterious Affair at Styles in both Chinese and English versions. Participants also completed reading comprehension questions and language proficiency assessments, including the LexTALE vocabulary test.

The corpus provides continuous eye-movement measures during long-form reading, allowing the investigation of cross-language differences under ecologically valid conditions. After excluding invalid trials and extremely low-frequency words, all participants met the criteria for individual computational modeling.

Computational Modeling and Analysis

An individualized computational model based on the E-Z Reader framework was applied to estimate latent cognitive parameters underlying reading behavior. The model incorporated lexical properties (word length and frequency) and eye-movement measures (first fixation duration, gaze duration, and skipping rate). Four parameters were estimated for each participant: lexical access time (a1), lexical frequency sensitivity (a2), saccadic preparation time (m1), and visual attention span (θ) (Dias et al., 2021).

For each participant, L1 Chinese reading data were first fitted using Differential Evolution optimization to obtain individual cognitive profiles. These L1-derived parameters were then transferred directly to L2 English reading data without re-fitting to test whether native reading mechanisms could predict L2 processing. A native-English baseline model was used for comparison.

Subsequently, L2-specific parameters were independently estimated through the same fitting procedure. Differences between L1 and L2 parameters were calculated and correlated with English reading comprehension accuracy and LexTALE vocabulary scores to examine whether changes in online processing mechanisms were associated with L2 reading performance.

For a detailed explanation and the content of the Methods section, please refer to the Supplementary Material https://drive.google.com/file/d/1ejqAD0hyO0kphj5-R_Ez1rFXydEPomXH/view?usp=sharing

Results

Zero-parameter Transfer Prediction Test

To test whether L1 reading mechanisms could predict L2 reading behavior, individual L1 Chinese parameters were directly applied to English reading data without re-fitting. The transfer model produced significantly higher prediction errors than the native-English baseline model (transfer: $M = 3.98$, $SD = 0.24$; baseline: $M = 3.75$, $SD = 0.34$), $t(29) = 4.08$, $p < .001$, Cohen's $d = 0.80$. This result indicates that L1 reading parameters cannot directly predict L2 eye-movement behavior. For detailed results for this part, please refer to Supplementary Table 3.

Cross-language Differences in Individual Cognitive Parameters

As direct transfer prediction was unsuccessful, L2-specific parameters were estimated through independent model fitting. Compared with L1 reading, bilinguals showed significantly higher lexical access time (a1), reduced lexical frequency sensitivity (a2), and longer saccadic preparation time (m1) during L2 reading (all PS < .001). Visual attention span (θ) also decreased after cross-language adjustment, although its interpretation requires caution due to differences in orthographic structure.

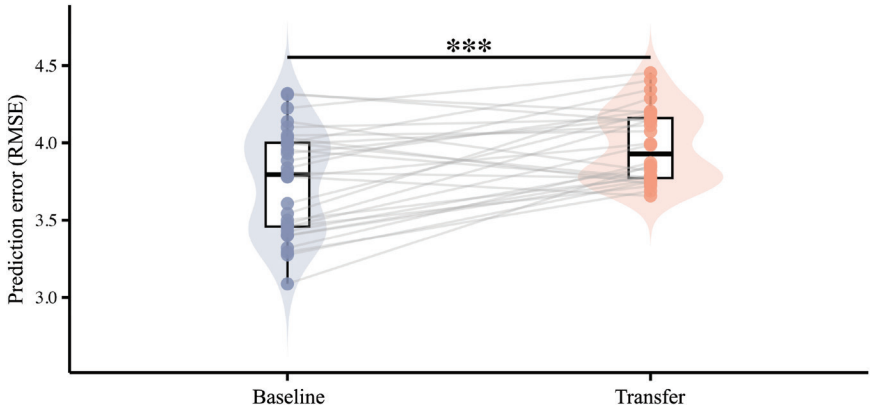

Figure 1. Results of the cross-language zero-parameter transfer prediction from L1 Chinese to L2 English

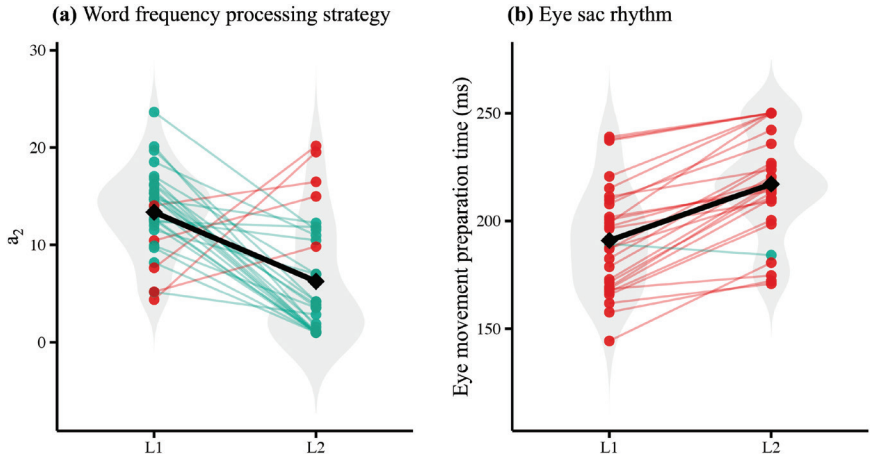

Figure 2. Individual differences in core cross-language cognitive parameters. (a) Paired-samples t-test results for lexical frequency sensitivity (a_2) between L1 and L2; (b) Paired-samples t-test results for saccadic preparation time (m_1) between L1 and L2

Relationship Between Cross-language Changes in Individual Cognitive Parameters and Individual Vocabulary Proficiency

Changes in lexical frequency sensitivity (Δa2) were positively correlated with English reading comprehension accuracy ($r = .42$, $p = .020$), whereas changes in saccadic preparation time (Δm1) showed a negative correlation ($r = -.50$, $p = .005$). No significant correlations were found between parameter changes and LexTALE vocabulary scores, suggesting that online processing efficiency and vocabulary knowledge may reflect partially independent aspects of L2 reading ability.

Discussion

The present study investigated whether native-language reading mechanisms could predict second-language reading behavior through computational modeling of naturalistic eye movements. The zero-parameter transfer analysis showed that L1 Chinese-derived parameters did not predict L2 English reading as accurately as the native-English baseline model. This finding suggests that reading mechanisms developed through L1 experience cannot be directly transferred to L2 processing. Instead, bilingual readers appear to adapt their processing strategies according to language-specific orthographic and lexical characteristics.

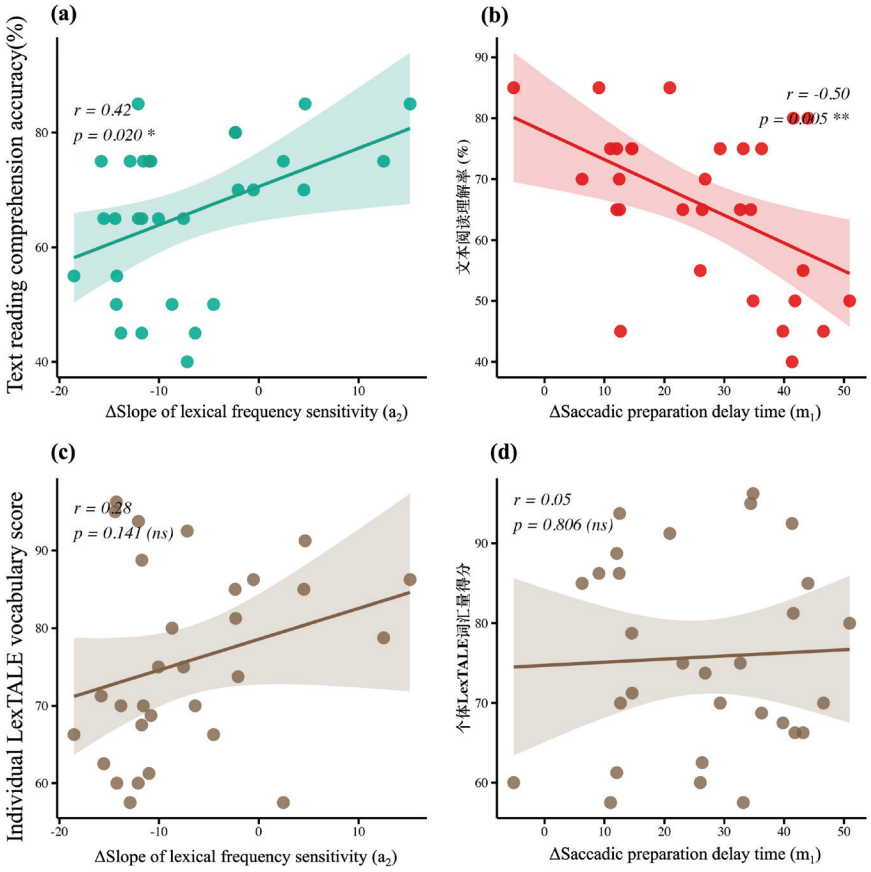

Figure 3. (a) Relationship between frequency processing and text reading comprehension accuracy; (b) Relationship between oculomotor rhythm and text reading comprehension accuracy; (c) Relationship between frequency processing and individual vocabulary proficiency; (d) Relationship between oculomotor rhythm and individual vocabulary proficiency

This result is consistent with previous research showing that reading involves both universal and language-specific mechanisms. Although frequency effects and predictability effects are observed across different writing systems (Li et al., 2022; Whitford & Titone, 2015), orthographic properties influence how readers allocate attention and control eye movements. Chinese and English differ substantially in visual and lexical structures: Chinese lacks explicit word boundaries and relies more heavily on character combinations and contextual information, whereas English uses alphabetic representations with clear spacing cues (Li & Pollatsek, 2020; McBride, 2016). Therefore,

bilingual readers may need to modify lexical processing and saccadic strategies when switching between languages.

The re-fitted L2 parameters further revealed significant changes in lexical frequency sensitivity and saccadic preparation time. The reduced frequency sensitivity during English reading suggests that bilingual readers may not utilize lexical frequency information in the same way across languages. Previous eye-movement studies have demonstrated that frequency effects are fundamental to lexical access across writing systems, although their magnitude can vary depending on linguistic characteristics (Yan et al., 2006). Similarly, the relationship between slower saccadic preparation and lower comprehension accuracy indicates that oculomotor efficiency contributes to successful discourse reading. This finding supports the view that eye movements reflect not only visual scanning behavior but also underlying cognitive resource allocation during reading (Rayner, 1998; Reichle et al., 2006).

Importantly, computational parameters derived from natural reading were associated with comprehension performance, whereas LexTALE vocabulary scores were not significantly related to these parameters. This suggests a potential dissociation between offline lexical knowledge and online reading procedures. Previous research has emphasized that proficient bilingual reading depends not only on vocabulary size but also on efficient lexical retrieval and processing strategies (Kiliańska-Przybyło & Ślęzak-Świat, 2021). In natural discourse reading, contextual information may also support comprehension by facilitating lexical access and reducing the impact of incomplete lexical knowledge (Liu Zhifang et al., 2020).

These findings provide implications for understanding interlanguage development. Previous bilingual research has shown that L2 proficiency influences cross-language competition and processing strategies (Pélissier et al., 2023). The current computational modeling approach extends this perspective by demonstrating that changes in reading procedures can be quantified through individual cognitive parameters. However, the present study has several limitations. The sample consisted of intermediate-to-advanced Chinese-English bilinguals, and future studies should include participants with different proficiency levels and native-language control groups. In addition, the current model represents an initial approximation of reading mechanisms and requires further validation with larger datasets and more advanced computational approaches.

Conclusion

Using the GECO-CN bilingual eye-tracking corpus and individualized computational modeling, this study showed that L1 Chinese reading parameters

cannot directly predict L2 English reading behavior. Instead, bilingual readers exhibit language-specific adaptations in lexical processing and oculomotor control. Changes in lexical frequency sensitivity and saccadic preparation were associated with reading comprehension, whereas vocabulary knowledge alone did not explain these online processing differences. These findings highlight the value of computational modeling for understanding dynamic bilingual reading mechanisms and suggest that L2 reading development involves changes in processing procedures beyond vocabulary accumulation.

References

Dias, E. C., Sheridan, H., Martínez, A., Sehatpour, P., Silipo, G., Rohrig, S., Hochman, A., Butler, P. D., Hoptman, M. J., Revheim, N., & Javitt, D. C. (2021). Neurophysiological, oculomotor, and computational modeling of impaired reading ability in schizophrenia. *Schizophrenia Bulletin*, *47*(1), 97–107. https://doi.org/10.1093/schbul/sbaa107

Kiliańska-Przybyło, G., & Ślęzak-Świat, A. (2021). Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide. Aline Godfroid. *Digital Scholarship in the Humanities*, *36*(1), 243–246. https://doi.org/10.1093/llc/fqaa075

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. https://doi.org/10.1038/s44159-022-00022-6

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, *127*(6), 1139–1162. https://doi.org/10.1037/rev0000248

Liu, Z., Tong, W., Zhang, Z., & Zhao, Y. (2020). The impact of contextual predictivity on word processing during reading: Eye-tracking evidence. *Acta Psychologica Sinica*, *52*(9), 1031–1047.

McBride, C. A. (2016). Is Chinese special? Four aspects of Chinese literacy acquisition that might distinguish learning Chinese from learning alphabetic orthographies. *Educational Psychology Review*, *28*(3), 523–549. https://doi.org/10.1007/s10648-015-9318-2

Pélissier, M., Haugland, D., Handeland, B., Urland, B. Z., Wetterlin, A., Wheeldon, L., & Frisson, S. (2023). Competition between form-related words in bilingual sentence reading: Effects of language proficiency. *Bilingualism: Language and Cognition*, *26*(2), 384–401. https://doi.org/10.1017/S1366728922000529

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, *7*(1), 4–22. https://doi.org/10.1016/j.cogsys.2005.07.002

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

Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2022). GECO-CN: Ghent eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods*, *55*(6), 2743–2763. https://doi.org/10.3758/s13428-022-01931-3

Whitford, V., & Titone, D. (2015). Second-language experience modulates eye movements during first- and second-language sentence reading: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1118–1129. https://doi.org/10.1037/xlm0000093

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. https://doi.org/10.1348/000712605X70066