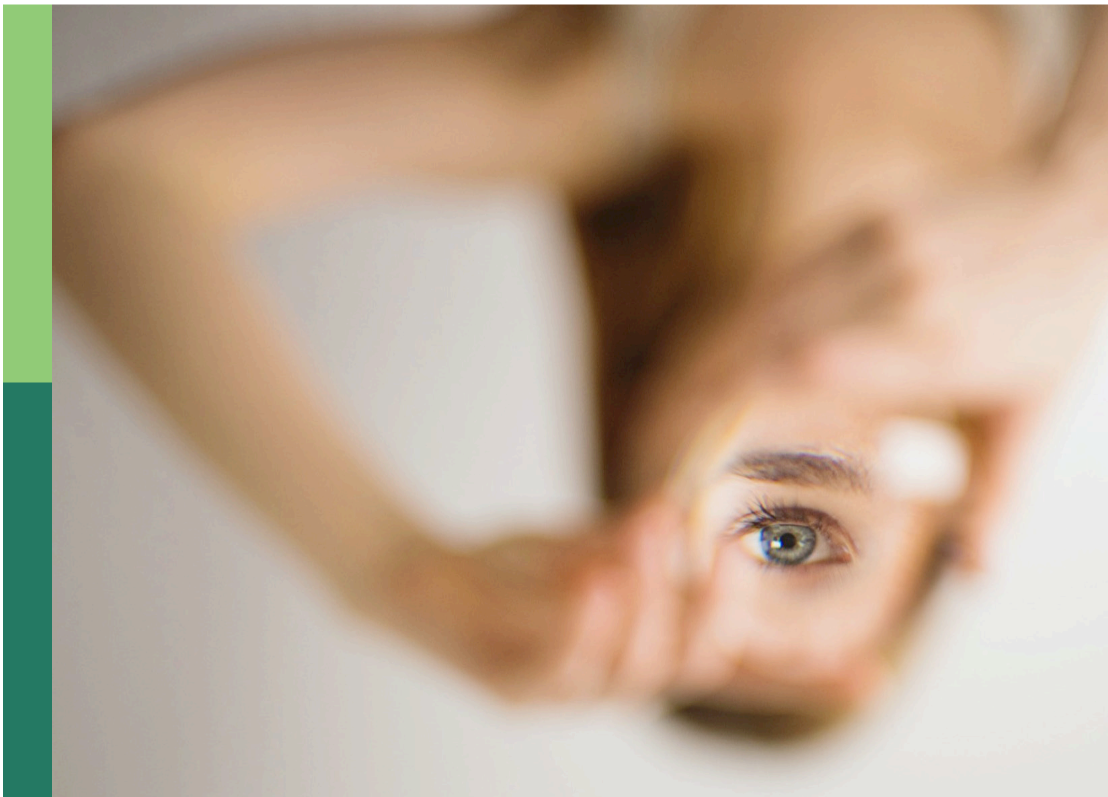
frontiers | Research Topics

Reading acquisition of Chinese as a second/foreign language, volume II

Edited by
Linjun Zhang, Zaizhu Han and Yang Zhang

Published in
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-3691-9
DOI 10.3389/978-2-8325-3691-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Reading acquisition of Chinese as a second/foreign language, volume II

Topic editors
Linjun Zhang — Peking University, China
Zaizhu Han — Beijing Normal University, China
Yang Zhang — University of Minnesota Twin Cities, United States

Citation
Zhang, L., Han, Z., Zhang, Y., eds. (2023). *Reading acquisition of Chinese as a second/foreign language, volume II*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3691-9

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 27 September 2023
DOI 10.3389/fpsyg.2023.1292975

OPEN ACCESS

EDITED AND REVIEWED BY
Caicai Zhang,
The Hong Kong Polytechnic University,
Hong Kong SAR, China

*CORRESPONDENCE
Linjun Zhang
✉ zhanglinjun75@gmail.com
Yang Zhang
✉ zhang470@umn.edu

RECEIVED 12 September 2023
ACCEPTED 19 September 2023
PUBLISHED 27 September 2023

CITATION
Zhang L, Han Z and Zhang Y (2023) Editorial: Reading acquisition of Chinese as a second/foreign language, volume II.
Front. Psychol. 14:1292975.
doi: 10.3389/fpsyg.2023.1292975

Editorial: Reading acquisition of Chinese as a second/foreign language, volume II

Linjun Zhang[1]*, Zaizhu Han[2] and Yang Zhang[3]*

[1]School of Chinese as a Second Language, Peking University, Beijing, China, [2]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China, [3]Department of Speech-Language-Hearing Sciences, Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, MN, United States

KEYWORDS

reading acquisition, Chinese as a second/foreign language, phonological awareness, morphological awareness, cognitive factors

Editorial on the Research Topic
Reading acquisition of Chinese as a second/foreign language, volume II

There has been a significant increase in the number of Chinese as a Second/Foreign Language (CSL/CFL) learners worldwide, particularly over the past two decades, due to a combination of economic, cultural, educational, and technological factors. However, learning Mandarin Chinese presents unique challenges for second language learners as they face specific hurdles such as tone acquisition and the complexity of Chinese morphology and writing (Kecskés and Sun, 2017; Gong et al., 2020; Ke, 2020; Wen, 2020; Chan et al., 2022). "*Reading Acquisition of Chinese as a Second/Foreign Language (CSL/CFL) – Volume II*" features a collection of 11 articles that represent continuing efforts within a general behavioral and neurocognitive framework to provide informative and insightful original findings on the Research Topic (Zhang et al., 2023).

Two child-focused studies addressed early learning challenges. Chan et al. investigated Chinese character acquisition in culturally diverse second language (L2) learners, comparing 491 L2 children (aged 3 to 6) to 240 first language (L1) peers in Hong Kong kindergartens using the Chinese Character Acquisition Assessment (Chan et al., 2020). The data revealed not only L2 learners' strength in sound and meaning but also struggles with character forms, emphasizing the importance of early oral support for L2 preschoolers. Zhou and McBride examined the connection between Cantonese phonological awareness, Mandarin Pinyin invented spelling, and English spelling in multilingual first and second language Cantonese-speaking second and third graders in Hong Kong. While both groups demonstrated similar skills in Cantonese phonological awareness and Mandarin Pinyin invented spelling, L1 speakers outperformed L2 speakers in Mandarin Pinyin tone skills, suggesting phonological transfer effects in multilingual contexts.

Three studies examined cross-language influences in adult learners. Hu and Zhao investigated code-switching costs in Chinese–English relative clause processing. The results highlighted the contribution of syntactic processing, particularly in head movement during relative clause comprehension, to these costs. Yan et al. examined how bilingualism and English proficiency may influence the processing of the Chinese sentence-final particle "le." Participants included English-dominant second language learners (L2), heritage learners, and monolingual Mandarin speakers. The study found that heritage learners showed sensitivity to semantic conditions similar to the target language, influenced by early exposure and positive cross-linguistic influence from English. This indicates that early bilingualism can shape the processing of Chinese sentence structures in heritage learners. Chai and Bao explored the influence of linguistic differences between learners' native languages (L1) and Chinese on character, vocabulary, and grammar acquisition. Using data from 58,240 Chinese as a second language (CSL) learners with diverse L1 backgrounds and proficiency levels, the study employs the World Atlas of Language Structures (WALS) index to measure linguistic distance. Results show that closer linguistic distance aids character and vocabulary acquisition across proficiency levels, but as proficiency increases, linguistic distance hinders grammar acquisition, indicating proficiency-dependent effects on CSL learners.

Three studies delve into cognitive factors in L2 Chinese reading. Yang et al. assessed 252 international students from Pakistan, Indonesia, Malaysia, and Laos and revealed that linguistic and cognitive skills account for 80% of L2 Chinese reading variation, with a significant role played by morphological awareness. Path analysis underscores the influence of these skills on reading comprehension through higher-order cognitive skills, highlighting the significance of morphological awareness in L2 Chinese reading. Feng and Jiang explored the context predictability effect (Li et al., 2022) and prediction error cost in reading, highlighting challenges and slower reading in advanced Chinese L2 learners during predictive processing. Wang et al. studied the impact of metacognition monitoring on L2 Chinese audiovisual comprehension. Findings suggest that absolute calibration accuracy significantly predicts L2 Chinese audiovisual comprehension, while relative calibration accuracy does not, emphasizing the importance of considering various aspects of metacognition monitoring in L2 learning strategies and addressing individual learner differences.

Three studies used electrophysiogical and neuroimaging methods. Liu et al. investigated how native Mandarin speakers and L2 learners process quadrisyllabic idiomatic expressions (QIEs) in Chinese. L2 learners processed high-frequency QIEs faster, while native speakers handle low-frequency ones more quickly, possibly due to semantic satiation. The fMRI data showed native speakers used the anterior cingulate for cognitive control during high- and low-frequency QIE comparisons. For comparing idiomatic and pseudo-idiomatic constructions, semantic processing in bilateral temporal poles dominated, suggesting that native speakers rely on conceptual understanding, rather than syntax, when processing Chinese idiomatic expressions. Ding et al. studied how conceptual concreteness may influence the acquisition and integration of novel words into semantic memory via thematic relations. Abstract words were harder to acquire and recall, but both concrete and abstract words integrated into semantic memory, as indicated by behavioral and ERP measures in a lexical decision task. Chen et al. investigated how Mandarin Chinese native speakers and highly proficient second language (L2) learners from South Korea process complex sentences with center-embedded relative clauses. Results showed both groups exhibited a consistent biphasic ERP waveform pattern, indicating interactive syntactic and semantic processing in Chinese. This challenges the idea of temporal and functional priority in syntactic processing found in morphologically rich languages and highlights cross-linguistic consistency.

These studies collectively contribute to a better understanding of the challenges and factors influencing reading acquisition for Chinese as a second language. They highlight the role of early reading challenges, cross-linguistic transfer, metacognition monitoring, idiomatic expression processing, thematic relations and semantic memory, linguistic distance, code-switching costs, sentence structure processing, context predictability, and the impact of heritage language exposure in the development of reading skills among L2 learners of Chinese. The findings range from uncovering neural and cognitive mechanisms to informing language teaching practices, calling for further research in this field into the intricacies of language learning, cross-linguistic transfer, and neurocognitive processes in diverse linguistic contexts.

Author contributions

LZ: Writing—original draft. ZH: Writing—original draft. YZ: Writing—review and editing.

Funding

Preparation of this manuscript was supported by the Humanities and Social Sciences Foundation of the Chinese Ministry of Education (19YJAZH106).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Chan, J., Woore, R., Molway, L., and Mutton, T. (2022). Learning and teaching Chinese as a foreign language: a scoping review. *Rev. Educ.* 10, e3370. doi: 10.1002/rev3.3370

Chan, S. W. Y., Cheung, W. M., Huang, Y., Lam, W.-I., and Lin, C.-H. (2020). Development and validation of a Chinese character acquisition assessment for second-language kindergarteners. *Lang. Test.* 37, 215–234. doi: 10.1177/0265532219876527

Gong, Y., Lai, C., and Gao, X. (2020). The teaching and learning of chinese as a second or foreign language: the current situation and future directions. *Front. Educ. China* 15, 1–13. doi: 10.1007/s11516-020-0001-0

Ke, C. (2020). *The Routledge Handbook of Chinese Second Language Acquisition*. London: Routledge.

Kecskés, I., and Sun, C. (2017). *Key Issues in Chinese as a Second Language Research*. London: Routledge.

Li, L., Zhao, S., Zhang, L., and Wang, J. (2022). Understanding mechanisms of prediction error cost in Chinese reading for older adults. *Adv. Psychol. Sci.* 30, 1–14. doi: 10.3724/SP.J.1042.2022.00001

Wen, X. (2020). Acquisition of Chinese as a first and second language. *MDPI AG* 5:32. doi: 10.3390/languages5030032

Zhang, L., Han, Z., and Zhang, Y. (2023). Editorial: reading acquisition of Chinese as a second/foreign language. *Front. Psychol.* 14, 1185195. doi: 10.3389/fpsyg.2023.1185195

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fpsyg.2022.1083574

OPEN ACCESS

EDITED BY
Yang Zhang,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Yuxin Hao,
Huaqiao University,
China
Zhilong Xie,
Jiangxi Normal University,
China

*CORRESPONDENCE
Xingsan Chai
✉ cxs66@blcu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

RECEIVED 29 October 2022
ACCEPTED 19 December 2022
PUBLISHED 13 January 2023

CITATION
Chai X and Bao J (2023) Linguistic distances between native languages and Chinese influence acquisition of Chinese character, vocabulary, and grammar. *Front. Psychol.* 13:1083574.
doi: 10.3389/fpsyg.2022.1083574

Linguistic distances between native languages and Chinese influence acquisition of Chinese character, vocabulary, and grammar

Xingsan Chai*† and Jie Bao†

Institute on Educational Policy and Evaluation of International Students, Beijing Language and Culture University, Beijing, China

How linguistic distance affects second language acquisition is a major concern in cross-language transfer research. However, no study has explored how systematic differences between Chinese and learners' native language (L1) influences Chinese character, vocabulary, and grammar acquisition, or how these influences change as Chinese proficiency improves. To address this, we employed the World Atlas of Language Structures (WALS) index method to multidimensionally quantify the linguistic distance between Chinese and L1, and examined the effect of systematic linguistic distance on acquisition of Chinese character (Quasi-Experiment 1), vocabulary (Quasi-Experiment 2), and grammatical knowledge (Quasi-Experiment 3) in Chinese as a second language (CSL) learners with elementary, intermediate, and advanced Chinese proficiency levels. We examined a random sample of 58,240 CSL learners' test scores from 24 different L1 backgrounds, and analyzed 2,250 CSL learners' Chinese character, vocabulary, and grammar scores in each of the three quasi-experiments. We found that closer linguistic distance facilitated more favorable Chinese character and vocabulary acquisition at elementary, intermediate, and advanced Chinese proficiency levels, and that the influence of linguistic distance on CSL learners' vocabulary acquisition tended to decrease as Chinese proficiency increased. Finally, linguistic difference did not significantly affect CSL learners' grammar acquisition at elementary proficiency, but as Chinese proficiency improved, an L1 interference effect occurred among CSL learners with a short linguistic distance from Chinese, which hindered grammar acquisition. These results suggest that linguistic distance has differential proficiency-dependent effects on Chinese character, vocabulary, and grammar acquisition.

KEYWORDS

linguistic distance, Chinese proficiency, language knowledge, Chinese as a second language acquisition, cross-language transfer

1. Introduction

The role of native language (L1) in second language acquisition is a core issue in cross-language transfer research (Odlin, 1989; Ellis, 1994). Cross-language transfer is the influence of a language acquired or learned earlier on the new learning or acquisition of another language (Jarvis and Pavlenko, 2008). In a positive cross-language transfer, the acquired language promotes target language acquisition; in a negative cross-language transfer, the acquired language hinders target language acquisition (Odlin, 2001). Structuralist linguistics and behaviorist psychology assert that greater similarity between L1 and a target language facilitates target language acquisition, whereas greater differences are detrimental to target language acquisition (Lado, 1957). However, several studies have found that L1's influence on learners' second language acquisition is limited by several factors (Odlin, 1989; Odlin and Jarvis, 2004; Jarvis and Pavlenko, 2008), and that the extent to which L1 influences target language acquisition is not necessarily the same for learners with different target language proficiency levels (Sjoholm, 1995; Jarvis, 1998; Cenoz, 2001; Ringbom, 2007). Currently, there is no consensus on L1's role in second language acquisition.

Linguistic distance is an important indicator of L1 and second language similarity, and thus an important independent variable in cross-language transfer studies. Many previous studies on acquiring Chinese as a second language (CSL) examined linguistic distance effects on Chinese character (Jiang, 2003; Zhang, 2017; Deng and Hu, 2022), vocabulary (Hong, 2013; Xu, 2014; Hsieh and Wang, 2020; Chai and Ma, 2022; Tang and Chan, 2022), and grammar (Yuan, 1998, 2004; Wu and Zhao, 2018; Hao et al., 2022) acquisition. Chinese comprises several subsystems, such as Chinese character, vocabulary, and grammar, that interact across levels (De Saussure, 1959). However, previous studies focused on similarities and differences in specific features of learners' L1 and Chinese linguistics, and neglected to examine the effect of systematic differences between learners' L1 and Chinese on language knowledge acquisition (Hao et al., 2021). In addition, the validity of extending homogeneous sample-based laboratory studies' findings to heterogeneous environments is controversial (Cui et al., 2018; Floccia et al., 2018). Therefore, it is of considerable theoretical and practical importance to study linguistic distance effects on Chinese character, vocabulary, and grammar acquisition using a large sample in a non-laboratory setting.

Multidimensional methods for quantifying linguistic distance have developed through an in-depth intersection of linguistics, statistics, computer science, and other disciplines. Numerous studies have examined the impact of linguistic distance on CSL acquisition under heterogeneous conditions based on language test data. However, the findings concentrate on acquisition of CSL language skills at the macro level, such as speaking, reading, writing, and listening (Wang and Cui, 2018). Both language knowledge and language skills are important components of language competence (Bachman, 1990), and language knowledge is the basis for developing language skills. Ideally, higher learner language skills proficiency reflects higher language knowledge proficiency. In practice, however, language skills and language knowledge development are not perfectly synchronized, because second language acquisition is influenced by several factors, such as knowledge characteristics within the language subsystem, learning difficulty, and learners' second language proficiency (De Jong et al., 2013; Dai, 2014). Currently, no comprehensive empirical study of large-scale, standardized language test data has investigated the impact of systematic differences between L1 and Chinese on CSL learners' Chinese character, vocabulary, and grammar acquisition.

Considering the contested findings in theoretical and empirical studies on L1's role in second language acquisition, and that Chinese characters, vocabulary, and grammar differ from those of alphabetic languages, previous studies' findings are unlikely to generalize to the role of L1 in CSL learners' language knowledge acquisition.

Therefore, this study used the Hànyu Shuǐpíng Kǎoshì (HSK; literally translated as Chinese Proficiency Test) to measure CSL learners' language proficiency, and systematically investigated linguistic distance influence patterns on Chinese character, vocabulary, and grammar acquisition at elementary, intermediate, and advanced proficiency, using multidimensional linguistic distance quantification. This study provides systematic and persuasive evidence in relation to current theoretical debates on cross-language transfer, and sheds light on the study of CSL acquisition and teaching.

2. Literature review

2.1. Cross-language transfer theory

Cross-language transfer is a core issue in second language acquisition field (Klein et al., 1986; Gass and Selinker, 1992). Behaviorism, cognition, and social schools differ in both their theoretical claims and empirical research findings regarding whether and how L1 influences second language acquisition (Cook and Singleton, 2014).

Behaviorism asserts that L1 is the primary cause of learners' language acquisition difficulties and errors, with a greater difference between L1 and the target language leading to more difficulty learning the target language (Lado, 1957; Stockwell et al., 1965). However, the behaviorist viewpoint raises theoretical and empirical questions. From a theoretical perspective, many researchers argue that behaviorist theories of cross-language transfer ignore learners' subjectivity, while viewing language acquisition as a stimulus-reflection habit-forming process (Krashen, 1985; Swain and Lapkin, 1995; Long, 1996). From an empirical perspective, numerous research findings demonstrate that cross-language differences do not necessarily lead to second language acquisition difficulties (Grauberg, 1971; Dulay and Burt, 1973), or serve as a main reason for second language acquisition difficulties (Kleinmann, 1978; Zobl, 1983; Pica, 1994).

The cognitive school's two main branches, universal grammar theory and connectionist framework theory, emphasize the learner's role as a cognitive subject in language acquisition (Cook and Singleton, 2014). The universal grammar theory posits two distinct views on whether L1 affects second language acquisition. One holds that universal grammar covers the second language's initial state, and that L1 has no effect on second language acquisition (Bley-Vroman, 1990; Epstein, 1996; Platzack, 1996), while the other holds that L1 differs from the initial state of the second language, and that L1 affects second language acquisition (Schwartz and Sprouse, 1996; Vainikka and Young-Scholten, 1996). Both positions are supported by empirical studies. Different from the universal grammar theory, which emphasizes innate determinism, the connectionist framework theory holds that second language acquisition difficulty and learning speed largely depend on target language input frequency (Goldschneider and DeKeyser, 2001; Ellis, 2006), where L1 plays a role in regulating learners' absorption of second language input. If the second language's input information is similar to that of L1, a positive cross-language transfer occurs; if the second language's input information differs from that of L1, a negative cross-language transfer occurs because the second language's output is not what is expected. More importantly, at the elementary proficiency level of second language acquisition, learners usually draw on their L1 knowledge to compensate because their knowledge of the second language is not sufficient to allow them to fully express themselves. Therefore, learners at an elementary proficiency level are more likely to be affected by their L1, which also explains why second language acquisition development shows asymptotic and dynamic characteristics (Cook and Singleton, 2014).

Unlike the cognitive school, which views language as a psychological phenomenon, the sociocultural school focuses on both the influence of learners' elements on second language acquisition and on how social and cultural aspects affect second language acquisition (Vygotsky, 1978; Ellis, 2008; Green and Abutalebi, 2013; Tong and Yip, 2015). Vygotsky's sociocultural theory is influential in this school (Lantolf and Appel, 1994; Lantolf and Thorne, 2006), and posits that learners' L1 serves as a mediation tool that helps them achieve their communicative purposes when learning a second language. Numerous cross-language transfer studies based on sociocultural theory also show that L1's influence on learners' second language acquisition varies with their second language proficiency; when learners are not proficient enough to control cognitive activity in the second language, they rely more on their L1 to complete challenging mental tasks (Frawley and Lantolf, 1985; Swain and Lapkin, 1998; Centeno-Cortés and Jiménez Jiménez, 2004).

In summary, various schools of cross-language transfer theory agree that L1 plays a role in second language acquisition (MacWhinney and Kroll, 2005). Most scholars acknowledge that cross-language transfer is a complex and dynamic cognitive process that is influenced by a variety of factors, including linguistics, psychology, and society, rather than a mechanical habit-forming process based on stimulus–response (Xu Q. et al., 2013). Therefore, L1's role in second language acquisition needs to be examined from a systematic and developmental perspective (Larsen-Freeman, 1997; Larsen-Freeman and Cameron, 2008).

2.2. Specificity of Chinese characters, vocabulary, and grammar

Chinese characters, vocabulary, and grammar are important components of Chinese language knowledge in relation to CSL acquisition. Chinese is regarded as a difficult language to learn, which is related to both the peculiarities of the Chinese writing system, and the peculiarities of Chinese vocabulary and grammar.

Chinese characters are the written symbols of Chinese and the basis for Chinese reading and writing. CSL learners, especially those from the non-Sinosphere, often find Chinese characters difficult to learn because of fundamental differences between Chinese characters and alphabets regarding stereoscopic structure and ideographic nature (Everson, 1998; Shi and Wan, 1998). First, compared with alphabets' linear structure, Chinese characters have a square-shaped and more complex structure. Chinese character units comprise strokes and radicals, where strokes are the characters' smallest units, and radicals are the characters' secondary units, composed of strokes (Fei, 1996). Although there are few basic strokes in Chinese characters, the strokes have different deformations when in different positions in Chinese characters. For example, the stroke "丨" is "丨" in the Chinese character "快" (kùai, quick) and is "亅" in the Chinese character "水" (shuǐ, water). In combining radicals into Chinese characters, the size and direction of strokes in each position also change with the radical's position and the positional relationship between radicals. For example, the relative position and size of radical "口" in Chinese characters "扣"(kòu, button) and "器"(qì, vessel) differ. Second, Chinese characters belong to the morpheme-syllabic system, with a strong connection between form and meaning, but a poor connection with pronunciation (Tan et al., 2005). Chinese characters comprise four types: pictographic, ideographic, self-explanatory, and pictophonetic, which are all ideographic in nature. For example, the meaning of the pictographic character "刀" (daō, knife) is derived from the graphic " ," meaning knife; the meaning of the ideographic character "武"(wǔ, military) comprises the meaning of two Chinese characters "止"(zhǐ, stop) and "戈"(gē, dagger-ax), which means the cessation of war; the meaning of the self-explanatory character "刃"(rèn, blade) comprises the meaning of the pictographic character "刀" plus the indicator "丶," which means knife blade; the pictophonetic character "湖"(hú, lake) has the semantic radical "氵," indicating that the meaning of the character is related to water. Among the four types of Chinese characters, the pictophonetic character is the only one with a phonetic representation function (Li et al., 1992). However, the phonetic radical of the pictophonetic character can no longer accurately represent the pronunciation. For example, "触"(chù, touch) and "浊"(zhuó, muddy), which share the phonetic radical

"虫" (chóng, insect), are not pronounced similarly. Chinese characters' peculiarities make their acquisition tremendously difficult for CSL learners, especially for learners whose L1 is an alphabet script.

Vocabulary is the carrier of meaning, the key to understanding, and the basis of expression. Vocabulary knowledge includes both breadth knowledge and depth knowledge (Meara and Jones, 1988; Zareva et al., 2005; Moinzadeh and Moslehpour, 2012), where breadth knowledge refers to vocabulary size (Qian, 1999), and depth knowledge includes word aggregation relationships (such as synonymous relationships and context relationships) and combinatorial relationships (collocation relationships) (Nation, 1990). From a word aggregation perspective, first, vocabulary learning difficulties are reflected in differences in word concepts' cognition in different languages. For example, when the Chinese use the word "鱼" (yú, fish) as a metaphor for people, it usually means dishonest, but in Russian, it means silent, and in Czech, it means robust. Second, the degree of word concept refinement differs across languages. Mandarin has a large number of synonyms, many of which have very little difference in meaning. For example: "承继" (chéngjì) and "继承" (jìchéng), both mean subsequent possession, but the former focuses on forward continuation, and the latter on backward continuation. CSL learners often struggle to grasp such subtle differences between words (Zhang, 2019). From a word combination relationships perspective, collocation knowledge reflects syntactic, semantic, and usage frequency information in the mental lexicon (Xing, 2013). Therefore, acquiring collocation knowledge includes mastering both target language collocation grammar rules and collocation words' semantic category (Shi et al., 2021). However, each language has unique collocation rules, and many collocations are also based on various ethnic groups' social psychology language habits, which cannot be logically explained (Yamashita and Jiang, 2010). For example, "black tea" in English corresponds to "红茶"(hóngchá) rather than "黑茶"(hēichá) in Chinese. This difference in collocations across languages poses another difficulty in CSL learners' vocabulary acquisition.

Grammar is the organizational rule of a language; an important sign of second language mastery is mastering its grammar (Li, 2016). Chinese grammar is difficult to learn because, first, unlike Indo-European languages that use a rich variety of morphology to express grammatical relationships, Chinese is an isolated language and lacks strict morphology, so word order and function word are important means of expressing grammatical relationships (Lv, 1979). The grammatical meaning often differs according to the word order. For example, in "我看" (wǒkàn, I see) and "看我" (kànwǒ, see me), the former indicates a subject-predicate relationship, while the latter indicates a verb-object relationship. The choice of whether to use function words, and the use of different function words, often indicates different grammatical relationships. For example, "买书"(mǎishū) is a verb-object relationship, while "买的书"(mǎideshū) is a subordinate relationship. Second, unlike in English and many other languages, Chinese has no one-to-one correspondence between word class and syntactic constituents (Zhu, 1985)—a word class can act as multiple syntactic constituents, and a syntactic constituent can also be acted on by multiple word classes. Chinese also has some syntactic peculiarities that set it apart from other subject-verb-object languages, such as prepositional phrases followed by verbs, and relative clauses placed before the head. Additionally, Chinese contains unique syntactic constructions, such as pivotal and ba-structure sentences. Therefore, Chinese grammar acquisition difficulties may vary for CSL learners with different L1 backgrounds, owing to these distinctive aspects of Chinese grammar.

However, although Chinese characters, vocabulary, and grammar have distinct characteristics, this does not mean that they are unrelated. From the perspective of Chinese language research, Chinese characters belong to the morpheme-syllabic script, meaning that Chinese characters are not only syllables that represent pronunciation, but also words or morphemes that represent meanings (Li, 2009). The integration of Chinese characters' form, pronunciation, and meaning has exerted great influence on Chinese words' syllable form, formation, and meaning composition. Additionally, as Chinese word construction is similar to phrase or sentence construction, syntactic structures and words share a selective restriction relationship (Shi and Yang, 2021); therefore, Chinese character, vocabulary, and grammar characteristics often influence each other. In addition, evidence from many empirical studies shows that knowledge of Chinese characters' sublexical and grammatical features is also activated during Chinese character processing (Tsai et al., 2004; Yan et al., 2009, 2012; Tsang et al., 2017; Yeh et al., 2017; Pan et al., 2019), which indicates that CSL learners learn Chinese characters, vocabulary, and grammar simultaneously.

In sum, language systems interact across levels, and Chinese character, vocabulary, and grammar characteristics and their acquisition often influence each other. Only when CSL learners' Chinese characters, vocabulary, and grammar are all well-developed can Chinese proficiency improve. Therefore, these three elements should not be separated when examining CSL learners' language knowledge acquisition.

2.3. Linguistic distance and CSL acquisition

2.3.1. Qualitative linguistic distance and CSL acquisition

Linguistic distance refers to the degree of actual difference between languages and is an important independent variable in the study of cross-language transfer, expressed through intra-linguistic factors, such as phonology, vocabulary, syntax, and writing forms (Ellis, 1994; Chiswick and Miller, 2005). Linguistic distance measures include qualitative and quantitative methods. Studies on CSL acquisition have primarily used qualitative methods, such as genealogical classification and linguistic typological classifications, to examine the influence of linguistic distance on Chinese character, vocabulary, and grammar acquisition.

AS Chinese characters are unique to Chinese, previous studies have compared performance of CSL learners from the Sinosphere and the non-Sinosphere to explore the impact of linguistic distance on Chinese character acquisition. First, linguistic distance has an impact on how quickly CSL learners acquire orthographic awareness of Chinese characters. Several studies have indicated that CSL learners from the Sinosphere develop orthographic awareness more quickly than other CSL learners (Lu, 2002; Feng, 2006; Liu, 2013; Zhang, 2016; Loh et al., 2018). Second, linguistic distance affects CSL learners' strategies for recognizing Chinese characters. CSL learners from the Sinosphere tend to memorize Chinese characters using their form, and process them using conformational structures; whereas CSL learners from the non-Sinosphere are more likely to be influenced by their L1, relying on phonological strategies to recognize Chinese characters, and processing them using strokes or radicals (Jiang, 2003; Yeh et al., 2003; Feng et al., 2005; Ke and Chan, 2017; Jiang et al., 2020). Third, evidence shows that CSL learners' L1 orthographic characteristics affect their Chinese character writing, where CSL learners with complex visual space L1 scripts similar to Chinese characters perform better than CSL learners with linear L1 scripts (Lin and Collins, 2012; Zhang and Roberts, 2021). In addition, many studies have compared Chinese character writing and reading performance in CSL learners from different L1 backgrounds and found that Chinese character writing is more difficult than reading. As CSL learners' Chinese proficiency improves, linguistic distance has fewer effects on Chinese character reading performance, but still has a significant effect on Chinese character writing (Jiang, 2000; Wu et al., 2006; Li et al., 2014; Zhang and Roberts, 2021).

Studies on Chinese vocabulary acquisition level have examined the impact of linguistic distance on the Chinese vocabulary performance of CSL learners from the Sinosphere and non-Sinosphere. Studies on acquiring vocabulary breadth knowledge show that CSL learners' vocabulary size increases as their Chinese proficiency improves, but at elementary proficiency, CSL learners from the Sinosphere master significantly more vocabulary than CSL learners from the non-Sinosphere (Zhang, 2006; Luo and Duan, 2019). Studies of acquiring vocabulary depth knowledge have generated some controversy regarding vocabulary semantic acquisition. Zhang et al. (2011) compared word semantic acquisition performance of polysemous words between CSL learners from the Sinosphere and the non-Sinosphere at elementary, intermediate, and advanced Chinese proficiency levels and found no significant performance differences in CSL learners from different L1 backgrounds at any Chinese proficiency level. The authors concluded that linguistic distance had no significant effect on CSL learners' polysemous word acquisition. By contrast, Hong and Chen (2011) found that CSL learners from both the Sinosphere and the non-Sinosphere relied on their L1 to establish L2 synonym semantic relations at elementary Chinese proficiency. Only at intermediate Chinese proficiency could CSL learners significantly acquire the ability to differentiate in lexical semantics. Wang and Hao (2014) reported similar results, finding that CSL learners first acquire knowledge of shared lexical items in both languages at elementary proficiency, and then begin to recognize target language-specific items at intermediate proficiency. The L1-specific items begin to interfere with developing a bilingual mental lexical; only at advanced proficiency do CSL learners gradually abandon the L1-specific items and reach a proficiency close to that of native speakers. Studies on vocabulary collocation knowledge acquisition show relatively consistent results. As Chinese proficiency improves, CSL learners' vocabulary collocation competence gradually improves (Cai, 2017; Chang, 2020; Shi et al., 2021); even at advanced Chinese proficiency levels, it is influenced by CSL learners' L1 characteristics, indicating that collocation knowledge with similar L1 and Chinese characteristics is easy to acquire, but that greater difference hinders acquisition (Cai, 2017; Luo and Duan, 2019; Chang, 2020).

Grammar acquisition studies have reported contradictory findings. Some found that similarities between CSL learners' L1 and Chinese grammatical features promote Chinese grammar acquisition. Guo and Liu (2017) investigated Chinese word order error statistics among CSL learners with intermediate and advanced Chinese proficiency levels, isolated language, agglutinative language, and inflected language L1 backgrounds, and found that the isolated language error rates were lowest with L1 more similar to Chinese, while the inflected language error rate was highest for L1s that differed most from Chinese, indicating that higher similarities between L1 and Chinese facilitate acquisition. Hu et al. (2018) examined the written production of topic-comment constructions by elementary and advanced CSL learners from English and Japanese L1 backgrounds and found that Japanese CSL learners had higher usage rates than English CSL learners at either proficiency level, because Japanese and Chinese are topic-salient languages. Zhang (2021) examined Chinese ellipsis object sentence acquisition by CSL learners with Korean and English L1 backgrounds and also found that CSL learners from English backgrounds with similar characteristics to Chinese showed better acquisition. These findings show that cross-language transfer plays a role in grammar acquisition.

By contrast, some studies found that greater differences between L1 and Chinese better facilitate acquisition. Yuan (2004) examined Chinese negative sentences acquisition in CSL learners from native German, French, and English L1 backgrounds with different learning durations, and found no significant differences in acquisition performance at any proficiency level between CSL learners from English backgrounds whose L1 had similar negation structures to Chinese, or CSL learners from French and German backgrounds whose negation structures were different. The author suggested that this might be because German and French negative structure is quite different from that of Chinese, so it attracts learners' attention at the beginning, and the difference is constantly strengthened in the process of learning, thus promoting CSL acquisition in learners with German and French backgrounds. Similarly, Wu and Zhao (2018) examined collocation acquisition of Chinese negation markers "不" and "没" with aspect markers "着," "了," and "过" by CSL learners from intermediate and advanced English, and Korean backgrounds, and also found that CSL learners from native English backgrounds notice the difference between the collocation of "不" and "没" because English lacks the two negative oppositional markers, thus facilitating acquisition. However, Yuan and Zhao (2005) examined acquisition of resumptive pronouns in Chinese relative clauses by CSL learners with English and Arabic backgrounds, and found that, despite that the use of resumptive pronouns in relative clauses is allowed in Arabic, learners with Arabic backgrounds did not show significantly higher accuracy in judgment tasks than learners with English backgrounds. The authors suggest that learners with Arabic L1 perceive a greater psycho-typological distance between Chinese and Arabic, which hinders positive transfer. Wu and Mo (2018) examined the use of ba-structure sentence among Danish and Korean CSL learners through grammar judgment and picture description tasks and found that, despite that Korean has object prepositions while Danish does not, learners with Danish backgrounds frequently used the ba-structure sentence and were more confident in their understanding of it. According to the retrospective interview, Korean and Chinese object prepositions share some characteristics, but also differ, so learners tended to employ avoidance strategies to lessen usage errors. Furthermore, several studies revealed linguistic distance's effect on advanced CSL learners' implicit grammatical processing from an electrophysiological perspective. Hao et al. (2022) used event-related potentials to investigate how linguistic distance affected advanced CSL learners from Indonesia and Thailand acquire Chinese "aspect" and discovered similar EEG patterns evoked by the two learner types for processing aspect violation sentences, but noted that even advanced CSL learners did not reach native speakers' automatic processing level. These studies hold that psychological typology, learning strategies, and other factors weaken the impact of L1 negative transfer, and they do not entirely deny the role of cross-language transfer.

Linguistic distance's influence on Chinese character, vocabulary, and grammar acquisition differs by nature and degree. The inconsistent study findings may be related to sampling differences across studies, different L1 backgrounds across studies, and studies being limited to only one language knowledge type of Chinese character, vocabulary, and grammar knowledge. Therefore, analyzing data in relation to CSL learners with larger sample sizes and richer L1 backgrounds is valuable in that it facilitates a fuller understanding of linguistic distance's influence on CSL knowledge acquisition.

2.3.2. Quantitative linguistic distance and CSL acquisition

Qualitative linguistic distance methods identify differences between languages, but cannot identify the magnitude of the differences (McCloskey, 1998; Chiswick and Miller, 2005); therefore, qualitative methods are significantly limited in comparative studies of learners from several different L1 backgrounds acquiring the same target language. Conversely, quantitative methods determine the numerical magnitude of differences between languages, which facilitates comparing similarities between L1 and target languages and provides a feasible method for calculating linguistic distance. Five quantitative methods for measuring linguistic distance include the dummy variable method, cognate method, test assessment method, automated similarity judgment program (ASJP) edit distance method, and the World Atlas of Language Structures (WALS) index method (Wang and Yang, 2019). The dummy variable method, similar to a qualitative method, dichotomizes linguistic distances and is thus rarely used (Ginsburgh and Weber, 2014). The cognate method calculates the proportion of cognates between languages based on a core word list, but is only applicable to studies between languages within the Indo-European family (Dyen et al., 1992; Schepens et al., 2016). The test assessment method uses language test scores as a linguistic distance measure (Hart-Gonzalez and Lindemann, 1993; Chiswick and Miller, 2005), but is controversial because the results are affected by test reliability, validity, and examinees' motivation (Van der Slik, 2010). The ASJP edit distance method is based on phonological differences in synonyms or near-synonyms between languages, where fewer conversions indicate a closer linguistic distance (Isphording and Otten, 2013). However, this method calculates the phonological distance between languages, making it more suitable for studies of dialects, languages in which pronunciation features are the main difference, or studies that focus on listening and speaking skills (Isphording and Otten, 2013; Cui et al., 2018). The WALS index method is presently the only method that calculates linguistic distance based on many aspects of language differences, including 192 linguistic features contained in the WALS online[1] database, which covers 11 categories: phonology, morphology, noun category, noun syntax, verb category, word order, simple sentence, complex sentence, vocabulary, sign language, and others. Therefore, the WALS index method accurately reflects and measures real differences between major languages. It has been used in many studies that compared Chinese with other languages (Bakker et al., 2009; Lupyan and Dale, 2010; Moran and Blasi, 2014; Wang and Cui, 2018; Wang and Yang, 2019; Xu and Zi, 2020).

Studies that quantify linguistic distance to identify second language acquisition influence have primarily been conducted in the language of economics field, with results that reported macro-level effects of L1 to Chinese linguistic distance on language skill acquisition, including listening, speaking, reading, and writing, and consistently reported that shorter linguistic distances between CSL learners' L1 and Chinese are associated with higher language skill proficiency (Isphording and Otten, 2013; Lindgren and Muñoz, 2013; Schepens et al., 2016; Cui et al., 2018). Language competence includes both language knowledge and language

1 https://wals.info

skills, with language knowledge being the basis for language skill development, but evidence suggests that developing language knowledge is not necessarily synchronized with developing language skills. Ding and Xiao (2016) tracked oral expressive skills' development in Italian CSL learners and found that not all dimensions of vocabulary knowledge were developed as oral proficiency improved. Additionally, Wu (2017) found that as CSL learners' Chinese proficiency improved, vocabulary and grammar development was not synchronized, and even showed a competitive relationship. Other studies of foreign target language acquisition have reported similar findings (Spoelman and Verspoor, 2010; Lowie et al., 2017). Moreover, the influence of L1 on language knowledge and language skills is not always the same. A meta-analysis (Jeon and Yamashita, 2014) showed that linguistic distance had a significant effect on reading comprehension, but no moderating effect on vocabulary or grammar.

Given this context, it is not possible to simply generalize the findings of studies examining macro-level language skills to the findings of studies examining micro-level language knowledge. Therefore, it is worthwhile to investigate the effect of linguistic distance on CSL acquisition by applying a quantitative method for determining linguistic distance.

3. The current study

The literature review showed that a large body of research has examined the influence of linguistic distance on Chinese character, vocabulary, and grammar knowledge acquisition among CSL learners. However, these studies limited the linguistic distance measurement to a comparison of specific features and did not systematically compare differences across languages as a whole. Nor did they systematically investigate linguistic distance's influence on the acquisition of different language knowledge. Although they applied quantitative methods to calculate linguistic distance and examined the impact of this factor on CSL learners' acquisition of Chinese based on language tests, they focused only on language skill levels.

Considering the connection between Chinese characters, vocabulary, and grammar, and the complex and dynamic characteristics of language acquisition, the current study used the WALS index method to calculate linguistic distance between CSL learners' L1 and Chinese, and used HSK test data to systematically investigate the influence of linguistic distance on the acquisition of Chinese characters, vocabulary, and grammar at different proficiency levels. The findings provide systematic empirical support for the role of differences between Chinese and L1 on Chinese language knowledge acquisition.

We conducted three quasi-experiments. Quasi-Experiment 1 examined whether and how linguistic distance affects Chinese character acquisition in CSL learners at elementary, intermediate, and advanced Chinese proficiency levels. Quasi-Experiment 2 examined whether and how linguistic distance affects Chinese vocabulary acquisition in CSL learners at elementary, intermediate, and advanced Chinese proficiency levels. Quasi-Experiment 3 examined whether and how linguistic distance affects Chinese grammar acquisition in CSL learners at elementary, intermediate, and advanced Chinese proficiency levels.

4. Quasi-experiment 1

4.1. Methods

4.1.1. Participants

Data comprised a subset of a large database (gathered in 2009) that contains information on 80,506 examinees who participated in HSK tests in different regions of China. A brief and non-mandatory questionnaire collected personal background information during online HSK registration, which included basic demographic characteristics, such as gender, age, native language, and place of birth.

We excluded participants with invalid data, including those whose L1 was Chinese or a dialect, or who misfiled or omitted information, resulting in valid data from 58,240 examinees. The examinees spoke 24 L1s ($M = 2,426.67$ speakers per language) that, according to the WALS (Dryer and Haspelmath, 2013), belong to 11 language families (i.e., Afro-Asiatic, Altaic, Austro-Asiatic, Indo-European, Japanese, Korean, Kartvelian, Niger-Congo, Sino-Tibetan, Tai-Kadai, and Uralic).

From each group with elementary, intermediate, and advanced Chinese proficiency, we extracted for inclusion in the analysis 250 CSL learners with a short distance between their L1 and Chinese, 250 with a middle distance, and 250 with a long distance, for a total of 2,250 CSL learners for analysis, owing to the uneven database distribution of CSL learners at different Chinese proficiency levels and different linguistic distance levels (1,121 females; age range 9.70 to 61.80, mean age = 23.44 years, standard deviation = 5.96).

4.1.2. Instruments

The HSK is a national standardized test designed to measure Chinese proficiency in non-native speakers, including international students, overseas Chinese students, and students from ethnic minorities in China. The HSK test score is required for undergraduate or graduate admission to Chinese universities. It also serves as a crucial basis for some business organizations and multinational corporations in China to assess CSL learners' Chinese communication skills when hiring employees (Meyer, 2014; Teng, 2017).

HSK development has had three stages. HSK Version 1.0 was developed and implemented by Beijing Language and Culture University (BLCU) in 1984. In 2007, BLCU released HSK Version 2.0 to better serve global promotion of the Chinese language. Since 2010, the Confucius Institute Headquarters has improved and perfected the test system and released HSK Version 3.0 (Hanban, 2011).

The study's HSK data were collected from 78 testing centers in China in 2009. We selected these test data because BLCU has accumulated 26 years of theory and experience on this test (1984–2010), and researchers have conducted sufficient empirical studies on its reliability, construct validity, and test score equivalence to determine that the test results are a reliable indicator of CSL learners' Chinese proficiency (Chai, 2006; Chen, 2006).

To meet different CSL learner groups' measurement needs, the designers developed three independent tests: the Elementary HSK, the Intermediate HSK, and the Advanced HSK. The Elementary HSK is only suitable for assessing beginners' Chinese proficiency; the Advanced HSK assesses CSL learners whose Chinese proficiency is close to that of native Chinese speakers; therefore, the number of examinees for these tests is limited. The Intermediate HSK measures the widest range of Chinese language proficiencies and has the largest number of examinees, so it provides a higher degree of Chinese language proficiency differentiation among CSL learners.

The Intermediate HSK comprises four sections: listening comprehension, grammar structure, reading comprehension, and comprehensive completion, which assess knowledge of Chinese characters, vocabulary, grammar, and listening and reading skills through nine subtests. The Intermediate HSK has a total of 170 points, including 16 for Chinese characters, 20 for vocabulary, and 30 for grammar. Quasi-Experiment 1 focused on the Chinese character subtests.

4.1.3. Variables

In Quasi-Experiment 1, the HSK Chinese character test score was the dependent variable. Two independent between-subjects variables were the linguistic distance between L1 and Chinese, and Chinese proficiency.

4.1.3.1. Chinese character test scores

The HSK Chinese character subtest measures orthographic competence by examining CSL learners' Chinese character writing accuracy under contextual constraints. A higher score on the HSK Chinese character subtest indicates better orthographic proficiency. The Chinese character subtest took 15 min to complete. Figure 1 shows a sample HSK Chinese character subtest.

4.1.3.2. Chinese proficiency

Chinese proficiency was grouped by total HSK score minus the Chinese character test score, to exclude the interference of Chinese character proficiency in the study results. There was a significant positive correlation between total HSK and total HSK minus the Chinese character test scores ($\rho = 0.995$, $p < 0.001$), indicating that excluding the Chinese character scores from the total HSK scores did not change the essential characteristics of the data sample structure.

Therefore, we first ranked the scores from highest to lowest, and then operationally defined the lowest one-third and highest one-third of the examinees as having elementary and advanced Chinese proficiency, respectively, and the middle one-third as having intermediate proficiency. According to this standard, we classified examinees in the full database with scores below 61 as having elementary Chinese proficiency ($n = 9{,}493$); those with scores between 62 and 107 as having intermediate Chinese proficiency ($n = 33{,}093$); and those with scores above 108 as having advanced Chinese proficiency ($n = 15{,}654$).

4.1.3.3. Linguistic distance

We computed linguistic distance between participants' L1 and Chinese based on the linguistic structural features contained in the online WALS. First, based on 153 Chinese features, we compared feature similarities and differences between CSL learners' L1 and Chinese. Common features of both languages were assigned 0, and different features were assigned 1. We then calculated the average value of the assignment results of each L1, with lower linguistic similarity between L1 and Chinese indicating a higher WALS value. The linguistic distance values ranged from 0.33 to 0.64 ($M = 0.49$, $SD = 0.08$). Based on the linguistic distance value ranges, we divided the 24 L1s into three equal groups; the short-distance group included participants with linguistic distance values below 0.44, the middle-distance group had values between 0.44 and 0.54, and the long-distance group had distance values above 0.54. Table 1 shows the descriptive statistics for each linguistic distance level group in the overall database sample.

4.1.4. Data analysis

We used a two-way between-subjects ANOVA to test for a main effect of linguistic distance on Chinese character acquisition and an interaction effect between linguistic distance and Chinese proficiency. It should be noted that several authors have argued that violation of normality is not a serious problem (Sokal and Rohlf, 1995, p. 407; Zar, 2010, p. 137) in terms of the central limit theory. Some authors even argue that the normality assumption is not needed with adequately large samples (Fitzmaurice et al., 2012; Ghasemi and Zahediasl, 2012). Therefore, we used two-way between-subjects ANOVA for data analysis, even if the data did not meet the normality assumption.

We used the Sidak method to correct for significance levels when performing *post hoc* multiple comparisons of linguistic distance main effects, and simple effects tests for the interaction between linguistic distance and Chinese proficiency. We calculated effect sizes using partial eta squared ($\eta 2p$) and classified effect sizes as very small (0–0.02); small (0.02–0.15); moderate (0.15–0.35); and large (0.35–1.0; Cohen, 1992). We considered two-tailed probability values <0.05 statistically significant.

All analyzes were performed using IBM SPSS Statistics version 26.0; data visualization was performed using the R statistical programming language.

4.2. Results

To examine whether and how linguistic distance affects Chinese character acquisition by elementary, intermediate, and advanced proficiency CSL learners, we analyzed Chinese character test scores from 2,250 CSL learners in nine groups. Table 2 shows the Chinese character test score descriptive statistics for the nine CSL learner groups.

4 · 4 · 4 · 4 · 4 · 4 · 4

第 二 部 分

说明：155－170 题，每段话中都有若干个空儿（空儿中标有题目序号）。请根据上下文的意思**在答卷上**的每一个空格中填写一个恰当的汉字。

155－157

尊敬的陈青先生、各位朋友与来宾：

记得在两周之前，我们在这里欢聚一堂，热烈欢迎陈青先生。今天，在陈先生访问即将结束之时，我们再次聚在这里，感到 **155** 别亲切和高兴。

陈先生是我们 **156** 悉的老朋友，我们之间有着十分深厚的友谊和长期的合作关系。临别在即，除了表示热烈欢送外，我们祝陈先生一路 **157** 安、身体健康，并诚挚地欢迎陈先生有机会再来！

谢谢！

158－161

尊敬的领导：

您好！我是一名刚刚从湖北经济学院毕业的本科生，我是学习酒店管理的，我热爱我的 **158** 业。在四年的学习生活中，我所学习的 **159** 容包括了书本知识和实际技能。

经过这四年的学习，我对这一领域的知识有了一定程度的理解和掌握，并取得了优异的成绩，我还参加了很多社会实践活动。

我真诚地希望能够到贵酒店实 **160** 我的梦想，能够为贵酒店奉献我的一份力 **161**。同时也希望能够和贵酒店的高层人员团结合作，和大家一起努力使酒店以后的事业蒸蒸日上。如果我有幸加入到贵酒店，我一定会努力做好我的工作。

希望各位领导能够对我予以考虑，我热切期盼你们的回信，谢谢！

此致

敬礼！

谢波

162－170

尊敬的领导：

您好！请允 **162** 我向您介 **163** 自己。

我叫王冲，毕业于四川省信息工程学院计算机系。我尊敬老师，团结同学，有着强烈的集体责 **164** 感。通过四年的努力学习，我取得了一定的成绩，并获得了多种技能证书。在假期中，我认真参加了社会实践，学到了在书上或课堂上学不到的知识，**165** 然时间很短，但体会很深，无 **166** 是业务能力，还是社交能力，都有一定的 **167** 高，具备了一定的工作经验。请您 **168** 信我，给我一个发 **169** 的机会，我会以一颗真诚善良的心、饱满的工作热 **170**、勤奋务实的工作作风、突出的工作业绩回报贵单位。期待您的回复。

最后，衷心祝愿贵单位事业发达、蒸蒸日上！

此致

敬礼！

求职人：王冲

M12TB01X

（完）

20

FIGURE 1
Sample test in the HSK character subtest.

Two-way between-subjects ANOVAs revealed a significant main effect of linguistic distance ($F_{(2, 2,241)} = 51.69$, $p < 0.001$, $\eta_p^2 = 0.044$) and Chinese proficiency ($F_{(2, 2,241)} = 1102.23$, $p < 0.001$, $\eta_p^2 = 0.496$), with a significant interaction effect ($F_{(4, 2,241)} = 2.86$, $p = 0.02$, $\eta_p^2 = 0.005$). To further explore specific differences between the groups, we conducted a simple effects test, which showed that at the elementary Chinese proficiency level, the short-distance group Chinese character scores were significantly higher than those of the

TABLE 1 Participant descriptive statistics with the linguistic distance between L1s and Chinese.

Linguistic distance levels	L1	Values	*N*
Short	Vietnamese	0.33	3,888
	Tai	0.34	3,468
	Korean	0.35	29,012
	Indonesian	0.40	1,837
	Yoruba	0.42	2
	Japanese	0.43	10,966
Middle	Burmese	0.45	159
	English	0.47	1,889
	Hausa	0.49	2
	Fijian	0.49	3
	Tagalog	0.50	42
	Russian	0.50	3,742
	Finnish	0.51	43
	Spanish	0.52	515
	Hungarian	0.52	34
	Hindi	0.53	230
	Hebrew	0.53	34
	Turkish	0.53	486
Long	Greek	0.56	16
	Persian	0.56	37
	German	0.57	668
	Swahili	0.60	33
	French	0.62	1,123
	Georgian	0.64	11

middle-distance group (mean difference = 0.69, $p = 0.02$) and long-distance group (mean difference = 0.72, $p = 0.01$). Although the middle-distance group's Chinese character scores were higher than those of the long-distance group, the difference was not significant (mean difference = 0.024, $p = 1.00$). At the intermediate Chinese proficiency level, the short-distance group's Chinese character scores were significantly higher than those of the middle-distance group (mean difference = 1.36, $p < 0.001$); the short-distance group scores were also significantly higher than those of the long-distance group (mean difference = 1.80, $p < 0.001$). Although the middle-distance group's performance was higher than that of the long-distance group, the difference was not significant (mean difference = 0.44, $p = 0.23$). At the advanced Chinese proficiency level, the short-distance group's Chinese character scores were significantly higher than those in the middle-distance group (mean difference = 1.43, $p < 0.001$); the short-distance group's Chinese character scores were also significantly higher than those in the long-distance group (mean difference = 1.64, $p < 0.001$), but there was no significant difference between the middle-distance group and the long-distance group (mean difference = 0.21, $p = 0.79$; see Figure 2).

TABLE 2 Chinese character test score descriptive statistics.

Chinese proficiency level	Linguistic distance level	Mean values of Chinese characters
Elementary	Short	3.52 (SD = 2.19)
	Middle	2.83 (SD = 2.11)
	Long	2.80 (SD = 1.80)
Intermediate	Short	7.37 (SD = 3.07)
	Middle	6.01 (SD = 3.11)
	Long	5.58 (SD = 2.80)
Advanced	Short	10.93 (SD = 2.83)
	Middle	9.50 (SD = 3.74)
	Long	9.29 (SD = 3.26)

The findings from Quasi-Experiment 1 showed how linguistic distance affected CSL learners' Chinese character acquisition, where performance differed across the elementary, intermediate, and advanced Chinese proficiency conditions, but the short-distance group's Chinese character performance was noticeably better than that of the middle-and long-distance groups. Therefore, differences in linguistic distance's influence on CSL learners' Chinese character acquisition predominately manifested between the short-and middle-distance groups, and between the short-and long-distance groups.

5. Quasi-experiment 2

5.1. Methods

5.1.1. Participants

Quasi-Experiment 2 data came from the same database as Quasi-Experiment 1. For comparable results between the two experiments, we extracted 250 CSL learners with a short distance between L1 and Chinese, 250 with a middle distance, and 250 with a long distance from each of the elementary, intermediate, and advanced Chinese proficiency groups, for a total of 2,250 individual HSK vocabulary scores for analysis (1,161 females; age range 9.90 to 67.4 years, mean age = 23.33, standard deviation = 5.85).

5.1.2. Variables

The Quasi-Experiment 2 dependent variable was HSK vocabulary test scores. The two independent between-subject variables were the linguistic distance between L1 and Chinese, and Chinese proficiency.

5.1.2.1. Vocabulary test scores

Vocabulary scores were derived from the HSK vocabulary subtests, which measure vocabulary knowledge depth and breadth, with higher HSK vocabulary scores indicating higher vocabulary proficiency. The vocabulary subtest took 20 min to complete. A sample of the test is shown in Figure 3.

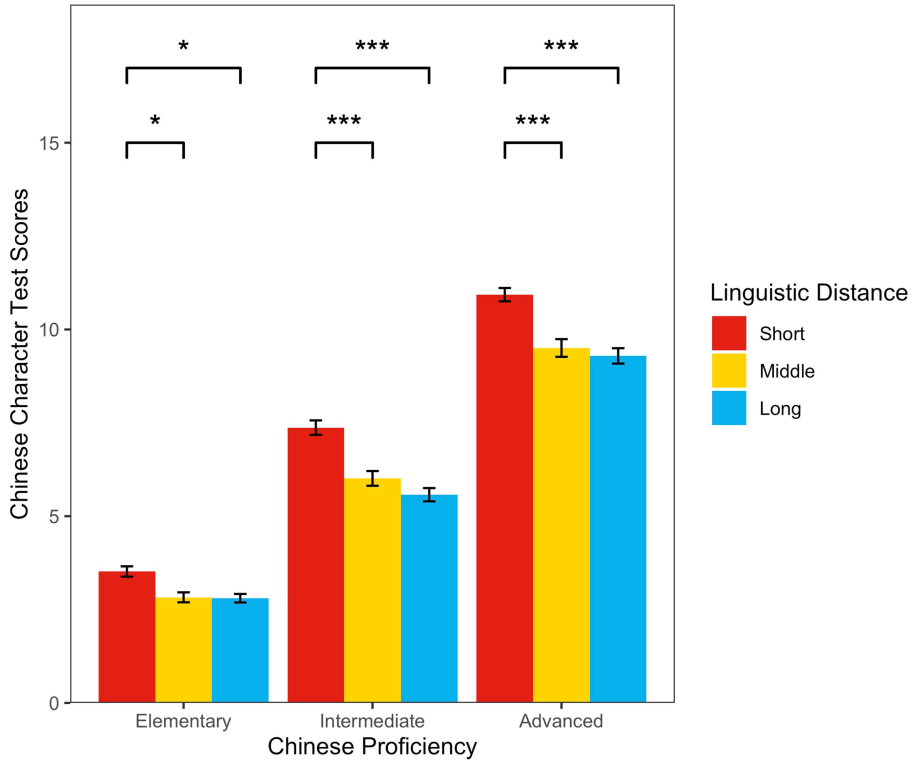


FIGURE 2
The relationship between Chinese proficiency, linguistic distance, and Chinese character test scores.

5.1.2.2. Chinese proficiency

Similar to the approach used in Quasi-Experiment 1, Quasi-Experiment 2 divided Chinese proficiency into three groups according to the total HSK scores minus the vocabulary scores. We found a significant positive correlation between total HSK score and total HSK score minus the vocabulary score ($\rho = 0.996$, $p < 0.001$). Finally, in the full database, we classified scores below 61 as elementary Chinese proficiency ($n = 9,404$); scores between 62 and 107 as intermediate Chinese proficiency ($n = 34,041$); and scores above 108 as advanced Chinese proficiency ($n = 14,795$).

5.1.2.3. Linguistic distance

The calculation and grouping methods for linguistic distance were the same as in Quasi-Experiment 1.

5.1.3. Data analysis

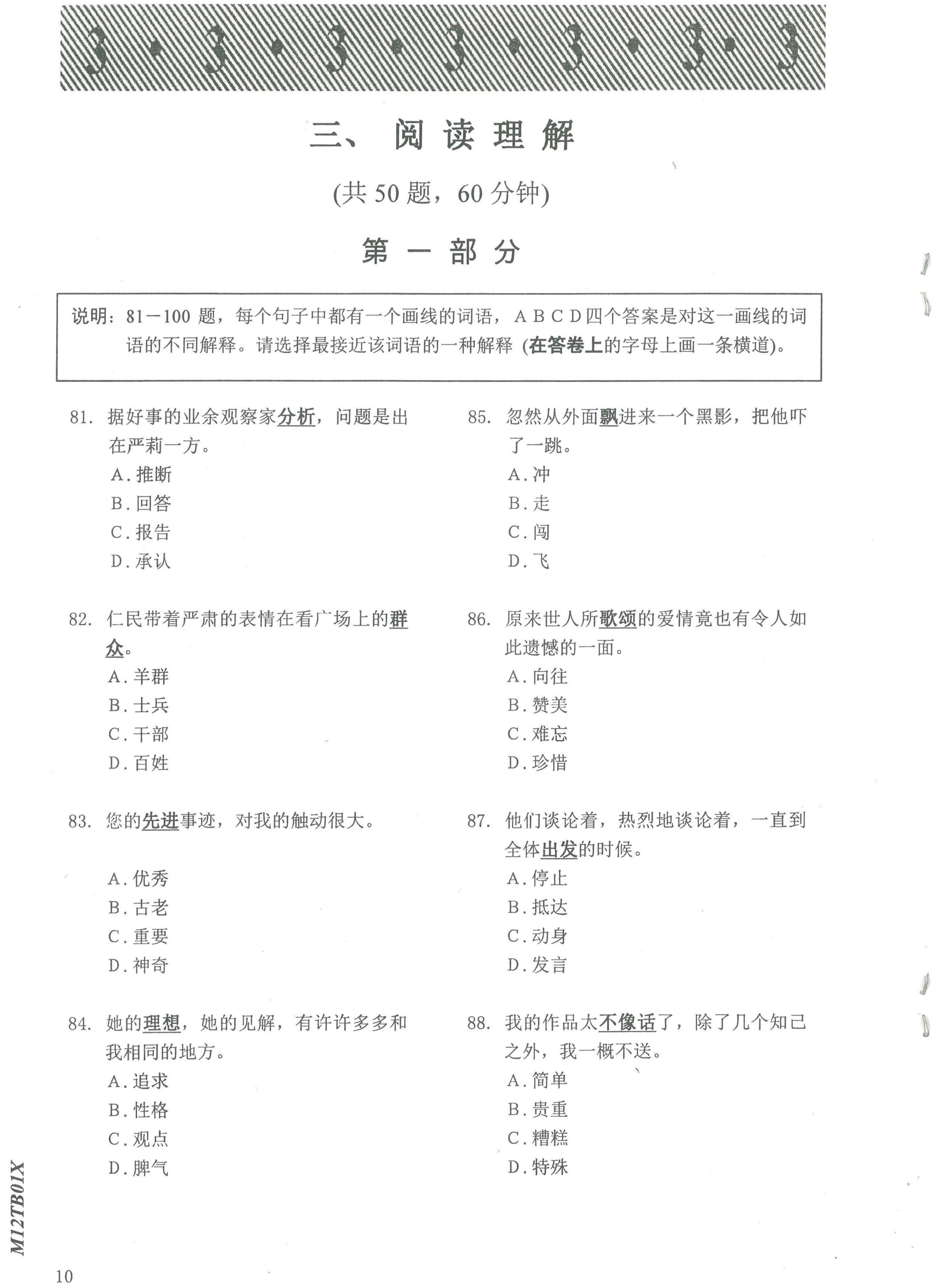
Data analysis methods and procedures were the same as in Quasi-Experiment 1.

5.2. Results

To examine whether and how linguistic distance affects Chinese vocabulary acquisition by elementary, intermediate, and advanced proficiency CSL learners, we analyzed HSK vocabulary test scores from 2,250 CSL learners in nine groups. Table 3 shows Chinese vocabulary test score descriptive statistics for the nine CSL learner groups.

Two-way between-subjects ANOVAs revealed a significant main effect of linguistic distance ($F_{(2\ ,\ 2,241)} = 43.41$, $p < 0.001$, $\eta_p^2 = 0.037$) and Chinese proficiency ($F_{(2\ ,\ 2,241)} = 2135.19$, $p < 0.001$, $\eta_p^2 = 0.656$), with a significant interaction effect ($F_{(4\ ,\ 2,241)} = 2.86$, $p = 0.02$, $\eta_p^2 = 0.005$).

A simple effect test showed that for elementary Chinese proficiency, the short-distance group's vocabulary scores were significantly higher than those of the middle-distance group (mean difference = 1.13, $p < 0.001$), and those of the long-distance group (mean difference = 1.38, $p < 0.001$). However, there was no significant difference between the middle-distance and long-distance groups (mean difference = 0.25, $p = 0.67$). For intermediate Chinese proficiency, the short-distance group's vocabulary scores were significantly higher than those of the middle-distance (mean difference = 1.05, $p < 0.001$), and long-distanc (mean difference = 1.62, $p < 0.001$) groups. However, there was no significant difference between the middle-distance and long-distance groups' scores (mean difference = 0.58, $p = 0.06$). For advanced Chinese proficiency, there was no significant difference between the short-distance and middle-distance groups' vocabulary scores (mean difference = 0.10, $p = 0.97$), but the two groups' scores were significantly higher than those of the long-distance group (mean difference = 0.98, $p < 0.001$, and mean difference = 0.88, $p = 0.001$, respectively) (see Figure 4).



三、阅读理解

(共 50 题，60 分钟)

第一部分

说明：81—100 题，每个句子中都有一个画线的词语，A B C D 四个答案是对这一画线的词语的不同解释。请选择最接近该词语的一种解释（**在答卷上**的字母上画一条横道）。

81. 据好事的业余观察家**分析**，问题是出在严莉一方。
A. 推断
B. 回答
C. 报告
D. 承认

82. 仁民带着严肃的表情在看广场上的**群众**。
A. 羊群
B. 士兵
C. 干部
D. 百姓

83. 您的**先进**事迹，对我的触动很大。
A. 优秀
B. 古老
C. 重要
D. 神奇

84. 她的**理想**，她的见解，有许许多多和我相同的地方。
A. 追求
B. 性格
C. 观点
D. 脾气

85. 忽然从外面**飘**进来一个黑影，把他吓了一跳。
A. 冲
B. 走
C. 闯
D. 飞

86. 原来世人所**歌颂**的爱情竟也有令人如此遗憾的一面。
A. 向往
B. 赞美
C. 难忘
D. 珍惜

87. 他们谈论着，热烈地谈论着，一直到全体**出发**的时候。
A. 停止
B. 抵达
C. 动身
D. 发言

88. 我的作品太**不像话**了，除了几个知己之外，我一概不送。
A. 简单
B. 贵重
C. 糟糕
D. 特殊

M12TB01X

10

FIGURE 3
Sample test in the HSK vocabulary subtest.

Quasi-Experiment 2 showed that linguistic distance affected CSL learners' vocabulary acquisition, and that vocabulary performance differed across distance groups under the elementary, intermediate, and advanced Chinese proficiency conditions. First, the short-distance group's vocabulary scores were significantly higher than those of the middle-distance group in the elementary and intermediate Chinese proficiency conditions, but there was no significant difference between the two groups in the advanced Chinese proficiency condition. This indicates that difference in vocabulary acquisition performance between the short-and

TABLE 3 Chinese vocabulary test score descriptive statistics.

Chinese proficiency level	Linguistic distance level	Mean values of Chinese vocabulary
Elementary	Short	7.44 (SD = 2.80)
	Middle	6.31 (SD = 2.42)
	Long	6.06 (SD = 2.57)
Intermediate	Short	11.91 (SD = 3.13)
	Middle	10.86 (SD = 3.42)
	Long	10.29 (SD = 3.12)
Advanced	Short	16.33 (SD = 2.22)
	Middle	16.22 (SD = 2.50)
	Long	15.34 (SD = 2.57)

middle-distance groups decreased as Chinese proficiency increased. Second, there was no significant difference between the middle-distance and long-distance groups in the elementary and intermediate proficiency conditions, but the middle-distance group had significantly higher vocabulary scores than the long-distance group under the advanced Chinese proficiency condition. This indicates that as Chinese proficiency improved, vocabulary acquisition performance differences between the middle-distance and the long-distance groups increased. Finally, although the difference in vocabulary scores between the short-and long-distance groups showed a decreasing trend, the short-distance group's vocabulary scores were significantly higher than those of the long-distance group at any Chinese proficiency level.

6. Quasi-experiment 3

6.1. Methods

6.1.1. Participants

Quasi-Experiment 3 data came from the same database as the previous two quasi-experiments, and the method and number of participants were also selected in the same way as for Quasi-Experiments 1 and 2 (1,136 females; age range 11.2 to 61.9 years, mean age = 23.53, standard deviation = 5.91).

6.1.2. Variables

The Quasi-Experiment 3 dependent variable was HSK grammar scores, and the two independent between-subject variables were linguistic distance between L1 and Chinese, and Chinese proficiency.

6.1.2.1. Grammar test scores

Grammar scores were derived from the HSK grammar subtests, which measure Chinese grammar knowledge by examining grammar usage accuracy, where higher grammar subtest scores indicate higher grammar proficiency. The grammar subtest takes 30 min to complete. Figure 5 shows a grammar subtest sample.

6.1.2.2. Chinese proficiency

Quasi-Experiment 3 grouped Chinese proficiency the same as in quasi-experiments 1 and 2. There was a significant positive correlation between total HSK score and total HSK score minus the grammar subtest score ($\rho = 0.993$, $p < 0.001$). Scores less than 56 were defined as elementary proficiency ($n = 10{,}510$); scores between 57 and 98 as intermediate proficiency ($n = 33{,}305$); and scores above 99 as advanced proficiency ($n = 14{,}425$).

6.1.2.3. Linguistic distance

The linguistic distance calculation and grouping methods were the same as in Quasi-Experiments 1 and 2.

6.1.3. Data analysis

The data analysis approach and process were the same as for Quasi-Experiments 1 and 2.

6.2. Results

To examine whether and how linguistic distance affected Chinese grammar acquisition by elementary, intermediate, and advanced proficiency, we analyzed the HSK grammar scores of 2,250 CSL learners in nine groups. Table 4 shows the grammar test score descriptive statistics for the nine groups.

Two-way between-subjects ANOVA revealed a significant main effect of linguistic distance ($F_{(2, 2{,}241)} = 16.111$, $p < 0.001$, $\eta_p^2 = 0.014$) and Chinese proficiency ($F_{(2, 2{,}241)} = 2499.552$, $p < 0.001$, $\eta_p^2 = 0.690$), and a significant interaction effect ($F_{(4, 2{,}241)} = 3.323$, $p = 0.01$, $\eta_p^2 = 0.006$).

The simple effect test showed that the elementary proficiency group had no significant difference in grammar scores across the long-, middle-, and short-distance groups (mean difference = 0.58, $p = 0.23$; mean difference = 0.43, $p = 0.49$; mean difference = 0.15, $p = 0.96$, respectively). The intermediate proficiency group's long-distance grammar scores were significantly higher than those of the short-distance group (mean difference = 1.03, $p = 0.01$). The middle-distance group's grammar scores were also significantly higher than those of the short-distance group (mean difference = 0.86, $p = 0.03$), but there was no significant difference between the long-and middle-distance groups (mean difference = 0.17, $p = 0.94$). For advanced proficiency, the long-distance group's grammar scores were significantly higher than those of the short-distance group (mean difference = 1.23, $p = 0.001$); the middle-distance group's grammar scores were also significantly higher than those of the short-distance group (mean difference = 1.83, $p < 0.001$). Although the middle-distance group's average score was higher than that of the long-distance group, the difference was not significant (mean difference = 0.60, $p = 0.20$) (see Figure 6).

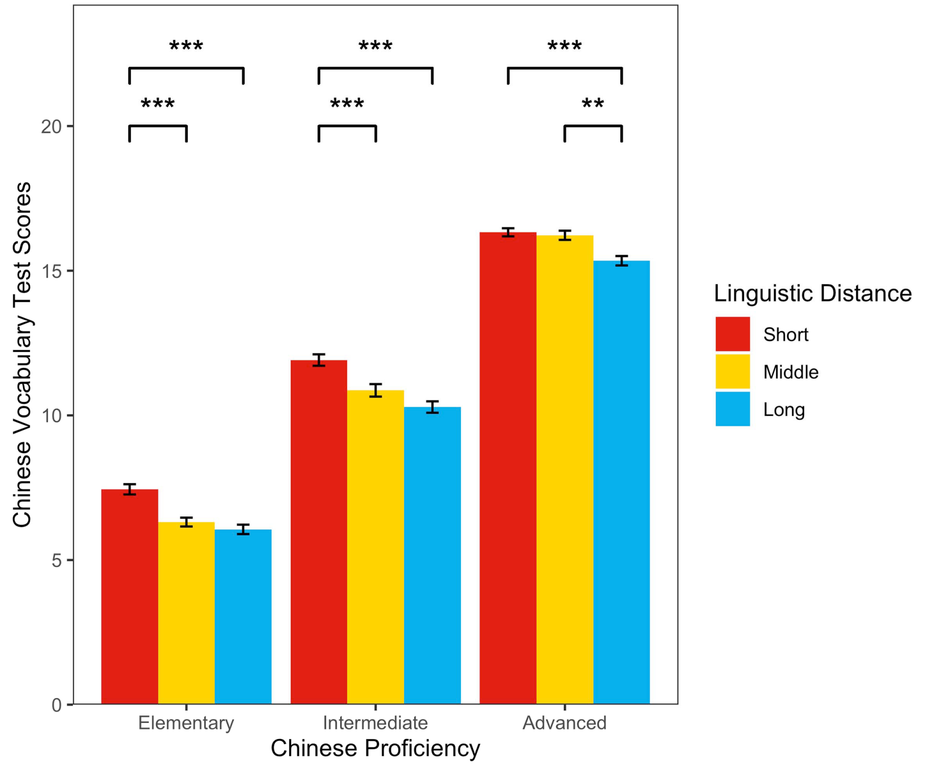


FIGURE 4
The relationship between Chinese proficiency, linguistic distance, and Chinese vocabulary test scores.

Quasi-Experiment 3 showed that linguistic distance affected CSL learners' grammar acquisition, and that distance groups' grammar scores were not consistent across the elementary, intermediate, and advanced proficiency groups. Under elementary Chinese proficiency, there was no significant difference in grammar scores across the distance groups. However, under intermediate and advanced proficiency, the middle-and long-distance groups' grammar scores were significantly higher than those of the short-distance group, indicating that CSL grammar acquisition differences increased as Chinese proficiency improved, as shown between the long-and short-distance groups, and the middle-and short-distance groups.

7. General discussion

This study explored the effects of linguistic distance on CSL learners' Chinese character, vocabulary, and grammar knowledge acquisition by Chinese proficiency level. The results fully demonstrated the complexity and dynamics of L1 difference effects in language knowledge acquisition. First, the Chinese character acquisition results showed that shorter linguistic distance between L1 and Chinese was associated with better acquisition across the elementary, intermediate, and advanced Chinese proficiency levels. Second, the vocabulary acquisition results showed that, at elementary, intermediate, and advanced proficiency levels, shorter linguistic distance between L1 and Chinese was associated with better acquisition. However, the difference in vocabulary acquisition performance between the short-and middle-distance groups gradually decreased as learners' Chinese proficiency improved, while the difference in vocabulary acquisition performance between the long-and middle-distance groups gradually increased. Third, the grammar acquisition results showed that linguistic distance did not significantly affect CSL learners' grammar acquisition for those with elementary Chinese proficiency, but at intermediate and advanced Chinese proficiency, longer linguistic distance was beneficial to grammar acquisition. These results indicate that only Chinese character and vocabulary acquisition support the cross-language transfer theory hypothesis; grammar acquisition does not support the theory.

7.1. Linguistic distance and Chinese character acquisition

Quasi-Experiment 1 examined linguistic distance effects on CSL learners' Chinese character acquisition at different Chinese proficiency levels. The results showed that, for elementary, intermediate, or advanced Chinese proficiency levels, Chinese character acquisition was more favorable with a shorter linguistic

2 • 2 • 2 • 2 • 2 • 2 • 2

二、 语 法 结 构

(共 30 题，20 分钟)

第 一 部 分

说明：51－60 题，在每一个句子下面都有一个指定词语，句中 A B C D 是供选择的四个不同位置。请判断这一词语放在句子中哪个位置上恰当。

例如：

55. 我们 A 一起 B 去上海 C 旅游 D 过。
没有

"没有"只有放在句中 A 的位置上，使全句变为"我们没有一起去上海旅游过"，才合乎语法，所以第 55 题唯一恰当的答案是 A。你应该**在答卷上**找到号码 55，在字母 A 上画一条横道。横道一定要画得粗一些、重一些。

55. ■ [B] [C] [D]

51. 这有什么不会 A 的，简单 B，比"桥牌" C、"杜勒克"都 D 容易。
得很

52. 这 A 是 B 一个难以 C 完成的 D 任务。
显然

53. 答应了我们，你们就 A 应该认真地去 B 做，你们 C 能说了话 D 不算数呢？
怎么

54. 这时 A 一个小孩 B 走了 C 过去，D 说："我来吧。"
只见

55. 海难 A 中唯一 B 一个 C 幸存 D 者被海水冲到了一个无人小岛上。
的

56. 泸州 A 以"酒城" B 闻名于世，若 C 提到泸州不能不 D 提到酒。
向来

57. 一个人开车在乡间迷了路，他 A 边开车 B 边查看地图 C，结果 D 把车驶离了狭窄的乡间小路。
一下子

58. 他 A 逗得孩子们 B 肚子 C 笑 D 疼了。
都

59. 我 A 认为不能让孩子 B 读完 C 的教材不是好 D 教材。
一口气

60. 她的父亲，A 年轻时 B 能一气喝两斤白酒，C 现在 D 能喝一斤的老人，站在屋前破口大骂。
仍然

M12TB01X

6

FIGURE 5
Sample test in the HSK vocabulary subtest.

distance between L1 and Chinese, but this effect was only observed between the short-and middle-distance groups and the short-and long-distance groups. There were six L1s in the short-distance group comprising Vietnamese, Thai, Korean, Indonesian, Yoruba, and Japanese individuals. Japan, Korea, and Vietnam are Sinosphere member countries, while Indonesia and Thailand are both neighbors of China. Thus, our findings are consistent with the findings of most previous studies that Chinese characters are better acquired in Sinosphere CSL learners than in non-Sinosphere CSL learners (Jiang, 2003; Feng et al., 2005; Zhang, 2017; Loh et al., 2018; Deng and Hu, 2022). This result is partially consistent with cross-language transfer theory assumptions. According to previous studies, CSL learners in short-distance groups (mainly from the Sinosphere) have a basic understanding of the rules for writing Chinese characters because their native script has Chinese character forms, and they can acquire Chinese character meanings much easier because they have more frequent exposure to Chinese culture. Therefore, short-distance CSL learners have an advantage in both Chinese character experience and literacy (Feng et al., 2005; Feng, 2006; Lin and Collins, 2012). By contrast, the middle-and long-distance CSL learners have alphabetic language backgrounds; the stereoscopic structure of Chinese characters is more complex compared to the linear structure of alphabetic characters (Everson, 1998; Shi and Wan, 1998). Additionally, literary strategies in relation to Chinese characters differ greatly from those of alphabetic characters (Tan et al., 2005). The form-phonemic connection of Chinese characters is poor; however, CSL learners from alphabetic backgrounds are often influenced by their L1 and tend to use native-like form-phonemic strategies to recognize Chinese characters (Jiang, 2003). Moreover, there is a long-standing lack of emphasis on Chinese characters when teaching CSL (Li, 2017). All these factors increase the difficulty of Chinese character acquisition for CSL learners from non-Sinosphere countries, resulting in their inability to avoid the influence of negative cross-language transfer on Chinese character acquisition, even if they reach a relatively advanced Chinese proficiency level.

TABLE 4 Chinese grammar test score descriptive statistics.

Chinese proficiency level	Linguistic distance level	Mean values of Chinese grammar
Elementary	Short	10.68 (SD = 3.14)
	Middle	10.83 (SD = 3.86)
	Long	11.26 (SD = 3.67)
Intermediate	Short	16.96 (SD = 4.19)
	Middle	17.82 (SD = 4.41)
	Long	17.99 (SD = 4.25)
Advanced	Short	23.52 (SD = 3.63)
	Middle	25.36 (SD = 3.18)
	Long	24.75 (SD = 2.96)

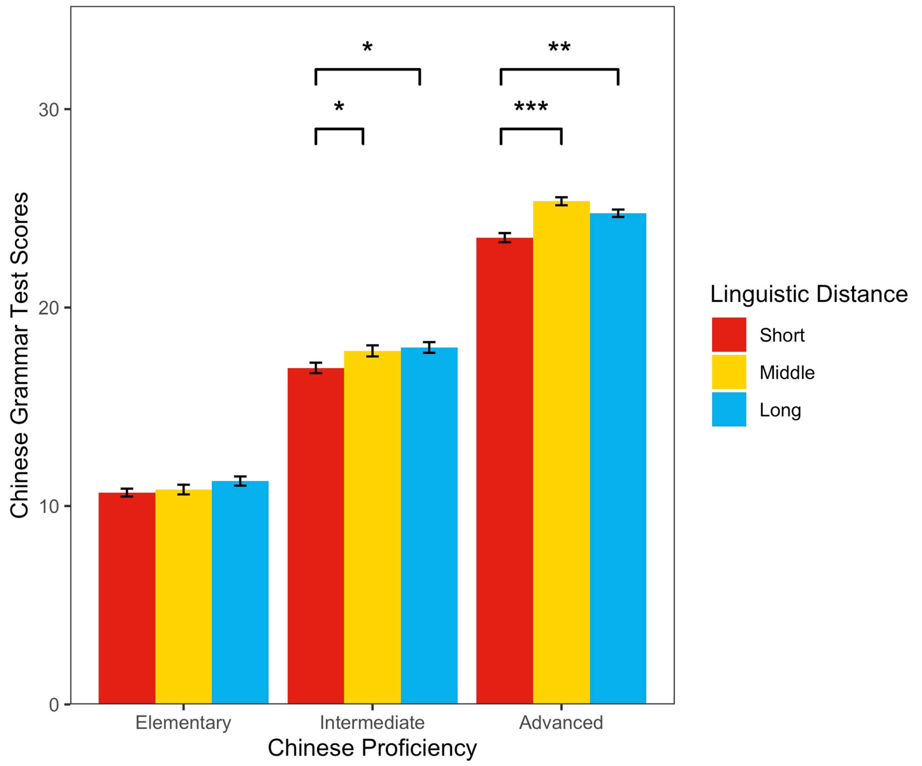


FIGURE 6
The relationship between Chinese proficiency, linguistic distance, and Chinese grammar test scores.

7.2. Linguistic distance and Chinese vocabulary acquisition

Quasi-Experiment 2 examined the influence of linguistic distance on CSL learners' vocabulary acquisition at different proficiency levels. The results showed that CSL learners with elementary and intermediate proficiency and short linguistic distance had a vocabulary acquisition advantage, and that CSL learners with advanced proficiency and either middle or short linguistic distance showed significantly better performance than those with a long linguistic difference. First, these results are partially consistent with the finding that shorter linguistic distance is more favorable to vocabulary acquisition, which accords with cross-language transfer theory and most previous study findings (Zhang, 2006; Cai, 2017; Luo and Duan, 2019; Chang, 2020). Previous studies have shown that, because Sinosphere L1s share many Chinese origin words and morphemes, CSL learners from Sinosphere countries have a certain awareness of Chinese morphemes and can use morphemic strategies to learn new words at the elementary stage (Hong, 2011; Xu, 2014). Shorter linguistic distance to Chinese often means more exposure to Chinese culture; since words in different languages can be linked through concepts (Zhang et al., 2011), shorter linguistic distance also means easier conceptual word linkage, which makes positive cross-language transfer easier.

Second, vocabulary scores of all distance groups significantly increased with improvements in CSL learners' Chinese proficiency levels. Vocabulary score differences between the short-and middle-distance groups significantly decreased, but vocabulary score differences between the long-and middle-distance groups significantly increased. We believe this result reflects a difference in the linguistic distance effect on vocabulary acquisition speed for non-Sinosphere CSL learners. As the middle-distance group had a linguistic distance advantage over the long-distance group, vocabulary knowledge expansion with improved Chinese proficiency could reduce the effect of negative cross-language transfer; whereas CSL learners in the long-distance group, whose L1 differed more from Chinese, were more affected by negative cross-language transfer, requiring them to overcome more vocabulary acquisition difficulties, which resulted in significantly slower vocabulary acquisition speeds for CSL learners in the long-distance group compared to the middle-distance group. Combined with the trend (although not significant) for decreased vocabulary score differences between the short-and long-distance groups, we believe that linguistic distance's influence on CSL learners' vocabulary acquisition generally diminished with increased Chinese proficiency.

Additionally, in comparing the results with those of Quasi-Experiment 1, we found similarities in the effects of linguistic distance on Chinese character and vocabulary acquisition, where shorter linguistic distance between L1 and Chinese was associated with more favorable acquisition, and showed a significant acquisition advantage for Sinosphere CSL learners at any proficiency level. This is most likely a result of Chinese characters and words being similarly acquired because of their blurred boundaries (Li, 2009). However, there were also differences in linguistic distance effects on character and vocabulary acquisition. Quasi-Experiment 2 showed no significant difference in vocabulary performance between the short-and middle-distance groups for CSL learners with advanced Chinese proficiency, while Quasi-Experiment 1 showed a significant difference. Considering that the HSK Chinese character subtest examines learners' Chinese character or morpheme writing, and the vocabulary subtest focuses on word or morpheme recognition, this result supports, to some extent, the finding that Chinese character or morpheme writing is more influenced by linguistic distance than is reading (Wu et al., 2006; Liu, 2008; Li et al., 2014). Based on previous studies, we believe that the main reason for this phenomenon is that Chinese character recognition can be accomplished by using only part of a character's information to remember its pronunciation or meaning, so it is only necessary to break these characters down into strokes; whereas writing requires learners to recall and reproduce the characters through their pronunciation and meaning, which requires mastery of not only the character radicals, but also the smaller stroke units (Jiang, 2000; Yeh et al., 2003; Xu Y. et al., 2013; Ke and Chan, 2017). Therefore, writing Chinese characters or morphemes is more difficult, resulting in a more profound negative cross-language transfer effect on Chinese character writing.

7.3. Linguistic distance and Chinese grammar acquisition

Interestingly, Quasi-Experiment 3 showed a pattern of cross-language transfer effects that completely differed from those found in Chinese character and vocabulary acquisition. There was no significant linguistic distance effect on grammar acquisition for CSL learners with elementary Chinese proficiency. This may be because cross-language transfer is evident in early stages, but in the middle-and long-distance groups, linguistic differences' promoting effect weakened the L1 negative transfer effects and facilitated acquisition (Yuan, 2004; Wu and Zhao, 2018). Meanwhile, numerous studies have demonstrated that elementary proficiency learners rely more on lexical semantic information than syntactic information (Clahsen and Felser, 2006, 2018; Rah and Adone, 2010); therefore, grammatical knowledge development frequently lags behind that of content meaning knowledge, such as Chinese characters and vocabulary, which may be another reason why L1 differences are unlikely to show a significant grammar acquisition effect for CSL learners with elementary Chinese proficiency.

Although linguistic distance influenced CSL learners' grammar acquisition at intermediate and advanced proficiency levels, short distance did not contribute to grammar acquisition at these proficiency levels. In other words, short distance did not produce a positive cross-language transfer, but rather had a hindering effect.

Previous studies suggest that this result may stem from a greater difference between learners' L1 and Chinese, which makes learners more likely to pay attention to the differences, facilitating acquisition (Yuan, 2004; Wu and Zhao, 2018); or it could be that a high degree of cross-language similarity makes learners more likely to ignore differences, making acquisition more difficult (Laufer, 1990; Ellis, 1999). This result can only occur when language features' similarity interference effects are significantly stronger than language differences' facilitation effects, and significantly stronger than L1 transfer effects. Additionally, with increased grammar learning content, the probability of overgeneralizing will be greater (Ellis, 1994); compared with obvious differences between learners' L1 and target language, the probability of overgeneralizing subtle differences between learners' L1 and target language is greater (Ellis, 1999). In other words, as Chinese proficiency improves and CSL learners are exposed to more grammatical knowledge, short-distance group CSL learners are not only subject to interference from certain grammar with highly similar features to Chinese, but also to interference from their native grammar. This may result in a longer period of confusion before learners fully master Chinese grammar. Therefore, it is reasonable to believe that enhanced interference effects of short-distance L1 on CSL learners' grammar acquisition inhibits grammatical competence development as their Chinese proficiency improves.

Previous studies showed that avoidance strategies and psycho-typological distance weaken L1 negative transfer and result in a non-significant acquisition advantage for short-distance CSL learners. However, as the HSK is a high-risk test with objective items, examinees are less likely to use avoidance strategies. The role of learners' perceived differences in language typology is beyond the scope of our study; these issues can be further explored in future studies.

Our study also found no significant differences in Chinese character, vocabulary, and grammar scores between different distance groups at each Chinese proficiency level, except for vocabulary scores for CSL learners with advanced Chinese proficiency. This finding indicates that linguistic distance is important for distinguishing language knowledge acquisition performance in both Sinosphere and non-Sinosphere CSL learners, but does not play an obvious role in distinguishing language knowledge acquisition performance within non-Sinosphere CSL learners. Given the lack of studies that specifically focus on differences in Chinese acquisition within non-Sinosphere CSL learners, we cannot yet offer a general explanation.

Our results show that linguistic distance has different patterns of influence on Chinese character, vocabulary, and grammar acquisition. As Ellis (1999) asserted, L1's influence on second language acquisition is not always a one-way positive or negative process, but involves a reciprocally dynamic process (Xu Q. et al., 2013).

7.4. Differences in the influence of linguistic distance on language knowledge and language skills

Previous studies showed that shorter linguistic distances between L1 and Chinese facilitate CSL learners' language skill development, such as listening, speaking, reading, and writing (Wang and Cui, 2018; Wang and Yang, 2019; Xu and Zi, 2020). Using HSK test data, we investigated the impact of linguistic distance between L1 and Chinese on character, vocabulary, and grammar knowledge acquisition. The results showed differences in linguistic distance's effect on the total characteristics of three types of language knowledge, and differences in the degree of influence on three types of language knowledge acquisition at different Chinese language proficiency levels. Our findings indicate that, at least under the influence of linguistic distance, language knowledge acquisition is not consistent with language skills acquisition, because the language knowledge acquisition process is more complex (Ding and Xiao, 2016; Wu, 2017).

8. Limitations and future directions

The current study has limitations. First, the linguistic features provided by the WALS online system for calculating cross-language similarity between L1 and Chinese vary in both the number of indicators and their categories. We could only examine test data from 24 L1 backgrounds that do not differ significantly from the number of Chinese linguistic features.

Second, the study only examined acquisition performance of intermediate HSK examinees at different proficiency levels, so the results do not reflect higher proficiency learners' acquisition characteristics. Future research using larger samples with richer L1 backgrounds is recommended.

Third, internal factors have important effects on second language acquisition, such as learners' L1, language proficiency and acquisition age (Ellis, 1994). This study focused on linguistic distance between learners' L1 and Chinese and Chinese proficiency level influences on Chinese learning. Future research could investigate the moderating effect of age on language knowledge acquisition under different linguistic distances and Chinese proficiency levels.[2]

9. Implications

To the best of our knowledge, this is the first empirical study to systematically examine the effect of linguistic distance between L1 and Chinese on Chinese character, vocabulary, and grammar knowledge acquisition by Chinese proficiency level, using HSK test data and a multidimensional quantitative linguistic distance method. The study findings have theoretical and practical implications.

2 We thank the editor for this recommendation.

From a theoretical perspective, first, this study provides additional systematic empirical evidence regarding the long-standing question about the role of learners' LI in second language acquisition in cross-language transfer theory. Combined with linguistic distance's varying degree of influence on each language knowledge for different Chinese proficiency levels, our findings fully reflect the complex and dynamic understanding of the cognitive and sociocultural schools of thought regarding the role of L1 in second language acquisition (Vygotsky, 1978; Ringbom, 2007). Second, our study also provides support for asymmetry in CSL learners' Chinese character or morpheme writing and recognition development (Li et al., 2014; Zhang and Roberts, 2021). Third, this study not only provides empirical evidence for Chinese character, vocabulary, and grammar acquisition performance in CSL learners from the Sinosphere, but also provides a reference for the less-studied Chinese character, vocabulary, and grammar acquisition patterns in CSL learners from non-Sinosphere countries. Finally, our findings provide new insights into the role of CSL learners' L1 in second language acquisition. L1 plays different roles in language skills and language knowledge acquisition, because the effects of L1 differences on language knowledge acquisition are more complex.

From a practical perspective, this study adds a new perspective to existing cross-language transfer studies that relied on specific feature comparisons to measure linguistic distance between L1 and Chinese. Second, our findings suggest that CSL teachers should pay more attention to developing learners' Chinese character writing, and use comparative analysis methods to help learners notice minute grammatical differences. Third, we recommend that students be taught in separate classes based on whether they are from Sinosphere or non-Sinosphere countries, to improve overall teaching and learning efficiency.

10. Conclusion

Our study conducted three quasi-experiments to systematically examine the influence of linguistic distance on Chinese character, vocabulary, and grammar knowledge acquisition, and their developmental characteristics among CSL learners with different Chinese proficiency levels. For the Chinese character and vocabulary acquisition, linguistic distance's effect on CSL learners with elementary, intermediate, and advanced Chinese proficiency levels was largely consistent with cross-language transfer theory, which assumes that shorter linguistic distance between L1 and Chinese facilitates acquisition. However, the effect of linguistic distance on CSL learners' grammar acquisition at the elementary Chinese proficiency level was not significant, whereas the effect at intermediate and advanced proficiency levels showed a short-distance L1 interference effect. Further, as CSL learners' Chinese language proficiency improved, the linguistic distance effect on Chinese character acquisition remained largely unchanged, while the effect on vocabulary acquisition gradually decreased and the effect on grammar acquisition gradually increased. The results show that linguistic distance has differential proficiency-dependent effects on Chinese character, vocabulary, and grammar acquisition.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

XC and JB conceived and designed the work. JB performed the statistical analysis and wrote the first draft of the manuscript. XC revised the manuscript critically. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This project was supported by the Key Projects of Beijing Language and Culture University Fund (Award Number 19ZDJ04).

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of Beijing Language and Culture University. We would like to thank the examinees and test centers who helped us gather the data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., et al. (2009). Adding typology to lexicostatistics: a combined approach to language classification. *Linguis. Typol.* 13, 169–181. doi: 10.1515/LITY.2009.009

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguis. Anal.* 20, 3–49. doi: 10.1017/CBO9781139524544.005

Cai, Y. (2017). Second language acquisition of Chinese verb-noun collocations, Master dissertation, University of. Massachusetts Amherst.

Cenoz, J. (2001). "The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition," in *Cross-linguistic Influence in third Language Acquisition: Psycholinguistic Perspectives*. eds. J. Cenoz, B. Hufeisen and U. Jessner (Clevedon, UK: Multilingual Matters).

Centeno-Cortés, B., and Jiménez Jiménez, A. F. (2004). Problem-solving tasks in a foreign language: the importance of the L1 in private verbal thinking. *Int. J. Appl. Linguist.* 14, 7–35. doi: 10.1111/j.1473-4192.2004.00052.x

Chai, X. (2006). "关于HSK(初、中等)平行信度的实证研究 [empirical research on the parallel test reliability of the elementary-intermediate HSK]" in 汉语水平考试研究 *[HSK research]*. ed. K. Zhang (Beijing: The Commercial Press)

Chai, X., and Ma, M. (2022). Exploring the relationships between L2 Chinese character writing and reading acquisition from the embodied cognitive perspectives: evidence from the HSK big data. *Front. Psychol.* 12:779190. doi: 10.3389/fpsyg.2021.779190

Chang, X. (2020). 搭配强度与结构类型对中高级泰国汉语学习者限制性搭配加工的影响 [a study of the influence of collocational strength and structure types on the processing of restricted collocations by intermediate and advanced Chinese learners from Thailand]. 语言教学与研究 *[Lang. Teach. Linguis. Stud.]* 3, 22–32.

Chen, H. (2006). "关于考生团体异质性程度对HSK(初、中等)信度的影响 [on the influence of the test takers' population's degree of difference on the elementary-intermediate HSK's reliability]" in 汉语水平考试研究*[HSK Research]*. ed. K. Zhang (Beijing: The Commercial Press).

Chiswick, B. R., and Miller, P. W. (2005). Linguistic distance: a quantitative measure of the distance between English and other languages. *J. Multiling. Multicult. Dev.* 26, 1–11. doi: 10.1080/14790710508668395

Clahsen, H., and Felser, C. (2006). Continuity and shallow structures in language processing. *Appl. Psycholinguist.* 27, 107–126. doi: 10.1017/S0142716406060206

Clahsen, H., and Felser, C. (2018). Some notes on the shallow structure hypothesis. *Stud. Second. Lang. Acquis.* 40, 693–706. doi: 10.1017/S1366728915000942

Cook, V., and Singleton, D. (2014). *Key Topics in Second Language Acquisition*. (Bristol: Multilingual matters).

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Cui, M., Zhang, W., and Sun, T. (2018). 语言距离, 母语差异与汉语习得: 基于语言经济学的实证研究 [linguistic distance, language learners' mother tongue, and Chinese acquisition: an empirical study from the perspective of language economic]. 世界汉语教学 *[Chin. Teach. World]* 2, 280–288. doi: 10.13724/j.cnki.ctiw.2018.02.015

Dai, M. (2014). 二语习得研究中的接口假说 [the interface hypothesis in SLA research revisited]. 外语学刊 *[Foreign Lang. Res.]* 4, 109–117. doi: 10.16263/j.cnki.23-1071/h.2014.04.023

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., and Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Appl. Psycholinguist.* 34, 893–916. doi: 10.1017/S0142716412000069

De Saussure, F. (1959). *Course in General Linguistics*. New York: Columbia University Press.

Deng, S., and Hu, W. (2022). An examination of Chinese character writing errors: developmental differences among Chinese as a foreign language learners. *J. Chin. Writ. Syst.* 6, 39–51. doi: 10.1177/25138502221066

Ding, A., and Xiao, X. (2016). 意大利学习者初级汉语口语词汇能力发展研究 [on the development of oral lexical competence of Italian Chinese language learners]. 世界汉语教学 *[Chin. Teach. World]* 2, 239–252. doi: 10.13724/j.cnki.ctiw.2016.02.012

Dryer, M. S., and Haspelmath, M. (2013). The world atlas of language structures online. WALS.info Available at: http://wals.info (Accessed July 18, 2021).

Dulay, H. C., and Burt, M. K. (1973). Should we teach children syntax? *Lang. Learn.* 23, 245–258. doi: 10.1111/j.1467-1770.1973.tb00659.x

Dyen, I., Kruskal, J. B., and Black, P. (1992). An Indoeuropean classification: a lexicostatistical experiment. *Trans. Am. Philos. Soc.* 82, iii–132. doi: 10.2307/1006517

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R. (1999). *Understanding Second Language Acquisition*. Shanghai: Shanghai Foreign Language Education Press.

Ellis, R. (2008). *The Study of Second Language acquisition, 2nd Edn*. New York, NY: Oxford University Press.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Appl. Linguis.* 27, 164–194. doi: 10.1093/applin/aml015

Epstein, A. S. (1996). *Models of Early Childhood Education*. Ypsilanti, MI: High/Scope Press.

Everson, M. E. (1998). Word recognition among learners of Chinese as a foreign language: investigating the relationship between naming and knowing. *Mod. Lang. J.* 82, 194–204. doi: 10.2307/329208

Fei, J. (1996). 现代汉字部件探究 [research of modern Chinese characters components]. 语言文字应用 *[Appl. Linguis.]* 2, 20–26.

Feng, L. (2006). 外国留学生汉字正字法意识及其发展研究 [a study of Chinese orthographic awareness by foreign students and its development]. 云南师范大学学报 *[J. Yunnan Norm. Univ. (Teaching & Studying Chinese as a Foreign Language Edition)]* 1, 12–17. doi: 10.16802/j.cnki.ynsddw.2006.01.003

Feng, L., Lu, H., and Xu, C. (2005). 部件位置信息在留学生汉字加工中的作用 [the role of information about. Radical position in processing Chinese characters by foreign students]. 语言教学与研究 *[Lang. Teach. Linguis. Stud.]* 3, 66–72.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied Longitudinal Analysis*, Vol. *998*. Hoboken, NJ: John Wiley & Sons.

Floccia, C., Sambrook, T., Delle Luche, C., Kwok, R., Goslin, J., White, L., et al. (2018). Vocabulary of 2-year-olds learning English and an additional language: norms and effects of linguistic distance. V: GENERAL DISCUSSION. *Monogr. Soc. Res. Child Dev.* 83, 68–80. doi: 10.1111/mono.12352

Frawley, W., and Lantolf, J. P. (1985). Second language discourse: a Vygotskyan perspective. *Appl. Linguis.* 6, 19–44. doi: 10.1093/applin/6.1.19

Gass, S. M., and Selinker, L. (1992). *Language Transfer in Language Learning*: Revised Edition.*5* Amsterdam: John Benjamins Publishing.

Ghasemi, A., and Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *Int. J. Endocrinol. Metab.* 10, 486–489. doi: 10.5812/ijem.3505

Goldschneider, J. M., and DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: a meta-analysis of multiple determinants. *Lang. Learn.* 51, 1–50. doi: 10.1111/1467-9922.00147

Grauberg, W. (1971). "An error analysis in German of first year university students," in *Papers for the second international congress of applied linguistics*. eds. G. Perren and J. Trim (London: Cambridge University Press).

Guo, F., and Liu, H. (2017). 中高级水平学习者汉语语序偏误的类型学分析 [typology analysis of Chinese order error of intermediate and advanced level Chinese learners]. 汉语学习 *[Chin. Lang. Learn.]* 2, 98–105.

Ginsburgh, V. A., and Weber, S. (2014). Culture, linguistic diversity, and economics. *Handb. Econ. Art Cult.* 2, 507–543. doi: 10.1016/B978-0-444-53776-8.00019-2

Green, D. W., and Abutalebi, J. (2013). Language control in bilinguals: the adaptive control hypothesis. *J. Cogn. Psychol.* 25, 515–530. doi: 10.1080/20445911.2013.796377

Hao, Y., Duan, X., and Yan, Q. (2022). Processing aspectual agreement in a language with limited morphological inflection by second language learners: an ERP study of mandarin Chinese. *Brain Sci.* 11:524. doi: 10.3390/brainsci11091236

Hao, Y., Wang, X., and Yu, Q. (2021). Typological characteristics of interlanguage: across native language types and L2 proficiency levels. *Lingua* 257:103085. doi: 10.1016/j.lingua.2021.103085

Hanban (2011). Chinesetest. Cn. Available at: http://www.chinesetest.cn/gonewcontent.do?id=2317021 (Accessed July 18, 2021).

Hart-Gonzalez, L., and Lindemann, S. (1993). *Expected Achievement in Speaking Proficiency* Foreign Service Institute, Department of State, School of Language Studies.

Hong, W. (2011). 语素因素对留学生近义词学习影响的实证研究 [An experimental research on the morphemic effect on synonyms learning by CSL learners]. 语言教学与研究 *[Chin. Teach. World]* 3, 424–432.

Hong, W. (2013). 汉语作为第二语言的近义词教学实验研究 [an empirical study on the teaching of near-synonyms to learners of Chinese as a second language]. 世界汉语教学 *[Chin. Teach. World]* 3, 424–432. doi: 10.13724/j.cnki.ctiw.2013.03.011

Hong, W., and Chen, N. (2011). 汉语二语者近义词差异的习得考察 [a study on the L2 acquisition of differences in similar sense and dissimilar sense of Chinese near-synonyms]. 语言文字应用 *[Appl. Linguis.]* 2, 99–106. doi: 10.16499/j.cnki.1003-5397.2013.02.015

Hsieh, M. L., and Wang, Y. F. (2020). Acquiring the polysemous adverb HAI in Chinese by English-speaking, Japanese-speaking, and Korean-speaking CSL learners. *Concentric* 46, 173–205. doi: 10.1075/consl.00017.hsi

Hu, L., Chang, H., and Zheng, L. (2018). 母语为英语和日语的学习者对汉语话题-述题结构的习得研究: 可加工性理论视角[Acquisition of Chinese topic-comment constructions by English-and Japanese-speaking learners: processability theory approach]. 语言教学与研究*[Lang. Teach. Linguis. Stud.]* 3, 17–28.

Isphording, I. E., and Otten, S. (2013). The costs of babylon—linguistic distance in applied economics. *Rev. Int. Econ.* 21, 354–369. doi: 10.1111/roie.12041

Jarvis, S. (1998). *Conceptual Transfer in the Interlingual Lexicon*. Bloomington, IN: Indiana University Linguistics Club Publications.

Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.

Jeon, E. H., and Yamashita, J. (2014). L2 reading comprehension and its correlates: a meta-analysis. *Lang. Learn.* 64, 160–212. doi: 10.1111/lang.12034

Jiang, N., Hou, F., and Jiang, X. (2020). Analytic versus holistic recognition of Chinese words among L2 learners. *Mod. Lang. J.* 104, 567–580. doi: 10.1111/modl.12662

Jiang, N. (2000). Lexical representation and development in a second language. *Appl. Linguis.* 21, 47–77. doi: 10.1093/applin/21.1.47

Jiang, X. (2003). 不同母语背景的外国学生汉字知音和知义之间关系的研究 [the relationship between knowing pronunciation and knowing meaning of Chinese characters among CSL learners]. 语言教学与研究 *[Lang. Teach. Linguis. Stud.]* 6, 51–57.

Ke, S., and Chan, S. D. (2017). Strategy use in L2 Chinese reading: the effect of L1 background and L2 proficiency. *System* 66, 27–38. doi: 10.1016/j.system.2017.03.005

Klein, W., Klein, W. W., and Wolfgang, K. (1986). *Second Language Acquisition*. New York: Cambridge University Press.

Krashen, S-D. (1985). *The Input Hypothesis: Issues and Implications*. Boston: Addison-Wesley Longman Limited.

Kleinmann, H. H. (1978). *The Strategy of Avoidance in Adult Second Language Acquisition*. Barcelona: Second Language Acquisition Research Issues and Implications.

Lado, R. (1957). *Linguistics Across Cultures*. Ann Arvor, MI: University of Michigan Press.

Lantolf, J. P., and Appel, G. (1994). *Vygotskian Approaches to Second Language Research*. Connecticut: Greenwood Publishing Group.

Lantolf, J. P., and Thorne, S. L. (2006). *Sociocultural Theory and Genesis of Second Language Development*. Oxford: Oxford University Press.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Appl. Linguis.* 18, 141–165. doi: 10.1093/applin/18.2.141

Larsen-Freeman, D., and Cameron, L. (2008). *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.

Laufer, B. (1990). Why are some words more difficult than others? — some intralexical factors that affect the learning of words. *Int. Rev. Appl. Linguis. Lang. Teach.* 28, 293–308. doi: 10.1515/iral.1990.28.4.293

Li, L., Li, X., and Ao, L. (2014). 汉语水平与母语背景对汉字正字法意识的影响 [influences of Chinese proficiency and native language background on Chinese orthographic awareness of foreign students]. 心理研究 *[Psychol. Res.]* 6, 37–42.

Li, Q. (2016). 对外汉语教学语法体系: 目的、标准和特点 [the pedagogical grammar system of Chinese as a second language: aim, standards and characteristics]. 国际汉语教学研究 *[J. Int. Chin. Teach.]* 1, 45–55.

Li, Q. (2017). 汉语教学:本位观与"字本位" [Chinese teaching: unit-based view and Chinese character as the basic unit]. 国际汉语教学研究 *[J. Int. Chin. Teach.]* 3, 13–19.

Li, R. (2009). 论汉语和汉字的关系及相关的研究 [on the correlativity between Chinese and the Chinese characters]. 语言教学与研究 *[Lang. Teach. Linguis. Stud.]* 4, 9–17.

Li, Y., Kang, J., Wei, L., and Zhang, S. (1992). 现代汉语形声字研究 [a study of modern Chinese phonogram]. 语言文字应用 *[Appl. Linguist.]* 1, 74–83.

Lin, C. H., and Collins, P. (2012). The effects of L1 and orthographic regularity and consistency in naming Chinese characters. *Read. Writ.* 25, 1747–1767. doi: 10.1007/s11145-011-9340-9

Lindgren, E., and Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *Int. J. Mult.* 10, 105–129. doi: 10.1080/14790718.2012.679275

Liu, L. (2008). 笔画数与结构方式对留学生汉字学习的影响 [the impacts of the stroke number and structural. Way of Chinese characters on foreign students in learning Chinese characters]. 语言教学与研究 *[Lang. Teach. Linguist. Stud.]* 1, 89–96.

Liu, Y. (2013). 汉语学习者汉字构形意识的发展研究 [a study of the development of CSL learner's Chinese orthographic awareness]. 云南师范大学学报*(*对外汉语教学与研究版*) [J. Yunnan Norm. Univ. (Teaching and Studying Chinese as a Foreign Language Edition)]* 6, 35–42.

Loh, E. K. Y., Liao, X., and Leung, S. O. (2018). Acquisition of orthographic knowledge: developmental difference among learners with Chinese as a second language (CSL). *System* 74, 206–216. doi: 10.1016/j.system.2018.03.018

Long, M. (1996). "The role of the linguistic environment in second language acquisition," in *Handbook of Second Language Acquisition*. eds. W. C. Ritchie and T. K. Bhatia (San Diego: Academic Press), 413–468.

Lowie, W., van Dijk, M., Chan, H., and Verspoor, M. (2017). Finding the key to successful L2 learning in groups and individuals. *Stud. Sec. Lang. Learn. Teach.* 7, 127–148. doi: 10.14746/ssllt.2017.7.1.7

Lu, S. (2002). 母语为拼音文字的学习者汉字正字法意识发展的研究 [a study on the orthographic awareness among the students possessing alphabetic writing]. 语言教学与研究 *[Lang. Teach. Linguis. Stud.]* 3, 53–57.

Luo, Y., and Duan, Y. (2019). 不同母语背景汉语学习者书面产出词汇复杂性差异研究 [a study on the lexical complexity in written production by CSL learners with different L1 backgrounds]. 华文教学与研究 *[TCSOL Stud.]* 3, 51–62. doi: 10.16131/j.cnki.cn44-1669/g4.2019.03.008

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One* 5:e8559. doi: 10.1371/journal.pone.0008559

Lv, S. (1979). 汉语语法分析问题 *[Chinese Grammar Analysis Problems]*. Beijing: The Commercial Press.

MacWhinney, B., and Kroll, J. F. (2005). "A unified model of language acquisition," in *Handbook of Bilingualism: Psycholinguistic Approaches*, *Vol. 4967*. eds J. F. Kroll and A. M. B. de Groot Oxford: Oxford University Press. 50–70.

McCloskey, D. (1998). *The Rhetoric of Economics*. Madison: University of Wisconsin Press.

Meara, P., and Jones, G. (1988). Vocabulary size as a placement indicator. *Comp. Anal.* 9, 80–87.

Meyer, F. K. (2014). *Language Proficiency Testing for Chinese as a Foreign Language: An Argument-based Approach for Validating the Hanyu Shuiping Kaoshi (HSK)*. Pieterlen: Peter Lang AG.

Moinzadeh, A., and Moslehpour, R. (2012). Depth and breadth of vocabulary knowledge: which really matters in reading comprehension of Iranian EFL learners? *J. Lang. Teach. Res.* 3:1015. doi: 10.4304/jltr.3.5.1015-1026

Moran, S., and Blasi, D. (2014). "Cross-linguistic comparison of complexity measures in phonological systems," in *Measuring Grammatical Complexity*. eds. F. J. Newmeyer and L. B. Preston (Oxford, UK: Oxford University Press).

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. (New York: Heinle and Hienle).

Odlin, T. (1989). *Language Transfer*. Cambridge: Cambridge University Press.

Odlin, T. (2001). *Cross-linguistic Influence in Language Learning*. Shanghai: Shanghai Foreign Language Education Press.

Odlin, T., and Jarvis, S. (2004). Same source, different outcomes: a study of Swedish influence on the acquisition of English in Finland. *Int. J. Multiling.* 1, 123–140. doi: 10.1080/14790710408668183

Pan, J., Yan, M., Laubrock, J., and Shu, H. (2019). Lexical and sublexical phonological effects in Chinese silent and oral reading. *Sci. Stud. Read.* 23, 403–418. doi: 10.1080/10888438.2019.1583232

Pica, T. (1994). Research on negotiation: what does it reveal about second-language learning conditions, processes, and outcomes? *Lang. Learn.* 44, 493–527. doi: 10.1111/j.1467-1770.1994.tb01115.x

Platzack, C. (1996). "The initial hypothesis of syntax," in *Generative Perspectives on Language Acquisition*. ed. H. Clahsen (Amsterdam/Philadelphia: John Benjamins).

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Can. Mod. Lang. Rev.* 56, 282–308. doi: 10.3138/cmlr.56.2.282

Rah, A., and Adone, D. (2010). Processing of the reduced relative clause versus main verb ambiguity in L2 learners at different proficiency levels. *Stud. Second. Lang. Acquis.* 32, 79–109. doi: 10.1017/S027226310999026X

Ringbom, H. (2007). *Cross-linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters LTD.

Schepens, J. J., Van der Slik, F., and Van Hout, R. (2016). L1 and L2 distance effects in learning L3 Dutch. *Lang. Learn.* 66, 224–256. doi: 10.1111/lang.12150

Schwartz, B. D., and Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second. Lang. Res.* 12, 40–72. doi: 10.1177/026765839601200103

Stockwell, R. P., Bowen, J. D., and Martin, J. W. (1965). *The Grammatical Structures of English and Spanish*. Chicago: University of Chicago Press.

Sjoholm, K. (1995). *The influence of Crosslinguistic, semantic, and input factors on the Acquisition of English phrasal verbs: a comparison between Finnish and Swedish learners at an intermediate and advanced level*. Turku: Abo Akademi University Press.

Swain, M., and Lapkin, S. (1995). Problems in output and the cognitive processes they generate: a step towards second language learning. *Appl. Linguis.* 16, 371–391. doi: 10.1093/applin/16.3.371

Swain, M., and Lapkin, S. (1998). Interaction and second language learning: two adolescent French immersion students working together. *Mod. Lang. J.* 82, 320–337. doi: 10.1111/j.1540-4781.1998.tb01209.x

Shi, D., and Wan, Y. (1998). 关于对外汉字教学的调查报告 [report on teaching Chinese characters to foreign students]. 语言教学与研究 *[Lang. Teach. Linguist. Stud.]* 1, 36–48.

Shi, G., and Yang, C. (2021). 汉语动词句法框架信息加工机制及二语习得研究 [research on the information. Processing mechanism and second language acquisition of Chinese verbal syntactic structures]. 语言教学与研究 *[Lang. Teach. Linguist. Stud.]* 1, 33–43.

Shi, G., Yang, C., and Li, M. (2021). 汉语二语学习者词汇搭配知识地道性发展研究 [a study on the idiomatic. Development of CSL learners' vocabulary collection knowledge]. 国际中文教育(中英文) *[Int. Chin. Lang. Educ.]* 3:5-12+29.

Sokal, R. R., and Rohlf, F. J. (1995). *Biometry: The Principles and Practice of Statistics in Biological Research*, *3rd Edn*. New York, NY: W.H. Freeman and Company.

Spoelman, M., and Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: a longitudinal case study in the acquisition of Finnish. *Appl. Linguist.* 31, 532–553. doi: 10.1093/applin/amq001

Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., and Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8781–8785. doi: 10.1073/pnas.0503523102

Tang, M., and Chan, S. D. (2022). Effects of word semantic transparency, context length, and L1 background on CSL learners' incidental learning of word meanings in passage-level reading. *J. Psycholinguist. Res.* 51, 33–53. doi: 10.1007/s10936-021-09786-z

Teng, Y. (2017). "Hanyu Shuiping Kaoshi (HSK): past, present, and future," in *Chinese Language Learning Sciences*. eds. D. Zhang and C. H. Lin (Singapore: Springer Singapore).

Tong, X., and Yip, J. H. Y. (2015). Cracking the Chinese character: radical sensitivity in learners of Chinese as a foreign language and its relationship to Chinese word reading. *Read. Writ.* 28, 159–181. doi: 10.1007/s11145-014-9519-y

Tsai, J. L., Lee, C. Y., Tzeng, O. J. L., Hung, D. L., and Yen, N. S. (2004). Use of phonological codes for Chinese characters: evidence from processing of parafoveal preview when reading sentences. *Brain Lang.* 91, 235–244. doi: 10.1016/j.bandl.2004.02.005

Tsang, Y. K., Wu, Y., Ng, H. T. Y., and Chen, H. C. (2017). Semantic activation of phonetic radicals in Chinese. *Lang. Cogn. Neurosci.* 32, 618–636. doi: 10.1080/23273798.2016.1246744

Vainikka, A., and Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second. Lang. Res.* 12, 7–39. doi: 10.1177/026765839601200102

Van der Slik, F. W. (2010). Acquisition of Dutch as a second language: the explanative power of cognate and genetic linguistic distance measures for 11 west European first languages. *Stud. Second. Lang. Acquis.* 32, 401–432. doi: 10.1017/S0272263110000021

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Mental Process*. Cambridge, MA: Harvard University Press.

Wang, L., and Cui, C. (2018). 基于语言障碍指数的"一带一路"语言服务难度评级研究 [a study on the difficulty rating of "one belt, one road" language services based on the language barrier index]. 中国翻译 *[Chin. Transl. J.]* 2, 72–77.

Wang, X., and Yang, H. (2019). 区域国别视角下语言距离对中国向西亚北非出口贸易的影响及潜力分析 [the impact of language distance on china's export trade to west asia and North Africa and its potential in perspective of "area and country"]. 上海对外经贸大学学报 *[J. Shanghai Univ. Int. Bus. Econ.]* 2, 99–108. doi: 10.16060/j.cnki.issn2095-8072.2019.02.009

Wang, Z., and Hao, Y. (2014). 母语环境下美国汉语学习者心理词汇发展的实证研究 [an empirical study on. The development of mental lexicon by CSL learners in native language environment]. 语言教学与研究 *[Lang. Teach. Linguist. Stud.]* 1, 29–37.

Wu, H., and Zhao, Y. (2018). 英语, 韩语母语者汉语否定标记习得研究 [L2 Chinese acquisition of negative markers by English-and Korean-speaking]. 世界汉语教学 *[Chin. Teach. World]* 2, 256–269. doi: 10.13724/j.cnki.ctiw.2018.02.012

Wu, J., and Mo, X. (2018). 语言类型学视角下外国留学生的 "把" 字句习得研究 [the acquisition of Ba-construction by CFL learners: a typological perspective]. 汉语学习 *[Chin. Lang. Learn.]* 5, 77–85.

Wu, M., Gao, D., Xiao, X., and Zhang, R. (2006). 欧美韩日学生汉字认读与书写习得研究 [a study of learning, reading and writing Chinese characters by CSL learners from Korea, Japan, Western countries]. 语言教学与研究 *[Lang. Teach. Linguist. Stud.]* 6, 64–71.

Wu, J. (2017). 英语母语者汉语书面语动态发展个案研究 [a longitudinal study of CSL writing by an English native speaker]. 现代外语 *[Mod. Foreig. Lang.]* 2 254–264.

Xing, H. (2013). 词语搭配知识与二语词汇习得研究[collocation knowledge and second language lexical. Acquisition]. 语言文字应用*[Appl. Linguist.]* 4, 117–126. doi: 10.16499/j.cnki.1003-5397.2013.04.023

Xu, J., and Zi, Z. (2020). 语言的经济价值分析: 基于语言与中国对外服务贸易的实证研究[an analysis of the language economic value: an empirical study based on China's foreign trade in services]. 解放军外国语学院学报 *[J. PLA Univ. Foreign Lang.]* 3, 66–73.

Xu, Q., Cai, J., and Liu, Z. (2013). 语言迁移研究近20年的新发展: 回顾与思考 [new developments of language transfer studies over the last 20years: review and reflections]. 外语学刊 *[Foreign Lang. Res.]* 1, 103–109. doi: 10.16263/j.cnki.23-1071/h.2013.01.014

Xu, Y. (2014). 复合词结构类型对词义猜测的影响 [the influence of compound word structure on words meaning guessing]. 语言教学与研究 *[Lang. Teach. Linguist. Stud.]* 4, 17–23.

Xu, Y., Chang, L. Y., Zhang, J., and Perfetti, C. A. (2013). Reading, writing, and animation in character learning in Chinese as a foreign language. *Foreign Lang. Ann.* 46, 423–444. doi: 10.1111/flan.12040

Yamashita, J., and Jiang, N. A. N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Q.* 44, 647–668. doi: 10.5054/tq.2010.235998

Yan, M., Richter, E. M., Shu, H., and Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychon. Bull. Rev.* 16, 561–566. doi: 10.3758/PBR.16.3.561

Yan, M., Zhou, W., Shu, H., and Kliegl, R. (2012). Lexical and sublexical semantic preview benefits in Chinese reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1069–1075. doi: 10.1037/a0026935

Yeh, S. L., Chou, W. L., and Ho, P. (2017). Lexical processing of Chinese sub-character components: semantic activation of phonetic radicals as revealed by the Stroop effect. *Sci. Rep.* 7:15782. doi: 10.1038/s41598-017-15536-w

Yeh, S. L., Li, J. L., Takeuchi, T., Sun, V., and Liu, W. R. (2003). The role of learning experience on the perceptual organization of Chinese characters. *Vis. Cogn.* 10, 729–764. doi: 10.1080/13506280344000077

Yuan, B. (1998). Interpretation of binding and orientation of the Chinese reflexive ziji by English and Japanese speakers. *Second. Lang. Res.* 14, 324–340. doi: 10.1191/026765898670904111

Yuan, B. (2004). Negation in French–Chinese, German–Chinese and English–Chinese interlanguages. *Trans. Philol. Soc.* 102, 169–197. doi: 10.1111/j.0079-1636.2004.00134.x

Yuan, B., and Zhao, Y. (2005). Resumptive pronouns in English-Chinese and Arabic-Chinese interlanguages. *Int. Rev. Appl. Linguist Lang. Teach.* 43, 219–237. doi: 10.1515/iral.2005.43.3.219

Zar, J. H. (2010). *Biostatistical Analysis*, *5th Edn*. Upper Saddle River, NJ: Pearson Education.

Zareva, A., Schwanenflugel, P., and Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: variable sensitivity. *Stud. Second. Lang. Acquis.* 27, 567–595. doi: 10.1017/S0272263105050254

Zhang, H. (2006). 外国学生汉语词汇学习状况计量研究 [a quantitative study of vocabulary development of. Foreign students learning Chinese as a foreign language]. 世界汉语教学 *[Chin. Teach. World]* 1, 70–76.

Zhang, H. (2017). The influence of L1 background and other meta-linguistic and background variables on the learning of pinyin and Hanzi by Arabic and English learners of Chinese as a second language Doctoral dissertation University of York.

Zhang, H., and Roberts, L. (2021). The influence of L1 script directionality and L2 proficiency on Hanzi learning among Arabic and English learners of L2 Chinese. *Int. Rev. Appl. Linguist. Lang. Teach.* 4, 1039–1062. doi: 10.1515/iral-2018.0330

Zhang, J. (2016). 东南亚国家华裔、非华裔学生汉字正字法意识的对比研究 [a comparative study of Chinese character orthographic awareness of overseas learners with Chinese origin and non-Chinese origin in Southeast Asia]. 华文教学与研究 *[TCSOL Stud.]* 3, 17–26. doi: 10.16131/j.cnki.cn44-1669/g4.20160526.003

Zhang, J., Meng, H., and Liu, W. (2011). 汉语第二语言学习者单音多义词习得深度研究—以动词打为例 [the depth of monosyllabic polyseme acquisition of learners who speak Chinese as the second language—the case of da]. 语言文字应用 *[Appl. Linguist]*. 1, 112–121. doi: 10.16499/j.cnki.1003-5397.2011.01.016

Zhang, L. (2021). Existence and consequences of dormant features: evidence from L2 acquisition of Chinese object ellipsis. *Second. Lang. Res.* doi: 10.1177/02676583211066293

Zhang, Z. S. (2019). "Visualizing stylistic differences in Chinese synonyms," in *Computational and Corpus Approaches to Chinese Language Learning*. eds. X. Lu and B. Chen (Singapore: Springer), 145–172.

Zhu, D. (1985). 现代汉语语法研究 *[Studies on Modern Chinese Grammar]*. Beijing: The Commercial Press.

Zobl, H. (1983). Markedness and the projection problem. *Lang. Learn.* 33, 293–313. doi: 10.1111/j.1467-1770.1983.tb00543.x

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 18 January 2023
DOI 10.3389/fpsyg.2022.1039461

EDITED BY
Linjun Zhang,
Peking University,
China

REVIEWED BY
Yiu-Kei Tsang,
Hong Kong Baptist University,
Hong Kong SAR, China
Yu Li,
United International College,
China

*CORRESPONDENCE
Catherine McBride
cmcbridca@purdue.edu

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 08 September 2022
ACCEPTED 16 November 2022
PUBLISHED 18 January 2023

CITATION
Zhou Y and McBride C (2023) Invented spelling in English and pinyin in multilingual L1 and L2 Cantonese Chinese speaking children in Hong Kong.
Front. Psychol. 13:1039461.
doi: 10.3389/fpsyg.2022.1039461

Invented spelling in English and pinyin in multilingual L1 and L2 Cantonese Chinese speaking children in Hong Kong

Yanling Zhou[1] and Catherine McBride[2]*

[1]Department of Early Childhood Education, The Education University of Hong Kong, Hong Kong SAR, China, [2]Department of Human Development and Family Studies, Purdue University, West Lafayette, IN, United States

This research examined the relations among Cantonese phonological awareness, invented spelling in Pinyin (in Mandarin), and invented English spelling in 29 first language (L1) and 34s language (L2) Cantonese-speaking second and third graders in Hong Kong. The purpose of this study was to understand how phonological awareness skills across languages are associated in multilinguals. We compared the phonological skills in the two groups (i.e., L1 and L2 Chinese speaking children) for the three official languages (i.e., Cantonese, Mandarin, and English) spoken in Hong Kong. The two groups did not differ on Cantonese phonological awareness, Mandarin Pinyin invented spelling, or English invented spelling, but the L1 group performed significantly better than the L2 group on Mandarin Pinyin tone skills, with non-verbal intelligence and grade level statistically controlled. In both groups, all three of the phonological sensitivity measures were significantly correlated with one another. With group, grade, and nonverbal IQ statistically controlled, only Mandarin Pinyin invented spelling but not Cantonese phonological awareness uniquely explained English invented spelling performance. In contrast, Pinyin invented spelling was uniquely explained by both English invented spelling and Cantonese phonological awareness skills. Results highlight some phonological transfer effects across languages.

KEYWORDS

invented pinyin spelling, invented English spelling, Cantonese phonological awareness skills, L1 Cantonese speaking children, L2 Cantonese speaking children

Introduction

How are phonological awareness skills across different languages interrelated for multi-literacy acquisition in multilingual speakers? Nowadays, more and more children acquire multiple languages from early on across the globe. Ample research has supported the phenomenon of phonological skill transfer for linguistically close language pairs (e.g., Durgunoğlu et al., 1993), as well as linguistically distinct bilingual learning (e.g., Yang et al., 2017). These results seem to support Cummins' interdependence hypothesis of language transfer (Cummins, 1989), i.e., that phonological transfer is reciprocated among the languages within subject. However, most previous research has targeted L2 learners who acquire English or other alphabetic languages as an L2 (e.g., French, German, or Spanish). Despite a growing population of children whose first language is alphabetic and who acquire Chinese as an L2 (Zhou and McBride, 2018), little is understood about how such phonological transfer may affect such learners. In the context of Hong Kong, ethnic minority children represented by diverse alphabetic or Akshara speakers (e.g., Urdu, Hindi, Cebuano, etc.) lack a rich Chinese language and literacy environment at home. Yet they are expected, in the same way as their L1 Chinese speaking peers, to acquire Mandarin Chinese as a second or third language. Specifically, they use Cantonese, the home language of most families in Hong Kong, as the spoken language at school, but also learn Mandarin Chinese as an L3 in spoken and written forms, along with their Cantonese-speaking peers. In the present study, we aimed to understand how phonological skills in three languages are intercorrelated for L1 and L2 Chinese-speaking children. The three languages tested were Cantonese Chinese, Mandarin Chinese, and English. For one group of participants, all three of these languages were non-native languages, leading to intriguing questions in relation to overlap in phonological processes among them.

There were two primary focuses in this study: First, we examined how phonological awareness skills through a typical receptive phonological awareness task or an invented spelling task across the three languages, respectively, are correlated within the same children. Second, the study provides some insights into similarities and differences of such intercorrelations across L1 and L2 Chinese-speaking children.

Phonological awareness skills and invented spelling in English

Invented spelling is a process by which a speller applies basic or partial conventional spelling rules for a sound representation that is auditorily perceived. As the term 'invented spelling' was coined originally or at least primarily for spelling in English, invented spelling in English captures children's phonological awareness skills at the syllable and phoneme levels (e.g., Tangel and Blachman, 1992) and often serves as a gateway for reading and spelling, integrating phonological and orthographic knowledge (Invernizzi et al., 1994; McBride-Chang, 1998; Martins and Silva, 2006; Ouellette and Sénéchal, 2008; Ouellette et al., 2017; Treiman, 2018). Children's invented spelling in English reveals the level of their understanding and creative mastery of phoneme-grapheme mapping rules (Gentry, 1982; Bear and Templeton, 1998) given the relative opaqueness of the English orthography as compared with other transparent alphabetic orthographies such as German or French. Therefore, early invented spelling is seen as a significant predictor of subsequent reading and writing abilities (Graham and Santangelo, 2014; Ouellette and Sénéchal, 2017). Invented spelling in English typically is regarded as an optimal tool for understanding children's expressive phonological awareness skills because it requires children to demonstrate in writing explicitly how phonological representation is stored in their mind. It is also very useful for understanding lexical tone processing in Mandarin, since tonal notations are part of the Pinyin spelling process for this language. Invented spelling helps to indicate children's progress and development in moving from experimentation of word representation in print to a later incorporation of orthographic accuracy (Richgels, 1995), rather than a pure memorization process of conventional spellings.

Mandarin pinyin as a phonological facilitation in Chinese character acquisition

Similarly, invented Pinyin spelling is an effective tool to assess children's explicit Chinese phonological awareness skills, and a strong facilitator for L1 Mandarin speaking children's Chinese character reading skills (Lin et al., 2010; Ding et al., 2015, 2018) and L2 learners of Chinese (Ju et al., 2021; Zhang and Roberts, 2021). It captures syllable, phoneme, and tone knowledge (Lin et al., 2010). The pinyin system is a fabricated phonological system that was designed and implemented in the 1950s to facilitate early Mandarin Chinese literacy acquisition in China in order to address the issue of a relatively high rate of illiteracy (over 80%) in the country at the time (Zhou, 2013). The pinyin system is based on the Roman alphabetic system utilizing 26 letters representing consonants and vowels that map with Mandarin Chinese sounds. The morphosyllabic system of Chinese dictates that each character is an orthographic representation of a syllable; the syllable also represents a phonological unit. For majority of Chinese characters, there is a fixed onset-rime phonological structure which is usually attached to a lexical tone.

Children from all over China start to learn Pinyin in first grade at around the age of 6, for 10 weeks consecutively. In Mainland China, the school language is consistently Mandarin Chinese (Shu et al., 2003). That means that no matter what Chinese language or dialect children are speaking at home with their parents (e.g., Cantonese or Shanghainese, or Changsha Hua), students are required to speak and use Mandarin at school with their peers and teachers. In Hong Kong, the school language is usually Cantonese Chinese although the written language (vocabulary and grammar) that children need to acquire is in Mandarin Chinese (Cheung and Ng, 2003), since Mandarin Chinese is one of the three official languages and two literacies (Cantonese, English, and Mandarin) in Hong Kong.

Pinyin is one of several effective ways of learning Chinese characters (Packard, 1990; Chung, 2007; Everson, 2009; Poole and Sung, 2015; Zhang et al., 2017) together with other character learning methods such as writing focused (Guan et al., 2011) and character recognition focused approaches (Poole and Sung, 2015). However, In Hong Kong, children often begin to learn to read and to write Chinese characters and English words without any systematic phonological instruction at age 3.5 years as a cultural practice. Chinese characters are typically mapped to the Cantonese language for young children. Mandarin, another Chinese language, is also taught early in school in Hong Kong as a second language. However, pinyin, which is explicitly mapped only to Mandarin and not to Cantonese, is typically only taught as a part of the language arts subject beginning in primary school, several years after the initiation of Chinese literacy teaching, in Hong Kong.

The similarities between the pinyin system and English spelling

Pinyin and English spelling share substantial commonalities and basic onset-rime mechanisms across their respective notation systems. There are 21 pinyin onsets, and some look the same as English consonants (e.g., b, p, f, and k). In contrast, a few pinyin onsets may look somewhat unusual in comparison to the English orthography (e.g., zh). Some sound representations are similar to those in English (e.g., d, t, s, and z), and some sound very different for the two languages (e.g., x, q, ch, and sh). There are 36 pinyin rimes, including rimes with a single vowel phoneme rime (e.g., a, o, and e), a double vowel rime (e.g., ai, ei, ao, and iong) or a rime that ends with a nasal sound (e.g., in, ing). Moreover, pinyin spelling includes a representation of lexical tone, with four basic and one neutral tones possible. Each Chinese character has its own tone attached to the syllable. Thus, each Chinese character represents a holistic syllable. To summarize, phonologically, each Chinese syllable comprises an onset and a rime, as well as a lexical tone, only; there are no consonant clusters in Chinese, as there are in English. The impact of English instruction on Pinyin spelling.

Previous research has demonstrated how acquiring English as a L2 strengthens native Chinese-speaking children's pinyin skills (Chen et al., 2010; Ding et al., 2018). Ding et al. (2018) found that although invented Pinyin did not explain English conventional spelling in Mandarin-speaking children in Mainland China, their English invented spelling measure did explain unique variance in both their Chinese character writing and conventional English spelling performance. These researchers interpreted the English invented spelling findings as representing a linguistic transfer effect. Invented spelling indeed allows children the freedom to access their implicit knowledge of phonological representations of language or different languages in recoding sound patterns. Chen et al. (2010) found that learning English accelerates children's pinyin skills. This research suggests a positive one-way transfer from English instruction to pinyin learning.

Up until now, research has primarily focused on native Chinese-speaking children's pinyin knowledge. In contrast, very little is understood about the extent in which there is a transfer effect in multilingual children who are exposed to and learning Chinese as a second or third language (L2 or L3). In the present study, we aimed to obtain a comparative view of how phonological awareness skills in three languages might be associated among children living in Hong Kong. All of these children speak Cantonese as either L1 or L2. This focus can shed light on how multilingual children potentially organize their phonological system in the processes of multiple language and literacy acquisition. We examined phonological sensitivity in the form of Cantonese phonological awareness skills, as well as invented spelling in both pinyin (representing Mandarin, a prominent Chinese language) and English. As discussed below, invented spelling is a particularly sensitive way in which to capture phonological sensitivity (e.g., Mann et al., 1987; Lin et al., 2010). However, there is no such standard phonological written system available to, or at least common among, Cantonese-speaking children. Thus, our phonological sensitivity measure in Cantonese differed from the invented spelling measures in Mandarin and English, respectively.

How does mastery of the pinyin system affect Chinese-English biliteracy learning?

Teachers sometimes worry that the two systems, which are written using similar notations (i.e., most of the Roman alphabet), could potentially interfere with one another in the learning process. Among scholars focused on native Chinese children's learning, however, there has been a stronger argument that not only does pinyin learning not interfere with the acquisition of English learning (Lü, 2017), but it may actually have a positive effect on phonological awareness skills (e.g., McBride-Chang et al., 2004; Yin et al., 2011). For example, McBride-Chang et al. (2004) speculated that Beijing children, who typically perform significantly better on phonological skills at the syllable and phoneme levels than their Hong Kong counterparts, had an advantage precisely because of the Beijing children's pinyin knowledge. Similarly, Cheung et al. (2001) compared phonological awareness skills between Cantonese speaking children in Guangzhou, China, and Hong Kong. They found a possible advantage resulting from pinyin training in the Guangzhou children on onset, rime, and coda analyses compared with their peers in Hong Kong. In the present study, we compared pinyin knowledge in two sets of L2 Mandarin Chinese learners. One group of children are L1 Cantonese Chinese speakers, and the other group represents ethnic minorities in Hong Kong who typically speak a variety of languages used in India, Pakistan, Nepal, and surrounding areas as their home languages. The Hong Kong Chinese group also speaks a tonal language as their first language (L1), but the ethnic minority group typically speaks a non-tonal language as their L1.

Lexical tone as the Chinese phonological 'twist'

Indeed, lexical tone sensitivity is fundamental for Chinese phonological awareness. Previous studies have highlighted the importance and close associations between tone awareness and both L1 (So and Siegel, 1997; Fu and Huang, 2000; Leong et al., 2005; Wang et al., 2005; Shu et al., 2008; Li et al., 2012; Tong et al., 2015) and L2 (Zhou and McBride, 2018) Chinese word recognition. Yin et al. (2011) argued that the stronger the pinyin knowledge, the stronger the tone awareness skills. Lexical tones in Mandarin Chinese are particularly difficult for learning Chinese as a foreign language. Yet tone sensitivity is essential for Chinese character learning given the many homophones of Chinese and meaning changes in Chinese with lexical tone changes. Zhou and McBride (2018) demonstrated that lexical tone was the weakest area of pinyin knowledge in L2 Chinese-speaking children, even when these children had been in an immersion Chinese-English classroom together with their L1 Chinese speaking peers for more than 4 years.

The present study compared native Cantonese speaking children with ethnic minority children learning Cantonese, Mandarin, and English all as second, or foreign languages. L2 Chinese speaking children in Hong Kong constitute approximately 6.4% of the Hong Kong children population (Census and Statistics Department, HKSAR, 2011). These students encounter great challenges in language and literacy development in Hong Kong public schools (Arat et al., 2016; Zhou et al., 2018). However, these children are expected to acquire Chinese and English language and literacy skills alongside their L1 Chinese speaking peers. We tested (a) the extent to which the two groups differed in phonological sensitivity across the three languages of Cantonese (*via* a phonological awareness task), Mandarin (*via* an invented Pinyin knowledge task), and English (*via* an invented spelling task) and (b) how these phonological sensitivity measures explained invented spelling in Pinyin and English, respectively.

Materials and methods

Participants

Second and third grade local primary school children were recruited for this study. There were 29 Cantonese Chinese-speaking children (L1) including 16 boys and 13 girls (Mean age = 93.4 months SD = 9.37 months) and 34 s language Chinese speaking ethnic minority children (L2) including 17 boys and 17 girls (Mean age = 94.0 months, SD = 5.35 months). The L1 Chinese speaking group comprised homogenous Cantonese speaking children with a Cantonese speaking home language and literacy background. Among the L2 group, the ethnicity backgrounds of the children included 14 Indian, 10 Nepalese and 10 Pakistani, respectively. For these children, Cantonese was the medium of instruction at school, English was taught as a foreign language, and Mandarin was learned and used for Chinese language arts classes in school.

Procedure

Upon receiving parental consent, we worked with the school on scheduling testing sessions for individual children. Trained psychology undergraduate students and research assistants were employed for the data collection. They administered individual testing to the children at school or at after-school centers. Both L1 and L2 Chinese speaking children were divided into two groups, with some of the students beginning with Chinese tasks first and the remaining starting with English tasks for counterbalancing.

Measures

Raven's standard progressive matrices

Raven's Colored Progressive Matrices B-C was used to evaluate the non-verbal reasoning of children (Raven et al., 1995). Twelve incomplete geometric figures were presented to the children one by one. For each question, children were asked to choose one pattern from among 6–8 choices in order to make the figure complete. One point was given to each correct answer and the total score for this test was 12.

Chinese word reading

This test was composed of 211-items. Twenty-seven of them were single Chinese characters, 34 of them were two-character Chinese words which were frequently used, and the remaining 150 items were two-character Chinese words from the subset of The Hong Kong Test of Specific Learning Difficulties in Reading and Writing (Ho et al., 2000). Previous research has also adopted a similar measure to access the Chinese word reading ability of children (e.g., McBride-Chang et al., 2003; Zhou and McBride, 2018). Despite the number of characters for each item, one point was given to children who answered correctly for an item. Thus, the total points possible for this task was 211.

English word reading

Previous research studies had developed a task for examining the English reading ability of children (Tong and McBride-Chang, 2010; Zhou and McBride, 2018). Due to a previous observation of some likelihood that L2 Chinese speaking children may have higher proficiency in English, for the present study, we expanded the word list based on a previous measure adopted (Zhou and McBride, 2018). According to the children's reading ability, difficulties of word meaning and phoneme combinations, sixty-three words were selected to assess the English word reading ability of children. One point was given for each correct pronunciation, so the maximum points for this test was 63. The test was terminated immediately when the children mispronounced four consecutive words.

Chinese vocabulary knowledge

The test was divided into three subtasks to assess the knowledge of receptive vocabulary, expressive vocabulary, and vocabulary definitions of children. All the items in this section were consistent with the children's age of acquisition and ordered according to ascending difficulty. We tested vocabulary knowledge orally only, as a point of comparison to understand oral language skills. For receptive vocabulary, there were 21 items and the experimenter presented four pictures to children for each item. Children were required to choose a picture from among the four choices which best matched the meaning of the Chinese word that was verbally presented by the experimenter in Cantonese. One point was given for each correct response. For expressive vocabulary, there were 23 items and all the pictures were obtained from the Peabody Picture Vocabulary Test-Third Edition (PPVT-III; Dunn and Dunn, 1997). Children were required to name the object or situation in Cantonese corresponding to the picture shown. One point was given for each correct response. For vocabulary definitions, there were 27 items. The experimenter spoke a word in Cantonese to children and children were required to provide the definition of this word. Children were given 0–2 points per item based on the precision of their answer. If children answered five consecutive questions incorrectly in one of the subtasks, the experimenter stopped and returned to the remaining subtasks of the test. The total score for this test was 98.

English vocabulary knowledge

Similar to the procedure described above for Chinese vocabulary knowledge testing, tasks of receptive vocabulary, expressive vocabulary and vocabulary definitions were included in the test of English vocabulary knowledge. There were 21, 23, and 26 items in each subtask, respectively. The words for this test were extracted from the PPVT-III (Dunn and Dunn, 1997). The order of the words was based on the familiarity and difficulty levels. The English vocabulary knowledge test adopted the same scoring criteria that were used for Chinese vocabulary knowledge. The total score for this test was 96.

Cantonese phonological awareness

Items tapping syllable deletion and onset deletion were the measures used to assess children's Cantonese phonological awareness. Previous studies (McBride-Chang et al., 2005; Cheung et al., 2009) had adopted these two subtasks to assess the Cantonese phonological awareness of children. All the stimuli in this test were presented verbally. The syllable deletion task consisted of 29 items, and 15 of them were three-syllable real words while the remaining were pseudowords. Children were asked to drop one of the syllables of a word and then say what was left after removing that syllable. For instance, children were asked to say 漢堡包 /hon3/ /bou2/ /baau1/ without/baau1/. So, children needed to say漢堡 /hon3/ /bou2/ in order to receive one point. The onset deletion task consisted of 22 items, and ten of them were one-syllable real words while the remaining were one-syllable pseudowords. Children were asked to delete the first consonant of the word in each item and say the word without the initial sound. For instance, children needed to reply with /i1/ when they were told to remove the consonant of 詩 /si1/. One point was given for each correct response. Children started at the level consistent with their own grade level, and both subtasks were terminated when they reached the limit of the stopping rule. The total score for this test was 51.

Pinyin invented spelling (mandarin)

A previous study had developed this task for examining the Mandarin Pinyin invented spelling of children (Lin et al., 2010). For the present study, we also adopted a similar measure and scoring scheme. There were ten double syllable real pinyin words in the test and children were required to write down the Pinyin representation of the words that were verbally presented by the experimenter. Please see the Appendix for the list. As the verbally presented pinyin words (e.g., táng dài) were presented in the absence of a given context or sentence, these were conceptualized as appropriate for this task, since the items could be recognized as homophones of different words. With this approach, the Pinyin items were not too novel, reducing the cognitive load in processing unfamiliar sound combinations, but the items were also challenging. Relatively few words were written 100% correctly. For each pinyin syllable a child wrote down, we adopted scales for the onset (0–4 points), rime (0–4 points), lexical tones (0–3 points) and the order of Pinyin (0–1point). For scales for onset and rime, the scores were given based on the closeness of the written pinyin letter to the targeted pinyin. For example, 0 points were given if nothing was written; 1 point was given if a random letter was written down; 2 points were given if the component included a letter that was exactly or close to the target onset or rime, but had multiple irrelevant letters; 3 points were given if the component had a phonologically similar letter but not the accurate pinyin letter; 4 points were given if the component had exactly the targeted pinyin letter or letters. For the lexical tone evaluation, the correct tone was coded as 3 points and incorrect tones were coded as 1, but if nothing was written to indicate the tone, then 0 points were given. For each pinyin syllable, the maximum score was 12, so across the 20 Pinyin syllables altogether, the total possible score was 240. Based on this coding system, two variables were created, namely, (1) invented pinyin spelling in total that was inclusive of lexical tone score, and (2) the invented pinyin tone score alone.

English invented spelling

Unlike the English word dictation task, 8 pseudo words with single syllables were constructed, some of them containing digraphs and double consonants, for this task. Children were required to write the pseudo words when they heard the sounds. The scoring procedure was adopted from two previous studies on a 0–5 point scale (Mann et al., 1987; Tangel and Blachman, 1992). Children received 0 points when they did not make any response or answered the question randomly. One point and two points were allotted for each item on which children only correctly wrote the phonetically related letter(s) and initial letter, respectively; three points were given for items written in a pre-conventional manner with at least one phoneme correct; four points were given for items for which all phonemes were correctly written but in a pre-conventional manner; and five points were given for each item for which the pseudo word was correct. Thus, the total score for this test was 40.

Data analysis

A two-step statistical analysis was carried out. First, MANCOVAs were employed to determine whether there were differences between L1 and L2 groups on Chinese and English word reading related tasks. The result of Box's test did not allow us to reject the null hypothesis of equal covariance matrices for the MANCOVA tests. We then ran a hierarchical regression to test how phonological awareness skills in the other two languages predicted English invented spelling and Pinyin invented spelling for L1 and L2 groups similarly and differently. We computed dummy variables setting L1 as the reference group and L2 as the comparison group in order to examine the similarities and differences within the same model.

Results

Multiple ANCOVAs were carried out to examine the group differences on Chinese and English word reading related measures after controlling for non-verbal intelligence and grade. There was a significant major group effect, $F(1,8) = 36.6$, $p < 0.001$. As found previously (e.g., Zhou et al., 2018; Zhou and McBride, 2018), as the test of pairwise comparisons shows in Table 1, L2 Chinese speaking children demonstrated a clear advantage in English. They performed significantly better on English word reading and vocabulary knowledge than did the L1 Chinese speaking children. L1 Chinese speaking children demonstrated an advantage in Chinese word reading and vocabulary knowledge, compared with L2 Chinese speaking children. However, the two groups performed similarly on all phonological tasks across the three languages, including Cantonese phonological awareness skills and overall invented Mandarin Pinyin skills; the only difference was Chinese lexical tones. L1 Chinese speaking children performed significantly better than the L2 children in Mandarin tone awareness skills. Therefore, it appears that lexical tone in L2 Chinese speaking children who learn Chinese as a L2 is the weakest for acquiring Chinese phonological awareness skills, echoing previous findings for L2 Chinese speaking children whose first languages are nontonal (Zhou and McBride, 2018; Table 2).

The correlations shown in Table 3 indicate that after statistically controlling for grade and non-verbal intelligence, Chinese word reading was significantly associated with English word reading skills ($r = 0.54$, $p < 0.01$), Cantonese phonological awareness skills ($r = 0.54$, $p < 0.01$), Pinyin invented spelling skills ($r = 0.38$, $p < 0.05$), and English invented spelling skills ($r = 0.41$, $p < 0.05$) for L2 Chinese speaking children. However, L1 Chinese speaking children's Chinese word reading skills was negatively associated with their English word reading ($r = -0.24$, $p < 0.05$), and unassociated with phonological sensitivity in all three languages. English word reading was also significantly associated with Mandarin Pinyin invented spelling skills ($r = 0.54$, $p < 0.01$) and English invented spelling skills ($r = 0.71$, $p < 0.01$), but not with Cantonese phonological awareness skills in the L1 Chinese-speaking group.

For the L2 group, English word reading skills were significantly associated with phonological skills in all three languages [i.e., Mandarin Pinyin invented spelling skills ($r = 0.66$, $p < 0.01$), English invented spelling skills ($r = 0.74$, $p < 0.01$) and Cantonese phonological awareness skills ($r = 0.52$, $p < 0.01$)].

Phonological skills in the three languages were significantly intercorrelated in both the L1 and L2 groups. English invented spelling skills were significantly associated with Cantonese phonological awareness skills (L1: $r = 0.40$, $p < 0.05$; L2: $r = 0.51$, $p < 0.01$) and Mandarin Pinyin invented spelling skills (L1: $r = 0.48$, $p < 0.01$; L2: $r = 0.55$, $p < 0.01$). Cantonese phonological awareness skills were significantly associated with Mandarin Pinyin invented spelling skills (L1: $r = 0.55$, $p < 0.01$; L2: $r = 0.45$, $p < 0.05$).

Hierarchical regression analyses were carried out to examine how similarly and differently phonological skills in the three languages predicted the English invented spelling skills and Mandarin Pinyin Invented spelling skills for the L1 and L2 groups. We computed a dummy variable and set L1 as the reference group and L2 as the comparison group given the above-mentioned correlational differences in the groups. As shown in Table 3, the regression models revealed that for English invented spelling skills, at step 2 only the Cantonese phonological awareness skills task was a significant predictor for English invented spelling skills for both

TABLE 1 Descriptive statistics of L2 Chinese speaking and L1 Chinese speaking students' performance on Chinese and English related tasks after controlling for nonverbal intelligence and grades.

Variables	L2 ($N = 34$)		L1 ($N = 29$)		Pairwise comparison mean differences (L2–L1)	Reliability coefficiency
	Mean	SD	Mean	SD		
CWR	26.09	19.12	128.52	33.06	−95.62***	0.99
EWR	39.77	14.83	29.46	17.93	11.50*	0.98
CV	17.28	9.55	53.03	13.31	−33.79***	0.97
EV	46.39	10.13	34.97	18.76	13.44**	0.95
CPAS	28.82	11.01	35.69	11.10	−4.63	0.97
MPIST	30.88	12.73	46.67	10.10	−9.29**	0.96
MPIS	123.21	42.43	143.37	54.69	−4.59	0.96
EIS	18.06	5.67	17.59	6.70	1.92	0.78

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

TABLE 2 Partial correlation between all the literacy related measures in L2 Chinese speaking and L1 Chinese speaking children after controlling for nonverbal intelligence and grade (L2_Left bottom up triangle/ L1_ right corner down triangle).

Variables	1	2	3	4	5	6	7	8
1. Chinese Word Reading	–	−0.24	0.70***	−0.18	−0.07	0.26	−0.04	−0.17
2. English World Reading	0.54**	–	−0.21	0.80***	0.35	0.32	0.54**	0.71***
3. Chinese Vocabulary	0.59**	0.47**	–	−0.21	0.14	0.27	0.12	−0.08
4. English Vocabulary	0.22	0.53**	0.30	–	0.42*	0.19	0.53**	0.64***
5. Cantonese PA	0.54**	0.52**	0.41*	0.09	–	0.30	0.55**	0.40*
6. M-Pinyin Invented Spelling _Tone	0.12	0.35*	0.11	0.03	0.10	–	0.70***	0.32
7. M-Pinyin Invented Spelling Total	0.38*	0.66***	0.26	0.39	0.45*	0.60***	–	0.48**
8. English Invented Spelling	0.41*	0.74***	0.45***	0.28	0.51**	0.33	0.55***	–

groups without considering Mandarin Pinyin invented spelling, suggesting a moderate transfer effect. After introducing Mandarin Pinyin invented spelling skills and the dummy variable of Mandarin Pinyin invented spelling skills to the regression model at step 3, only the Pinyin invented spelling skills task was a significant predictor for English invented spelling skills for both groups. For Mandarin Pinyin invented spelling skills, Cantonese phonological awareness skill was a significant predictor for both groups. English invented spelling skills and its dummy variable were entered at Stage three of the regression. Both Cantonese phonological awareness skills and English invented spelling skills were significant predictors of Mandarin Pinyin invented spelling skills, which also highlights transfer of phonological awareness skills of both Cantonese and English in learning Mandarin Pinyin skills. The two groups did not differ on how phonological awareness skills in the three languages predicted English invented spelling skills and Mandarin Pinyin invented spelling skills, respectively (Table 4).

Discussion

The present study extended our understanding of the significant correlations of phonological skills in three different languages (Mandarin, Cantonese, English) in multilingual children across at least at two levels: First, the significant positive correlations across measures of phonological sensitivity suggest a kind of shared phonological mechanism that supports languages that are closely related to each other (i.e., Mandarin and Cantonese) and the languages that are distinct from each other (i.e., Mandarin/ Cantonese Chinese and English). Our results showed that strong phonological skills in oral Cantonese Chinese were associated with strong phonological skills in Mandarin and English, respectively. Pinyin knowledge as an acquired alphabetic system that is intended to support Chinese character recognition has a positive 'side effect' on English phonological awareness skill. In turn, strong awareness of English phono-grapheme rules as indicated *via* invented English spelling was positively associated with invented Pinyin spelling.

Second, this pattern was demonstrated not only among L1 Chinese speaking children but also for L2 learners. Although L1 Chinese speaking children demonstrated a clear advantage in pinyin spelling, the patterns for both groups demonstrated that at least for older children, even for children whose first language is different from these two languages, the pinyin and English invented spellings are two distinct alphabetic notation systems, though they are relatively strongly correlated. This is perhaps particularly important for L2 language Chinese young learners whose first languages are neither Chinese nor English. Pinyin learning can perhaps serve as a bridge for them in acquiring Mandarin Chinese phonological awareness skills. Similarly, for Cantonese speaking children who learn Mandarin Chinese as a written language, not only can pinyin knowledge possibly influence their English spelling skills, and, hence, their phonological awareness skills, but it also relates significantly to their L1 Cantonese phonological skills. Given that there is no official agreed upon Cantonese phonetic system taught at school, Mandarin pinyin skills can potentially have a positive effect for Chinese literacy acquisition. One of the limitations of the present study is that the examination of Cantonese phonological awareness skills was done verbally and Cantonese lexical tones were not included in the present study due to a lack of a coherent system that is explicitly introduced in the education system in Hong Kong. Future studies could explore phonological awareness skills at different levels including lexical tones, using auditory tasks (rather than written tasks) only for the three languages.

A side note from this study that should be highlighted is the fact that, although the total phonological sensitivity measures in Cantonese, Mandarin, and English were all moderately and significantly associated with reading in English for both L1 and L2 Chinese speaking children, none of these were associated with Chinese reading for L1; however, for L2 Chinese speaking subjects, the significant associations with Chinese reading and vocabulary remained moderate to strong. A lack of associations for Chinese word reading in L1 highlights the differing demands of literacy acquisition for Chinese as compared to English (Chung and Ho, 2010). English reading relies substantially on phonology. For L2 populations, similar to the results in a previous study (Zhou and McBride, 2018), the reliance on phonological awareness skills may be the major strategy, though not sufficient enough, to acquire Chinese literacy. In the Hong Kong context, where reading is not taught using a phonological coding system, as it is for the rest of Mainland China,

TABLE 3 Summary of Hierarchical regression predicting English Invented Spelling of L1 and L2 Chinese speaking children.

Variables		Change	*F* change	*β*	*t*
Step 1	0.20	0.20	4.84**		
Group				−0.18	−1.35
Ravens				0.45	3.50***
Grade				0.20	1.63
Step 2	0.37	0.18	7.93***		
Group				−0.43	−1.29
Ravens				0.26	2.09*
Grade				0.05	0.47
Cantonese PA				0.55	3.11*
DummyG_PA				−0.15	0.64
Step 3	0.49	0.12	6.17**		
Group				−0.14	0.39
Ravens				0.20	1.68
Grade				0.14	1.20
Cantonese PA				0.18	1.31
DummyG_PA				−0.01	−0.02
MPinyin Invent Total				0.39	2.25*
DummyG_ MPinyin Invent				0.12	0.32

Cantonese PA = Cantonese Phonological awareness; DummyG_ = L1 as reference group, 0, L2 as comparison group, 1 *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

TABLE 4 Summary of hierarchical regression predicting Mandarin Pinyin Invented Spelling of L1 and L2 Chinese speaking children.

Variables		Change	*F* change	*β*	*t*
Step 1	0.15	0.15	3.41-		
Group				0.02	0.16
Ravens				0.37	2.82**
Grade				−0.03	−0.21
Step 2	0.37	0.23	10.26***		
Group				−0.58	−1.73
Ravens				0.15	1.19
Grade				−0.16	−1.43
Cantonese PA				0.72	4.10***
DummyG_PA				−0.51	−1.55
Step 3	0.49	0.11	6.11**		
Group				−0.41	1.18
Ravens				0.04	0.31
Grade				−0.19	−1.75
Cantonese PA				0.48	2.51*
DummyG_PA				−0.42	−1.14
English Invented Spelling				0.44	2.54*
DummyG_EIS				−0.03	−0.09

Cantonese PA = Cantonese Phonological awareness; DummyG_ = L1 as reference group, 0, L2 as comparison group 1; EIS = English Invented Spelling. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

phonological skills may be required only in a very limited way in order to master word recognition.

Moreover, Mandarin lexical tones are clearly the most difficult aspect of Mandarin pinyin spelling to acquire for ethnic minority L2 Chinese speaking children when compared with L1 Chinese speaking children in Hong Kong. However, the comparable results between the two groups that were found for Cantonese phonological awareness, English invented spelling skills, and Mandarin pinyin, but not in Chinese word reading and vocabulary, were consistent with previous findings (Zhou and McBride, 2018) demonstrating that L1 and L2 Chinese speaking children did not differ on their phonological awareness skills in Chinese and English but did differ significantly on Chinese word reading and vocabulary. Lexical tones, which are absent in many alphabetic languages, are a critical indicator of Chinese vocabulary and word learning (Liu et al., 2010; Zhou and McBride, 2018). The complete absence of lexical tones in English and many other languages renders them more difficult for L2 Chinese speaking children to process, remember, and produce when compared with their L1 peers. The weakness in Mandarin lexical tones in the L2 children in the present study is also at least partly likely attributable to a lack of exposure and language experience in Mandarin Chinese as the school and peer language is mainly Cantonese. Mastering lexical tone may reflect overall oral language skills, which serve as a strong foundation for acquiring Chinese word reading (Zhou and McBride, 2018).

The present study aimed to explore possible transfer effects of phonological skills in English, Mandarin, and Cantonese in L1 and L2 Cantonese speaking children. In this study the identified transfer effect, both as a learning mechanism (Thorndike, 1913; Brown and Kane, 1988) and a linguistic effect (Corder, 1983; Cummins, 1989) enhanced our two understandings about phonological transfer. First, transfer evolves from implicit knowledge of one's native language to explicit knowledge of the L2 or L3 language (Corder, 1983); in the present study, this transfer effect emerged for both L1 and L2 Cantonese speaking children, even though English is a distant language from Mandarin and Cantonese, and words in the latter two languages have simpler phonological structures (i.e., consonant and vowel or onset and rime) than that for English words. Thus, the nature of combining speech sounds together to form individual words applies across all three languages in oral form as well as in the form of invented spellings of English and Mandarin.

Despite the relatively small sample size given various difficulties in recruiting L2 Cantonese learners at this grade level, we were able to identify positive correlations between spelling and phonological awareness in three languages in L1 and L2 Chinese speaking students in Hong Kong. The present study demonstrated that acquisition of Mandarin through the pinyin system as another language for both L1 and ethnic minority children is positively associated with spelling and phonological awareness skills in other languages among multilingual Hong Kong children. Future studies should consider further examining phonological awareness skills using oral tasks across all three languages within this multilingual population. The patterns identified through the present study, which shed light on the transferability of phonological sensitivity in multilingual children, should be further explored in future work.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by The Education University of Hong Kong. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

YZ analyzed the data and drafted the manuscript. CM had the data collected. All authors contributed to the article and approved the submitted version.

Funding

This research project was funded by the Public Policy Research Funding Scheme (Project Number: 2013.A4.001.13A) from the Central Policy Unit of the Hong Kong Special Administrative Region Government and by the General Research Fund of the Hong Kong Special Administrative Region Research Grants Council (Project Number: 14600818 and 18611218).

Acknowledgments

We thank all participating children, and all research assistants and student helpers on the project for their help in data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1039461/full#supplementary-material

References

Arat, G., Hoang, A. P., Jordan, L. P., and Wong, P. W. (2016). A systematic review of studies on ethnic minority youth development in Hong Kong: an application of the ecological framework. *China J. Soc. Work* 9, 218–237. doi: 10.1080/17525098.2017.1254716

Bear, D., and Templeton, S. (1998). Explorations in spelling: foundations for learning and teaching phonics, spelling, and vocabulary. *Read. Teach.* 52, 222–242.

Brown, A. L., and Kane, M. J. (1988). Preschool children can learn to transfer: learning to learn and learning from example. *Cognitive Psychology* 20, 493–523. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3191668

Census and Statistics Department, HKSAR (2011). Thematic report: ethnic minorities. Available at: https://www.censtatd.gov.hk/en/data/stat_report/product/B1120100/att/B11201002016XXXXB0100.pdf (Accessed December 28, 2022).

Chen, X., Xu, F., Nguyen, T.-K., Hong, G., and Wang, Y. (2010). Effects of cross-language transfer on first-language phonological awareness and literacy skills in Chinese children receiving English instruction. *J. Educ. Psychol.* 102, 712–728. doi: 10.1037/a0018802

Cheung, H., Chen, H. C., Lai, C. Y., Wong, O. C., and Hills, M. (2001). The development of phonological awareness: effects of spoken language experience and orthography. *Cognition* 81, 227–241. doi: 10.1016/S0010-0277(01)00136-6

Cheung, H., Chung, K. K., Wong, S. W., McBride-Chang, C., Penney, T. B., and Ho, C. S. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *J. Child Psychol. Psychiatry* 50, 726–733. doi: 10.1111/j.1469-7610.2008.02001.x

Cheung, H., and Ng, L. (2003). "Chinese reading development in some major Chinese societies: an introduction," in *Reading Development in Chinese Children*. eds. C. McBride-Chang and H.-C. Chen (Westport, CT: Praeger), 3–17.

Chung, K. K. (2007). Presentation factors in the learning of Chinese characters: the order and position of Hanyu pinyin and English translations. *Educ. Psychol.* 27, 1–20. doi: 10.1080/01443410601061306

Chung, K. K. H., and Ho, C. S.-H. (2010). Second language learning difficulties in Chinese children with dyslexia: what are the Reading-related cognitive skills that contribute to English and Chinese word Reading? *J. Learn. Disabil.* 43, 195–211. doi: 10.1177/0022219409345018

Corder, S. P. (1983). "A role for the mother tongue," in *Lanugage Transfer in Language Learning*. eds. S. Gass and L. Selinker (Rowley: Newbury House Publishers), 85–97.

Cummins, J. (1989). Language and literacy acquisition in bilingual contexts. *J. Multiling. Multicult. Dev.* 10, 17–31. doi: 10.1080/01434632.1989.9994360

Ding, Y., Liu, R. D., McBride, C. A., Fan, C. H., Xu, L., and Wang, J. (2018). Pinyin and English invented spelling in Chinese speaking students who speak English as a second language. *J. Psycholinguist. Res.* 47, 1163–1187. doi: 10.1007/s10936-018-9585-4

Ding, Y., Liu, R. D., McBride, C., and Zhang, D. (2015). Pinyin invented spelling in mandarin Chinese speaking children with and without reading difficulties. *J. Learn. Disabil.* 48, 635–645. doi: 10.1177/0022219414522704

Dunn, L. M., and Dunn, D. M. (1997). *Peabody Picture Vocabulary Test*. Circle Pines, MN: American Guidance Service. Inc. Publishing.

Durgunoğlu, A. Y., Nagy, W. E., and Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *J. Educ. Psychol.* 85, 453–465. doi: 10.1037/0022-0663.85.3.453

Everson, M. E. (2009). "Literacy development in Chinese as a foreign language," in *Teaching Chinese as a Foreign Language: Theories and Applications*. eds. Michael E. Everson and Yun Xiao (Boston, MA: Cheng & Tsui), 97–111.

Fu, C. L., and Huang, H. S. (2000). The effects of metalinguistic awareness training of the low-achievers at elementary schools. *Chin. J. Psychol.* 42, 87–100.

Gentry, J. R. (1982). An analysis of developmental spelling in GNYS at WRK. *Read. Teach.* 36, 192–200.

Graham, S., and Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Read. Writ.* 27, 1703–1743. doi: 10.1007/s11145-014-9517-0

Guan, C. Q., Liu, Y., Chan, D. H. L., Ye, F., and Perfetti, C. A. (2011). Writing strengthens orthography and alphabetic-coding strengthens phonology in learning to read Chinese. *J. Educ. Psychol.* 103, 509–522. doi: 10.1037/a0023730

Ho, C. S.-H., Chan, D. W.-O., Tsang, S.-M., and Lee, S.-H. (2000). *The Hong Kong Test of Specific Learning Difficulties in Reading and Writing*. Hong Kong.

Invernizzi, M., Abouzeid, M., and Gill, J. T. (1994). Using students' invented spellings as a guide for spelling instruction that emphasizes word study. *Elem. Sch. J.* 95, 155–167. doi: 10.1086/461796

Ju, Z., Zhou, Y., and DelMas, R. (2021). The contributions of separate pinyin skills and oral vocabulary to Chinese word reading of US mandarin immersion third graders. *Read. Writ.* 34, 2439–2459. doi: 10.1007/s11145-021-10150-9

Leong, C. K., Cheng, P. W., and Tan, L. H. (2005). The role of sensitivity to rhymes, phonemes and tones in reading English and Chinese pseudowords. *Read. Writ.* 18, 1–26. doi: 10.1007/s11145-004-3357-2

Li, H., Shu, H., McBride-Chang, C., Liu, H., and Peng, H. (2012). Chinese children's character recognition: Visuo-orthographic, phonological processing and morphological skills. *J. Res. Read.* 35, 287–307. doi: 10.1111/j.1467-9817.2010.01460.x

Lin, D., McBride-Chang, C., Shu, H., Zhang, Y., Li, H., Zhang, J., et al. (2010). Small wins big: analytic pinyin skills promote Chinese word reading. *Psychol. Sci.* 21, 1117–1122. doi: 10.1177/0956797610375447

Liu, P. D., McBride-Chang, C., Wong, A. M. Y., Tardif, T., Stokes, S. F., Fletcher, P., et al. (2010). Early oral language markers of poor reading performance in Hong Kong Chinese children. *J. Learn. Disabil.* 43, 322–331. doi: 10.1177/0022219410369084

Lü, C. (2017). The roles of pinyin skill in English-Chinese Biliteracy learning: evidence from Chinese immersion learners. *Foreign Lang. Ann.* 50, 306–322. doi: 10.1111/flan.12269

Mann, V. A., Tobin, P., and Wilson, R. (1987). Measuring phoneme awareness through invented spellings of kindergarten children. *Merrill-Palmer Q.* 33, 365–391.

Martins, M. A., and Silva, C. (2006). The impact of invented spelling on phonemic awareness. *Learn. Instr.* 16, 41–56. doi: 10.1016/j.learninstruc.2005.12.005

McBride-Chang, C. (1998). The development of invented spelling. *Early Educ. Dev.* 9, 147–160. doi: 10.1207/s15566935eed0902_3

McBride-Chang, C., Bialystok, E., Chong, K. K., and Li, Y. (2004). Levels of phonological awareness in three cultures. *J. Exp. Child Psychol.* 89, 93–111. doi: 10.1016/j.jecp.2004.05.001

McBride-Chang, C., Cho, J. R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., et al. (2005). Changing models across cultures: associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *J. Exp. Child Psychol.* 92, 140–160. doi: 10.1016/j.jecp.2005.03.009

McBride-Chang, C., Shu, H., Zhou, A., Wat, C. P., and Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition. *J. Educ. Psychol.* 95, 743–751. doi: 10.1037/0022-0663.95.4.743

Ouellette, G., Martin-Chang, S., and Rossi, M. (2017). Learning from our mistakes: improvements in spelling lead to gains in reading speed. *Sci. Stud. Read.* 21, 350–357. doi: 10.1080/10888438.2017.1306064

Ouellette, G., and Sénéchal, M. (2008). Pathways to literacy: a study of invented spelling and its role in learning to read. *Child Dev.* 79, 899–913. doi: 10.1111/j.1467-8624.2008.01166.x

Ouellette, G., and Sénéchal, M. (2017). Invented spelling in kindergarten as a predictor of reading and spelling in grade 1: a new pathway to literacy, or just the same road, less known? *Dev. Psychol.* 53, 77–88. doi: 10.1037/dev0000179

Packard, J. L. (1990). Effects of time lag in the introduction of characters into the Chinese language curriculum. *Mod. Lang. J.* 74, 167–175. doi: 10.1111/j.1540-4781.1990.tb02562.x

Poole, F., and Sung, K. (2015). Three approaches to beginning Chinese instruction and their effects on oral development and character recognition. *Eurasian J. Appl. Linguist.* 1, 59–75. doi: 10.32601/ejal.460599

Raven, J., Raven, J. C., and Court, J. H. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales (section 1, general overview)*. Oxford: Oxford Psychologists Press.

Richgels, D. J. (1995). Invented spelling ability and printed word learning in kindergarten. *Reading Research Quarterly* 30, 96–109.

Shu, H., Chen, X., Anderson, R. C., Wu, N., and Xuan, Y. (2003). Properties of school Chinese: implications for learning to read. *Child Dev.* 74, 27–47. doi: 10.1111/1467-8624.00519

Shu, H., Peng, H., and McBride-Chang, C. (2008). Phonological awareness in young Chinese children. *Dev. Sci.* 11, 171–181. doi: 10.1111/j.1467-7687.2007.00654.x

So, D., and Siegel, L. S. (1997). Learning to read Chinese: semantic, syntactic, phonological and working memory skills in normally achieving and poor readers. *Read. Writ.* 9, 1–21. doi: 10.1023/A:1007963513853

Tangel, D. M., and Blachman, B. A. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. *J. Read. Behav.* 24, 233–261. doi: 10.1080/10862969209547774

Thorndike, E. L. (1913). *Educational Psychology*. New York, NY: Teachers College, Columbia University.

Tong, X., and McBride-Chang, C. (2010). Chinese-English biscriptal reading: cognitive component skills across orthographies. *Read. Writ.* 23, 293–310. doi: 10.1007/s11145-009-9211-9

Tong, X., Tong, X., and McBride-Chang, C. (2015). A tale of two writing systems: double dissociation and metalinguistic transfer between Chinese and English word reading among Hong Kong children. *J. Learn. Disabil.* 48, 130–145. doi: 10.1177/0022219413492854

Treiman, R. (2018). Teaching and learning spelling. *Child Dev. Perspect.* 12, 235–239. doi: 10.1111/cdep.12292

Wang, M., Perfetti, C. A., and Liu, Y. (2005). Chinese-English biliteracy acquisition: cross-language and writing system transfer. *Cognition* 97, 67–88. doi: 10.1016/j.cognition.2004.10.001

Yang, M., Cooc, N., and Sheng, L. (2017). An investigation of cross-linguistic transfer between Chinese and English: a meta-analysis. *Asian-Pac. J. Second Foreign Lang. Educ.* 2, 1–21. doi: 10.1186/s40862-017-0036-9

Yin, L., Li, W., Chen, X., Anderson, R. C., Zhang, J., Shu, H., et al. (2011). The role of tone awareness and pinyin knowledge in Chinese reading. *Writ. Syst. Res.* 3, 59–68. doi: 10.1093/wsr/wsr010

Zhang, D., Lin, C. H., Zhang, Y., and Choi, Y. (2017). Pinyin or no pinyin: does access to word pronunciation matter in the assessment of Chinese learners' vocabulary knowledge? *Lang. Learn. J.* 47, 344–353. doi: 10.1080/09571736.2017.1289237

Zhang, H., and Roberts, L. (2021). A comparison of pinyin invented spelling and oddity test in measuring phonological awareness in L2 learners of Chinese. *J. Psycholinguist. Res.* 50, 375–396. doi: 10.1007/s10936-020-09700-z

Zhou, Y. G. (2013). *The Historical Evolution of Chinese Languages and Scripts*. Wuhan: Hubei Education Press.

Zhou, Y., and McBride, C. (2018). The same or different: an investigation of cognitive and metalinguistic correlates of Chinese word reading for native and non-native Chinese speaking children. *Biling. Lang. Congn.* 21, 765–781. doi: 10.1017/S1366728915000279

Zhou, Y., McBride, C., Leung, J. S. M., Wang, Y., Joshi, M., and Farver, J. (2018). Chinese and English reading-related skills in L1 and L2 Chinese-speaking children in Hong Kong. *Lang. Cognit. Neurosci.* 33, 300–312. doi: 10.1080/23273798.2017.1342848

Appendix I: Mandarin pinyin invented spelling list

1. bú yào
2. táng dǎi
3. bǐng gān
4. méi yǒu
5. sān gúo
6. zhī chí
7. qiǎo miào
8. lǜ sè
9. rì chū
10. jiàng xuě

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 February 2023
DOI 10.3389/fpsyg.2023.1133003

Check for updates

OPEN ACCESS

EDITED BY
Yang Zhang,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Tianxu Chen,
Minzu University of China, China
Jinfeng Ding,
Key Laboratory of Behavioral Science,
Institute of Psychology (CAS), China

*CORRESPONDENCE
Yang Zhao
✉ zhaoyang@pku.edu.cn

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 28 December 2022
ACCEPTED 02 February 2023
PUBLISHED 20 February 2023

CITATION
Wang Y, Hu J, An Z, Li C and Zhao Y (2023) The influence of metacognition monitoring on L2 Chinese audiovisual reading comprehension. *Front. Psychol.* 14:1133003.
doi: 10.3389/fpsyg.2023.1133003

The influence of metacognition monitoring on L2 Chinese audiovisual reading comprehension

Yamin Wang, Jingying Hu, Zhuoma An, Chaoran Li and Yang Zhao*

School of Chinese as a Second Language, Peking University, Beijing, China

Metacognition monitoring is the ability to evaluate the cognitive process actively. L2 learners with high metacognition monitoring ability can better monitor reading processes and outcomes consciously, thus facilitating self-regulated learning and improving reading efficiency. Previous studies mostly used offline self-reports to examine the metacognition monitoring in static text reading by L2 learners. This study investigated the effects of different indicators of metacognition monitoring on L2 Chinese audiovisual comprehension by online confidence judgment and audiovisual comprehension tasks. Target measures of metacognition monitoring included absolute calibration accuracy based on video or test and relative calibration accuracy measured by Gamma or Spearman correlation coefficient. 38 intermediate-advanced Chinese learners participated in the study. Multiple regression analysis showed three main results. First, absolute calibration accuracy can significantly predict L2 Chinese audiovisual comprehension, while relative calibration accuracy has no significant effect. Second, the predictive effect of video-based absolute calibration accuracy is affected by the video difficulty, that is, the greater the video difficulty, the greater the impact on the performance of audiovisual comprehension. Third, the predictive effect of test-based absolute calibration accuracy is influenced by the language proficiency, specifically, the higher the L2 Chinese proficiency, the stronger the prediction on the performance of audiovisual comprehension. These results support a multidimensional view of metacognition monitoring by specifying how different indicators of metacognition monitoring may predict L2 Chinese audiovisual comprehension. The findings have important pedagogical implications for strategy training of metacognition monitoring and point to the necessity to take task difficulty and individual differences among learners into full consideration.

KEYWORDS

metacognition monitoring, measurement indicators, audiovisual reading comprehension, video difficulty, language proficiency

1. Introduction

Metacognition monitoring refers to an individual's ability to actively assess cognitive processes, where people use appropriate and effective strategies to regulate their cognitive processes (Lin and Zabrucky, 1998). In the context of reading comprehension, metacognition monitoring is specifically concerned with the identification of difficulties and the use of strategies in the reading process, which focuses on how learners monitor their reading processes and outcomes consciously (Chen, 2009). Numerous studies have shown that metacognition monitoring has significant predictive power on the reading comprehension performance of native and L2 speakers (Garner, 1980; Maki et al., 1994; Van Gelderen et al., 2004; Taki, 2016; Silawi et al., 2020). Epstein et al. (1984) explained that learners' self-assessment of text comprehension often does not match the actual outcomes, and cognitive illusion can affect or even hinder the whole information processing and integration to some extent. This self-assessment ability is particularly crucial in the comprehension of audiovisual multimodal texts, which involves the selection and processing of multiple information such as pictures, sounds, and texts. In addition, learners also need to integrate and evaluate the complex and dynamic information in real time to understand and interpret text meaning (Coiro, 2011; Afflerbach et al., 2015; Fox and Alexander, 2017), which poses new challenges for metacognition monitoring in multimodal text reading for second language learners.

Previous studies mainly focused on the role of metacognition monitoring on L2 reading comprehension of static texts. Van Gelderen et al. (2004) explored the metacognition monitoring strategies during L2 English reading by questionnaire surveys and showed that metacognitive knowledge significantly predicted learners' reading comprehension performance. Taki (2016) adopted the same measure and found that metacognitive knowledge also had a significant predictive effect on Dutch L2 learners' reading test scores. Accordingly, the above studies examined the use of explicit strategic knowledge of L2 learners mainly by offline self-reports, but Ackerman and Goldsmith (2008) argued that the basis of learning process moderation lies in continuous real-time assessment and monitoring. Obviously, it is necessary to further explore the role of metacognition monitoring on L2 reading comprehension through online test methods. In addition, it has been shown that the effect of metacognition monitoring on reading comprehension is also influenced by text difficulty and language proficiency. Kim (2014) found that text difficulty affected the metacognition monitoring strategies in English L2 online reading comprehension, and learners tended to adopt more metacognitive strategies for regulation when reading more complex texts, which had a greater impact on their reading comprehension. Furthermore, low-level L2 learners were not good at using metacognitive strategies to monitor the reading comprehension process, while high-level learners could self-assess more accurately (Han and Stevenson, 2008; Tsai et al., 2010; Míguez-Álvarez et al., 2021; Gu and Wang, 2022). It is worth mentioning that Maki et al. (2005) combined two factors of text difficulty and language proficiency. They found that English native speakers with lower language proficiency over-assessed their reading performance when reading more difficult texts, while native speakers with higher language proficiency tended to underestimate their performance. It remains to be addressed whether the predictive effect of metacognition monitoring on L2 audiovisual reading comprehension is influenced by text difficulty and language proficiency.

Related studies have also investigated the effectiveness of metacognitive abilities in Chinese L2 static text reading, mainly using think-aloud protocol (Liu, 2002), questionnaires (Wu, 2016), or interviews (Zhu and Kong, 2017). The results showed that metacognitive ability could effectively predict Chinese L2 reading comprehension performance. However, there are still some unsolved problems: (1) previous studies mainly adopted offline self-reports, and online testing methods can be conducted to explore the whole process of metacognitive monitoring; (2) metacognitive monitoring in Chinese L2 audiovisual comprehension has not yet been investigated, which may bring greater challenges to L2 learners; (3) some important factors (such as video difficulty and language proficiency) have not been examined simultaneously in previous studies. The present study aims to examine the role of metacognition monitoring on Chinese L2 learners' audiovisual reading comprehension through online confidence judgment and audiovisual comprehension tasks and further explore the moderating effects of video difficulty and language proficiency.

2. Measuring metacognition monitoring

How to measure metacognitive monitoring effectively has become an essential issue in academic circles. Thiede et al. (2011) also argued that the influence of metacognition monitoring on learning outcomes cannot be emphasized enough, but it is how to be measured that is critical. Throughout previous research, the measures of metacognition monitoring in reading comprehension have consisted of two main categories: first, having learners make direct self-reports stating their use of strategic knowledge during reading (Block, 1992; Kroll and Ford, 1992; Zabrucky and Commander, 1993); and second, requiring learners to make real-time self-assessment, thus comparing the difference between learners' self-ratings and their actual performance (Maki et al., 1994; Lin and Zabrucky, 1998; Lin et al., 2001; Sarac and Tarhan, 2009; Kasperski and Katzir, 2013). The former mainly uses offline self-reports to retrace the use of strategic knowledge during reading, while the latter focuses on the self-assessment ability during reading. Pieschl (2009) defines *calibration* as the accuracy of the learners' perception of his or her own performance, that is, the ability to assess comprehension accurately. The researchers asked the participants to make a confidence judgment during the reading process to self-assess their reading situation, and calculate the *calibration bias*, which is the difference between the self-assessed correct rates and the actual ones (Glenberg and Epstein, 1987; Pressley and Ghatala, 1988; Zabrucky et al., 2009; Lin and Yu, 2015).

There are many different accounts of the actual measurement of metacognition monitoring in reading comprehension. It has been discussed extensively whether metacognition monitoring can effectively predict reading comprehension performance. From the perspective of the monitoring object, metacognition monitoring is divided into two types: text-based and test-based dimension. Specifically, text-based metacognition monitoring refers to the learners' self-assessment of text comprehension after reading the text and before taking the test, i.e., pre-test confidence judgment; in contrast, test-based metacognition monitoring refers to the learners' self-assessment of test performance after reading the text and completing the test, i.e., post-test confidence judgment (Lin and Zabrucky, 1998; Lin et al., 2001; Pieschl, 2009). Maki et al. (1994) showed that reading comprehension was significantly correlated with test-based metacognition monitoring, but not with text-based ones. Post-test confidence judgments are indeed more accurate than pre-test ones, because the post-test judgments could provide learners with more information to understand the test itself and thus better predict their actual performance (Pieschl, 2009). However, Zabrucky et al. (2009) found that both text-based and test-based metacognition monitoring were significantly associated with learners' reading comprehension performance. In summary, the predictive effect of text-based metacognition monitoring on reading comprehension is widely divergent and needs further verification.

In addition, metacognition monitoring can be divided into two types according to the differences in statistical methods: (1) Absolute Calibration Accuracy (also known as Calibration Bias) refers to the absolute difference between the learners' self-assessment and actual performance; the greater the absolute value of the subtraction between the two, the greater the gap between the learners' self-assessment and actual performance; (2) Relative Calibration Accuracy (also known as Resolution) refers to the extent to which learners' self-assessment scores reflect their actual reading comprehension scores, by calculating the correlation between the two correct rates. Higher correlation coefficients indicate greater agreement between learners' self-assessment and actual performance (Dunlosky and Thiede, 2013). These two calculations of metacognition monitoring are statistically independent of each other, and a high absolute calibration accuracy of learners does not mean a high relative calibration accuracy and vice versa (Koriat et al., 2002; Maki et al., 2005; Griffin et al., 2009). The multidimensional view of metacognitive monitoring argues that different indicators of metacognitive monitoring have different predictive effects on reading comprehension, both in terms of monitoring objects and statistical methods, and opposes the one-dimensional view that confuses all indicators (Moore et al., 2005; Chen and Li, 2008; Hadwin and Webster, 2013). Hadwin and Webster (2013) suggested that relevant researchers should use both measures of metacognition monitoring in their actual studies to explore the relationship between the different indicators of metacognition monitoring and reading comprehension.

The present study focuses on the metacognition monitoring of L2 Chinese audiovisual reading comprehension and examines the following four measures: (1) video-based absolute accuracy, the absolute difference between learners' self-assessment of video comprehension and their actual performance after watching the video and before taking the test; and (2) test-based absolute accuracy, the absolute difference between learners' self-assessment of their performance on the test items and their actual performance after watching the video and completing the test; and (3) test-based relative accuracy measured by Gamma, the correlation coefficient between learners' self-assessment of their performance on the test items and their actual performance after watching the video and completing the test; and (4) test-based relative accuracy measured by Spearman, the correlation coefficient between learners' self-assessment of their performance on the test items and their actual performance after watching the video and completing the test. Considering the four indicators of metacognition monitoring, this study will further investigate the following two questions:

1. Do the four indicators of metacognition monitoring have a significant predictive effect on the performance of L2 Chinese audiovisual reading comprehension?
2. Is the predictive effect mediated by video difficulty and Chinese language proficiency?

3. Materials and methods

3.1. Participants

The participants were 38 undergraduate students studying Chinese at a university in Beijing (18 male, 20 female; mean age = 21.53, $SD = 2.05$). All the participants were native Korean speakers and had studied Chinese for a mean of 6.81 years ($SD = 3.65$). They were intermediate and advanced Chinese L2 learners, with an average score of 25.47 ($SD = 3.07$, ranging from 19 to 30) in the fixed-ratio cloze test (full score: 30, Feng et al., 2020). The participants had no medical history of learning disabilities, attention deficit, hearing or visual impairment. They were recruited through experimental advertisements, gave informed consent to participate in the experiment, and were paid after the experiment.

3.2. Measures

3.2.1. Language proficiency test

In order to ensure the validity and immediacy of the test results, we did not use the participants' acquired HSK scores as the basis for measuring their language proficiency, whereas we remeasured the participants' Chinese language proficiency from both subjective and objective perspectives: (1) Language proficiency self-assessment questionnaire. For the subjective measures, we used the Chinese translation of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007). In addition to the Chinese language learning experience survey, the participants rated their Chinese language proficiency (on a 10-point scale) in four areas: listening, speaking, reading, and writing. We used the total score of the four items as the language proficiency self-ratings. (2) Fixed-ratio cloze questions. For the objective measures, we adopted the Chinese proficiency test developed by Feng et al. (2020), which required participants to complete a test of 30 fill-in-the-blanks within 15 min. If participants could not complete or fill in the blanks incorrectly, they were counted as errors (0 points), and the test had a total score of 30 points. The final score of the test is the number of correct answers (*Cronbach's alpha* = 0.71). The Pearson correlation coefficient between the subjective and objective measures of Chinese language proficiency in this experiment was 0.602 ($p < 0.001$), indicating a strong correlation. This study followed Silawi et al. (2020) in selecting a more accurate objective measure for data analysis.

3.2.2. Metacognition monitoring tasks

Participants were required to make confidence judgments before and after the test to further calculate the metacognition monitoring of the video and the test dimension. (1) Prediction confidence judgments. After watching the video and before taking the test, participants were asked, "How accurate do you think you will be when responding to comprehension questions regarding the video you have just watched?" They made judgments using a five-point scale ranging from 0 to 100% (Ackerman and Goldsmith, 2011; Silawi et al., 2020). (2) Postdiction confidence judgments. After completing each question of the video comprehension test, participants were asked, "How confident are you that you selected the correct answer?" These questions were also answered using a five-point scale from 0 to 100%.

3.2.3. Video comprehension test

Participants watched three videos in Chinese on a computer screen, each followed by five multiple-choice questions, which consisted of two detail questions, two inference questions, and one topic question (Cronbach's alpha = 0.66). The videos and test questions were adapted from *China Focus*, a set of comprehensive language textbooks designed specifically for L2 Chinese audiovisual and speaking courses. Three videos were selected for this experiment as practice video (1 min), test video 1 (3 min), and test video 2 (3 min). To control the difficulty of the videos, we revised the vocabulary and sentence patterns of the original videos and analyzed the difficulty of the revised texts with the assistance of the *Chi-Editor*, which is designed to provide CSL teachers with difficulty grading and intelligent adaptation of reading texts and contains three core modules: text rating, word annotation, and word archive (Jin et al., 2018). The results showed that the text difficulty of the practice video is the medium level (2.42), and the text length is 130 words; the text difficulty of test video 1 and video 2 are the medium level (2.01) and the high level (3.04) respectively, and have the same text length of 607 words. On this basis, we hired a professional dubbing artist (National Mandarin Proficiency Test Level 1B) to re-dub the videos and added subtitles (Song font 4, white on black background) to the three videos using the professional video editing software *Premiere*. At the end of all tests, participants were asked to answer the following question: "Which of the two previous videos did you find more difficult?" The options were "Test video 1", "Test video 2", and "Same difficulty". The results showed that 89.47% (34/38) of the participants thought that test video 2 was more difficult.

The three videos used in this experiment are all character documentaries. The practice video is selected from *China Focus: Inspirational Section* (Wang, 2016a), initially titled *The Podium of Life*, which is about disabled table tennis coaches. Test video 1 is selected from *China Focus: Professional Section* (Wang, 2016b), initially titled *Beijing Cabbie*, which is about Beijing taxi drivers. Test video 2 is from *China Focus: Arts Section* (Wang, 2017), originally titled *Make the Future Better*, which is about the founder of *Meitu Xiu Xiu* software. At the end of the experiment, it was revealed through the interviews that the participants had never watched the above three videos before. They were familiar with the themes of the three videos to a similar extent.

3.3. Procedures

Due to the epidemic, the participants could not come to the laboratory for the experiment, so all experiments in this study were completed online. Before the formal experiment, each participant filled out a questionnaire with basic information, including age, gender, education, duration of Chinese language learning, HSK level and score, etc. They then completed two online tests of language proficiency self-assessment questionnaire and fixed-ratio cloze questions. To ensure participants' concentration and freedom from outside interference during the experiment, they were required to conduct the experiment in a quiet space. The formal experiment was conducted on the Tencent conference platform, and their computer screens had to be shared with the experiment implementer in full screen throughout the experiment. The formal experiment consisted of two parts, that is, the practice session and the test session. Participants first watched the video online and comprehended its content. Before taking the test, they were given the prediction confidence judgment on how well they understood the video. Afterwards, they were asked to carefully complete five multiple-choice questions, each of which was presented on a single screen. Immediately after making their choice, they would assess how likely they were to get the question right and then move on to the next test question until the test was completed. With the exception of the practice session, neither the videos nor the test questions were allowed to be re-watched or redone. After viewing each video and completing the corresponding test questions, participants were given a 15 sec break. The experimental results were mainly recorded for the correctness rate of test questions and the participants' confidence judgment scores for each video and each test question. The specific experimental procedures and materials are available at the link: https://www.wenjuan.ltd/s/3umUZzA/.

3.4. Data analysis

Multiple regression analysis was conducted to examine the influence of metacognition monitoring on L2 Chinese audiovisual reading comprehension using the *lm* package in the R runtime environment (R version 4.0.3; R Core Team, 2020). The dependent variable of the model was the audiovisual comprehension performance, i.e., the actual percentage of correct responses to the multiple-choice questions in the formal experimental video, and the following fixed effects were included in the model: (1) Chinese language proficiency: fixed-ratio cloze test scores; (2) video difficulty: test video 1 is an easy video and test video 2 is a more difficult video; (3) video-based absolute accuracy: the absolute value of the self-assessed correctness rate minus the actual correctness rate for each video; (4) test-based absolute accuracy: the absolute value of the self-assessed correctness rate minus the actual correctness rate for each test question; (5) test-based relative accuracy measured by Gamma: the Gamma correlation coefficient between the self-assessed and actual correctness rate of five test questions in each video; (6) test-based relative accuracy measured by Spearman: the Spearman correlation coefficient between the self-assessed and actual correctness rate of five test questions in each video. Video difficulty was a categorical variable, and the rest were continuous variables. The optimal multiple linear regression model can be fitted better if there is a certain correlation between the independent variables and the dependent variable (correlation coefficient is greater than 0.3) and the correlation between each independent variable is not too high (correlation coefficient is less than 0.7) (Pallant, 2007). Therefore, we conducted the bivariate correlation analysis and eliminated the independent variables that were not strongly correlated with the dependent variable or had multicollinearity each other before establishing the regression model.

4. Results

4.1. Preliminary analyses

Descriptive statistics and correlation analyzes are shown in Table 1. Among the four indicators measuring metacognition monitoring, there was a significant positive correlation between video-based and test-based absolute accuracy, and the two indicators of relative accuracy showed a highly positive and significant correlation. The correlation between the two relative indicators was more significant than that

TABLE 1 Descriptive statistics and correlation matrix.

	M	*SD*	1	2	3	4	5	6
1. Audiovisual comprehension performance	73.16	26.64	1.00					
2. Chinese language proficiency	25.47	3.07	0.22	1.00				
3. Video-based absolute accuracy	21.84	17.41	−0.35**	−0.08	1.00			
4. Test-based absolute accuracy	20.00	17.24	−0.10	−0.02	0.37**	1.00		
5. Test-based relative accuracy measured by Gamma	0.41	0.75	−0.10	−0.16	−0.01	−0.26	1.00	
6. Test-based relative accuracy measured by Spearman	0.28	0.40	−0.01	−0.09	−0.07	−0.26	0.93***	1.00

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

TABLE 2 Multiple regression of audiovisual comprehension performance based on video-based absolute accuracy.

Variables	*β*	*SE*	*t*	*p*
Intercept	84.67**	25.72	3.29	0.002
Video-based absolute accuracy	−0.52	0.27	−1.94	0.056
Video difficulty	−97.85*	37.37	−2.62	0.011
Chinese language proficiency	−0.17	1.00	−0.17	0.863
Video-based absolute accuracy: video difficulty	−0.97**	0.31	−3.17	0.002
Video difficulty: Chinese language proficiency	3.42*	1.43	2.40	0.019

$R^2 = 0.5367$, $F(5, 70) = 16.21$, *$p < 0.05$, **$p < 0.01$.

between the two absolute indicators. The correlations between all variables were less than 0.7. In terms of the correlation between the independent and dependent variables, there was a significant negative correlation between video-based absolute accuracy and audiovisual comprehension performance, and a borderline significant correlation between Chinese language proficiency and audiovisual comprehension performance. Univariate linear regressions were conducted on the dependent variable of audiovisual comprehension performance using each of the four indicators as the independent variable. The results showed that both two indicators of relative accuracy were not significant. Based on the data from this study, neither of the two relative indicators significantly predicted the audiovisual comprehension performance, so the model data for the absolute accuracy of metacognition monitoring were mainly reported below.

4.2. Multiple regression analyses

The results of the multiple regression model with video-based absolute accuracy as the independent variable are shown in Table 2. The main effect of video-based absolute accuracy was marginally significant ($\beta = -0.52$, $SE = 0.27$, $t = -1.94$, and $p = 0.056$), indicating that larger absolute accuracy seemed to be associated with poorer audiovisual comprehension performance. The effect of video difficulty was significant ($\beta = -97.85$, $SE = 37.37$, $t = -2.62$, and $p = 0.011$), with the more difficult the video, the worse the comprehension performance. Importantly, we found a significant interaction effect between video-based absolute accuracy and video difficulty ($\beta = -0.97$, $SE = 0.31$, $t = -3.17$, and $p = 0.002$), with the follow-up contrasts demonstrating the greater the video difficulty, the greater the effect of video-based absolute accuracy on audiovisual comprehension performance. The effect of Chinese language proficiency was not significant ($\beta = -0.17$, $SE = 1.00$, $t = -0.17$, and $p = 0.863$), but the interaction effect of video difficulty and Chinese language proficiency was significant ($\beta = 3.42$, $SE = 1.43$, $t = 2.40$, and $p = 0.019$); specifically, the effect of Chinese language proficiency on audiovisual comprehension performance was enhanced with the increase of video difficulty.

A similar set of multiple regression analyzes were also conducted with test-based absolute accuracy as the independent variable (see Table 3). Test-based absolute accuracy contributed significantly to audiovisual comprehension performance ($\beta = -2.98$, $SE = 1.17$, $t = -2.55$, and $p = 0.013$), with the higher the deviation value of the self-assessment, the worse the audiovisual comprehension performance. The main effect of video difficulty was also significant ($\beta = -34.20$, $SE = 7.62$, $t = -4.49$, and $p < 0.001$), which indicated the more difficult the video, the worse the audiovisual comprehension performance. Chinese language proficiency was not a significant predictor in the model ($\beta = -0.87$, $SE = 1.18$, $t = -0.73$, and $p = 0.467$), while the interaction effect of test-based absolute accuracy and Chinese language proficiency was significant ($\beta = 0.14$, $SE = 0.04$, $t = 3.10$, and $p = 0.003$). Further comparisons revealed that the effect of test-based absolute accuracy on audiovisual comprehension performance increased with the improvement of Chinese language proficiency. The model yielded no significant interaction between test-based absolute accuracy and video difficulty ($\beta = -0.29$, $SE = 0.38$, $t = -0.77$, and $p = 0.442$).

TABLE 3 Multiple regression of audiovisual comprehension performance based on test-based absolute accuracy.

Variables	β	SE	t	p
Intercept	106.43**	31.28	3.40	0.001
Test-based absolute accuracy	−2.98*	1.17	−2.55	0.013
Video difficulty	−34.20***	7.62	−4.49	<0.001
Chinese language proficiency	−0.87	1.18	−0.73	0.467
Test-based absolute accuracy: video difficulty	−0.29	0.38	−0.77	0.442
Test-based absolute accuracy: Chinese language proficiency	0.14**	0.04	3.10	0.003

$R^2 = 0.5264$, $F(5, 70) = 15.56$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

5. Discussion

5.1. The predictive effect of metacognition monitoring on L2 Chinese audiovisual reading comprehension

The Self-regulation Learning Theory suggests that learners need to monitor their learning in real time to see if they reach their predetermined goals (Zimmerman and Schunk, 2001), emphasizing the role of metacognition monitoring in individual learning. Our findings showed that metacognition monitoring effectively predicts L2 Chinese audiovisual reading comprehension, which are consistent with those of Chinese L2 static text reading studies (Liu, 2002; Zhu and Kong, 2017). In this study, intermediate and advanced Chinese L2 learners could make appropriate self-assessment of audiovisual reading comprehension. Specifically, the test-based absolute accuracy is a better predictor of L2 Chinese audiovisual comprehension performance than the video-based absolute accuracy, consistent with studies of native speakers (Lin et al., 2001; Pieschl, 2009; Zabrucky et al., 2009). Both two absolute indicators derived by confidence judgment task represent the absolute difference between self-assessment and actual performance. Still, their connotations reflect different psychological features of metacognition monitoring, where the temporal dimension of making confidence judgments appears to be particularly important. In the first confidence judgment, participants need to estimate the expected comprehension performance before being exposed to the test. In contrast, in the second confidence judgment, they can integrate both text and test information to assess their actual performance and obviously can obtain more additional clues to help them make more accurate judgments. In addition, Sarac and Tarhan (2009) showed that English L2 learners made more accurate self-assessments based on their test performance but were unable to accurately assess their own comprehension of the text, which the authors attributed to a lack of discourse knowledge. Although the participants in this study were all intermediate to advanced learners who may have been familiar with most of the Chinese words and grammar in the videos, they had difficulty in following up on the video story in a specific organizational development way. Moreover, new information and new ideas emerged in real time. Therefore, it was difficult to make accurate judgments based solely on the videos.

The present study did not find significant predictive effects of the two relative indicators of metacognition monitoring on L2 audiovisual reading comprehension, which may indicate that the practical predictability of the relative indicators is less stable (Ackerman and Goldsmith, 2011). Masson and Rotello (2009) also found that Gamma coefficients computed from simulated sample data would reflect variation due to item bias rather than actual accuracy. These artifactual effects tend to be exaggerated in unequal variance distributions. In the case of this study, it may also be due to two reasons: on the one hand, the Gamma and Spearman coefficient are both calculated based on fixed-order variables, and the existence of missing values in the actual calculation process resulted in the relative indicators not being able to reflect the correlation between self-assessment and actual performance (Wiley et al., 2016); on the other hand, the relatively small number of test questions, with only five questions per video, may affect the stability and sensitivity of these coefficients (Spellman et al., 2008). Some researchers have also argued that we must be cautious in using relative indicators for metacognition monitoring in practical applications because their practical predictability may be affected by factors such as sample distribution, missing calculations, and the number of items (Masson and Rotello, 2009; Ackerman and Goldsmith, 2011).

It is worth mentioning that the four indicators of metacognition monitoring were examined comprehensively in this study, and there were differences in the predictive effects on audiovisual comprehension performance. Explicitly speaking, test-based absolute accuracy could significantly be predicted audiovisual comprehension performance and video-based absolute accuracy had weaker predictive power, while the two relative indicators did not have significant predictive effects. The above results provide support for the multidimensional view of metacognition monitoring (Moore et al., 2005; Chen and Li, 2008; Hadwin and Webster, 2013), demonstrating that the validity and stability of metacognition monitoring indicators vary, whether they are video-based or test-based, absolute or relative. Their practical predictability still need to be further examined in the context of specific experimental scenarios and instruments. Therefore, a one-dimensional view that simply handles different indicators without being distinguished should be opposed.

5.2. Real-time measures of metacognition monitoring and L2 audiovisual reading comprehension

Previous studies have focused on the role of metacognition monitoring in static text comprehension. In contrast, this study examined metacognition monitoring in L2 audiovisual reading comprehension using real-time measures, demonstrating the effectiveness of metacognition monitoring for intermediate and advanced Chinese L2 learners. Ackerman and Goldsmith (2008) argue that the basis of regulating the learning process lies in continuous real-time assessment. Online audiovisual reading comprehension shown in our experiments is a continuous, real-time dynamic learning process. Metacognition monitoring affects the whole process of cognitive activity, and the changed cognitive processing state will in turn influence metacognition monitoring. Our results showed that two of all four metacognition monitoring indicators measured by the real-time confidence judgments were good predictors of Chinese L2 audiovisual comprehension performance, demonstrating the reliability and validity of the online measure. It also proves that intermediate and advanced Chinese L2 learners can make accurate self-assessments especially continuous based on test performance during the selection and processing of multimodal information such as pictures, sounds, and words.

In real-life contexts, self-regulated learning by native or L2 speakers is online in real time, which requires them to monitor and regulate their learning process instantaneously. Metacognition monitoring reflected in this process not only motivates learners to adopt effective strategies but also enhances their self-efficacy, thus improving the effectiveness of online learning (Chen and Wen, 2010). Ackerman and Goldsmith (2011) also found that the difference between screen reading and paper reading is usually not cognitive but mainly reflected in metacognition, that is, learners were more inaccurate in predicting online reading performance and consumed more time in regulating learning time. The present study highlights the importance of metacognition monitoring for online audiovisual reading comprehension.

5.3. Moderating effects of video difficulty and L2 proficiency

The Cognitive Effort Hypothesis holds that learners' self-assessment accuracy of text comprehension is affected by text difficulty. To be specific, self-assessment of text comprehension is less accurate if the text being read is easy for the learner since they do not pay excessive attention to the details of easy texts (Maki et al., 1990). The present study found that the predictive effect of video-based absolute accuracy on Chinese L2 audiovisual comprehension was affected by video difficulty, in line with the Cognitive Effort Hypothesis. With the increase of video difficulty, the accuracy of Chinese learners' self-assessment of video comprehension is better, and the role of metacognition monitoring in predicting audiovisual comprehension is greater. Kim, (2014) also found that English L2 learners tended to use more metacognitive reading strategies when reading difficult texts, which had a greater impact on their reading comprehension. In addition, the participants selected for the experiment are intermediate to advanced Chinese L2 learners. When learners read the easy video (video 1), they were not inclined to consume too many cognitive resources in the control of video details or adopt more metacognitive reading strategies, but could get good audiovisual comprehension scores in the end. Further analysis showed that 20 participants (38 in total) achieved an actual correct rate of 100% on the video 1 test, with an overall average correct rate of 88.95%. To some extent, there was a ceiling effect, i.e., intermediate and advanced learners generally understood the lower difficulty videos better, so the effect of metacognition monitoring on audiovisual comprehension may be weakened.

The present study also found that the predictive effect of test-based absolute accuracy on L2 Chinese audiovisual comprehension was also moderated by Chinese language proficiency, i.e., the stronger the predictive effect of metacognition monitoring on audiovisual comprehension as learners' Chinese language proficiency increased. Tsai et al. (2010) found that language proficiency had a greater impact on English L2 speakers' use of reading metacognitive strategies compared to native speakers, and that highly proficient L2 speakers were better able to use more metacognitive strategies to monitor their reading comprehension process, which was conducive to improving their text comprehension and thus achieving higher reading scores. Míguez-Álvarez et al. (2021) also found that the monitoring accuracy of Spanish L2 speakers is influenced by their language level, that is, learners with lower L2 proficiency tended to overestimate their reading performance, while higher-level L2 learners are better able to monitor the reading process and make more accurate self-assessments. In addition, the absolute accuracy of metacognition monitoring was significantly higher in high-level learners than in low-level learners (Gu and Wang, 2022), because high-level learners were more likely to pay attention to and identify important information in reading materials and also invested more time and cognitive resources.

The above results suggest that the predictive effect of metacognition monitoring on Chinese L2 audiovisual comprehension is complex. The prediction of video-based absolute accuracy was affected by the video difficulty (the more difficult the video, the more significant the prediction effect). In contrast, the prediction of test-based absolute accuracy was moderated by L2 proficiency (the higher Chinese L2 proficiency, the more significant the prediction effect). Our findings also provide evidential support for a multidimensional view of metacognition monitoring (Moore et al., 2005; Chen and Li, 2008; Hadwin and Webster, 2013). As stated earlier, some differences exist in L2 learners' metacognition monitoring on the two dimensions of video comprehension and test performance, both in terms of self-assessment time and information resource capacity. Thus, their predictive effects are moderated by different variables.

6. Conclusion and implications

The current study systematically examined four indicators of metacognition monitoring in L2 Chinese audiovisual comprehension from the perspective of monitoring objects and statistical methods by online confidence judgment tasks. Furthermore, we explored the predictive effects of different indicators on L2 audiovisual comprehension performance. The main results showed that the absolute accuracy of metacognition monitoring had a significant predictive effect on L2 Chinese audiovisual comprehension, while the effect of relative accuracy was not significant. Specifically, the predictive effect of video-based absolute accuracy was influenced by the video difficulty (the more difficult the video, the more significant the prediction effect). In contrast, test-based absolute accuracy was moderated by L2 proficiency (the higher Chinese L2 proficiency, the more significant the prediction effect). Our findings support a multidimensional view of metacognition monitoring and oppose a one-dimensional view that simply makes no distinction among all four indicators.

Pedagogically, the empirical findings in this study not only emphasize the importance of metacognition monitoring in L2 audiovisual comprehension, but also provide some implications for teaching Chinese as a second language. Firstly, both Chinese teachers and learners should recognize that reading comprehension is not only a complex cognitive process but also a complex metacognitive process, and learners with higher metacognition monitoring ability can effectively regulate their own cognitive processes in real time and thus make more accurate self-assessments. Therefore, instructional interventions aimed at metacognition monitoring are essential. Teachers should provide learners with a variety of strategies needed to conduct more accurate self-assessment, such as planning and prediction of text content, assessment of the characteristics of text genre, assessment of a text content, construction and modification of a mental model, supervision of comprehension and text understanding, revision, and self-correction (Míguez-Álvarez et al., 2021). Secondly, this study also found that the predictive effect of metacognition monitoring on L2 audiovisual comprehension is complex and is affected by the video difficulty and language proficiency, which reveals that we need to appropriately adjust the difficulty of the video task or text according to L2 learners' language proficiency when training metacognition monitoring strategies, so as to motivate learners to mobilize their own initiative, monitor and regulate audiovisual comprehension in real time within a specific range of cognitive effort.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Academic Committee of School of Chinese as a Second Language, Peking University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YZ and YW designed the study and contributed to the final manuscript writing. YW, JH, ZA, and CL performed the study and collected the data. JH and YW analyzed the data. YW, JH, ZA, and CL prepared the draft. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by National Social Sciences Project "Application of Chinese Proficiency Grading Standards and Strategies of Textbook Research and Development (21STA032)".

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ackerman, R., and Goldsmith, M. (2008). "Learning directly from screen? Oh-no, I must print it! Metacognitive analysis of digitally presented text learning" in *Learning in the Technological Era III: Proceedings of the 2009 Chais Conference*. eds. Y. Eshet-Alkalai, A. Caspi and N. Geri (Raanana, Israel: Open University of Israel)

Ackerman, R., and Goldsmith, M. (2011). Metacognitive regulation of text learning: on screen versus on paper. *J. Exp. Psychol. Appl.* 17, 18–32. doi: 10.1037/a0022086

Afflerbach, P., Cho, B.-Y., and Kim, J.-Y. (2015). Conceptualizing and assessing higher-order thinking in Reading. *Theory Pract.* 54, 203–212. doi: 10.1080/00405841.2015.1044367

Block, E. L. (1992). See how they read: comprehension monitoring of L1 and L2 readers. *TESOL Q.* 26:319. doi: 10.2307/3587008

Chen, Q. (2009). Metacomprehension monitoring and regulation in reading comprehension. *Acta Psychol. Sin.* 41, 676–683. doi: 10.3724/SP.J.1041.2009.00676

Chen, Q., and Li, L. (2008). Rating comprehension and predicting performance: clarifying two forms of metacognition monitoring. *Acta Psychol. Sin.* 40, 961–968. doi: 10.3724/SP.J.1041.2008.00961

Chen, Q., and Wen, Z. (2010). Metacognition and effective learning:principles and conditions. *J. East China Norm. Univ. Sci.* 28:53. doi: 10.16382/j.cnki.1000-5560.2010.01.005

Coiro, J. (2011). Predicting Reading comprehension on the internet: contributions of offline Reading skills, online Reading skills, and Prior knowledge. *J. Lit. Res.* 43, 352–392. doi: 10.1177/1086296X11421979

Dunlosky, J., and Thiede, K. W. (2013). Four cornerstones of calibration research: why understanding students' judgments can improve their achievement. *Learn. Instr.* 24, 58–61. doi: 10.1016/j.learninstruc.2012.05.002

Epstein, W., Glenberg, A. M., and Bradley, M. M. (1984). Coactivation and comprehension: contribution of text variables to the illusion of knowing. *Mem. Cogn.* 12, 355–360. doi: 10.3758/BF03198295

Feng, L., Feng, H., Bai, S., and Wu, J. (2020). An analysis of proficiency test for CSL (Chinese as second language) based on fixed-ratio cloze questions. *Appl. Linguist.* 115, 69–79. doi: 10.16499/j.cnki.1003-5397.2020.03.011

Fox, E., and Alexander, P. A. (2017). "Text and comprehension: A retrospective, perspective, and prospective," *Handbook of Research on Reading Comprehension.* ed. S. E. Israel (New York, NY: Guilford Press), 335–352.

Garner, R. (1980). Monitoring of understanding: an investigation of good and poor readers' awareness of induced miscomprehension of text. *J. Read. Behav.* 12, 55–63. doi: 10.1080/10862968009547352

Glenberg, A. M., and Epstein, W. (1987). Inexpert calibration of comprehension. *Mem. Cogn.* 15, 84–93. doi: 10.3758/BF03197714

Griffin, T. D., Jee, B. D., and Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Mem. Cogn.* 37, 1001–1013. doi: 10.3758/MC.37.7.1001

Gu, S., and Wang, X. (2022). Absolute accuracy of EFL learners' meta-comprehension monitoring: based on eye tracking evidence. *Technol. Enhanc. Foreign Lang. Educ.* 204:15–22+114.

Hadwin, A. F., and Webster, E. A. (2013). Calibration in goal setting: examining the nature of judgments of confidence. *Learn. Instr.* 24, 37–47. doi: 10.1016/j.learninstruc.2012.10.001

Han, F., and Stevenson, M. (2008). Comprehension monitoring in first and foreign language reading. *Univ. Syd. Pap. TESOL* 18, 73–110.

Jin, T., Lu, X., Lin, Y., and Li, B. (2018). Chi-editor: an online Chinese text evaluation and adaptation system. Available at: https://www.languagedata.net/editor/ (Accessed July 8, 2021).

Kasperski, R., and Katzir, T. (2013). Are confidence ratings test-or trait-driven? Individual differences among high, average, and low Comprehenders in fourth grade. *Read. Psychol.* 34, 59–84. doi: 10.1080/02702711.2011.580042

Kim, N. H. (2014). Effect of text difficulty on the use of metacognition-driven strategy clusters in Reading online texts in English. *Mod. Stud. Engl. Lang. Lit.* 58, 353–383.

Koriat, A., Sheffer, L., and Ma'ayan, L. (2002). Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *J. Exp. Psychol. Gen.* 131, 147–162. doi: 10.1037/0096-3445.131.2.147

Kroll, M. D., and Ford, M. L. (1992). The illusion of knowing, error detection, and motivational orientations. *Contemp. Educ. Psychol.* 17, 371–378. doi: 10.1016/0361-476X(92)90075-A

Lin, L.-M., Moore, D., and Zabrucky, K. M. (2001). An assessment of students' calibration of comprehension and calibration of performance using multiple measures. *Read. Psychol.* 22, 111–128. doi: 10.1080/027027101300213083

Lin, L.-C., and Yu, W.-Y. (2015). A think-aloud study of strategy use by EFL college readers reading Chinese and English texts. *J. Res. Read.* 38, 286–306. doi: 10.1111/1467-9817.12012

Lin, L.-M., and Zabrucky, K. M. (1998). Calibration of comprehension: research and implications for education and instruction. *Contemp. Educ. Psychol.* 23, 345–391. doi: 10.1006/ceps.1998.0972

Liu, S. (2002). An examination of comprehension monitoring behaviors in Chinese learners' reading. *TCSOL Stud.* 2, 1–10. doi: 10.16131/j.cnki.cn44-1669/g4.2002.03.001

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., and Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 609–616. doi: 10.1037/0278-7393.16.4.609

Maki, R. H., Jonas, D., and Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychon. Bull. Rev.* 1, 126–129. doi: 10.3758/BF03200769

Maki, R. H., Shields, M., Wheeler, A. E., and Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *J. Educ. Psychol.* 97, 723–731. doi: 10.1037/0022-0663.97.4.723

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)

Masson, M. E. J., and Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 509–527. doi: 10.1037/a0014876

Míguez-Álvarez, C., Cuevas-Alonso, M., and Cruz, M. (2021). The relationship between Metacomprehension and Reading comprehension in Spanish as a second language. *Psicol. Educ.* 28, 23–29. doi: 10.5093/psed2021a26

Moore, D., Lin-Agler, L. M., and Zabrucky, K. M. (2005). A source of metacomprehension inaccuracy. *Read. Psychol.* 26, 251–265. doi: 10.1080/02702710590962578

Pallant, J. (2007). *SPSS survival manual: a step-by-step guide to data analysis using SPSS for windows (version 15)*. Maidenhead: Open University Press.

Pieschl, S. (2009). Metacognitive calibration—an extended conceptualization and potential applications. *Metacogn. Learn.* 4, 3–31. doi: 10.1007/s11409-008-9030-4

Pressley, M., and Ghatala, E. S. (1988). Delusions about performance on multiple-choice comprehension tests. *Read. Res. Q.* 23:454. doi: 10.2307/747643

R Core Team (2020). R: A language and environment for statistical computing. Available at: https://www.R-project.org/ (Accessed March 28, 2022).

Sarac, S., and Tarhan, B. (2009). Calibration of comprehension and performance in L2 reading. *Int. Electron. J. Elem. Educ.* 2, 167–179.

Silawi, R., Shalhoub-Awwad, Y., and Prior, A. (2020). Monitoring of reading comprehension across the first, second, and third language: domain-general or language-specific? *Lang. Learn.* 70, 886–922. doi: 10.1111/lang.12410

Spellman, B. A., Bloomfield, A., and Bjork, R. A. (2008). "Measuring memory and metamemory: Theoretical and statistical problems with assessing learning (in general) and using gamma (in particular) to do so," in *Handbook of Metamemory and Memory.* ed. J. Dunlosky and R. A. Bjork (New York, NY: Psychology Press), 95–114.

Taki, S. (2016). Metacognitive online reading strategy use: readers' perceptions in L1 and L2. *J. Res. Read.* 39, 409–427. doi: 10.1111/1467-9817.12048

Thiede, K. W., Wiley, J., and Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy: Metacomprehension. *Br. J. Educ. Psychol.* 81, 264–273. doi: 10.1348/135910710X510494

Tsai, Y.-R., Ernst, C., and Talley, P. C. (2010). L1 and L2 strategy use in reading comprehension of Chinese EFL readers. *Read. Psychol.* 31, 1–29. doi: 10.1080/02702710802412081

Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., et al. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first-and second-language reading comprehension: a componential analysis. *J. Educ. Psychol.* 96, 19–30. doi: 10.1037/0022-0663.96.1.19

Wang, T. (2016a). *China Focus: Inspirational Section*. Beijing: Beijing Language and Culture University Press.

Wang, T. (2016b). *China Focus: Professional Section*. Beijing: Beijing Language and Culture University Press.

Wang, T. (2017). *China Focus: Arts Section*. Beijing: Beijing Language and Culture University Press.

Wiley, J., Griffin, T. D., Jaeger, A. J., Jarosz, A. F., Cushen, P. J., and Thiede, K. W. (2016). Improving metacomprehension accuracy in an undergraduate course context. *J. Exp. Psychol. Appl.* 22, 393–405. doi: 10.1037/xap0000096

Wu, S. (2016). Influence of vocabulary, syntactic and metacognitive awareness of Japanese students on their Chinese Reading comprehension. *Lang. Teach. Ling. Stud.* 178, 59–66.

Zabrucky, K. M., Agler, L. L., and Moore, D. (2009). Metacognition in Taiwan: students' calibration of comprehension and performance. *Int. J. Psychol.* 44, 305–312. doi: 10.1080/00207590802315409

Zabrucky, K. M., and Commander, N. E. (1993). Rereading to understand: the role of text coherence and reader proficiency. *Contemp. Educ. Psychol.* 18, 442–454. doi: 10.1006/ceps.1993.1033

Zhu, Y., and Kong, L. (2017). A qualitative research on reading abilities of the oversea postgraduates in China. *TCSOL Stud.* 65, 26–33. doi: 10.16131/j.cnki.cn44-1669/g4.20160526.002

Zimmerman, B. J., and Schunk, D. H. (2001). *Self-regulated Learning and Academic Achievement: Theoretical Perspectives.* New York: Routledge.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 23 February 2023
DOI 10.3389/fpsyg.2023.1112611

OPEN ACCESS

EDITED BY
Yang Zhang,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Lang Chen,
Santa Clara University,
United States
Wenxing Yang,
Yangzhou University,
China

*CORRESPONDENCE
Tai-Li Chou
✉ tlchou25@ntu.edu.tw

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 30 November 2022
ACCEPTED 19 January 2023
PUBLISHED 23 February 2023

CITATION
Liu T-H, Lai C-H and Chou T-L (2023) The neurocognitive basis of Chinese idiomatic constructions and processing differences between native speakers and L2 learners of Mandarin.
Front. Psychol. 14:1112611.
doi: 10.3389/fpsyg.2023.1112611

The neurocognitive basis of Chinese idiomatic constructions and processing differences between native speakers and L2 learners of Mandarin

Te-Hsin Liu[1], Chia-Ho Lai[2] and Tai-Li Chou[3,4]*

[1]Graduate Program of Teaching Chinese as a Second Language, National Taiwan University, Taipei, Taiwan, [2]Department of Psychology, New York University, New York, NY, United States, [3]Department of Psychology, National Taiwan University, Taipei, Taiwan, [4]Graduate Institute of Linguistics, National Taiwan University, Taipei, Taiwan

Classic linguistic analyses assume that syntax is the center of linguistic system. Under this assumption, a finite set of rules can produce an infinite number of sentences. By contrast, construction grammar posits that grammar emerges from language use. Chinese quadrisyllabic idiomatic expressions (QIEs) offer a testing ground for this theoretical construct owing to their high productivity. To understand the cognitive processing of structure and meaning during reading comprehension, we used a semantic judgment task to measure behavioral performance and brain activation (functional MRI). Participants were 19 Mandarin native speakers and 19 L2 learners of intermediate and advanced levels of Mandarin. In the task, participants were instructed to indicate whether the interpretation of a QIE was correct. Our behavioral results showed that L2 learners processed high frequency QIEs faster than low frequency ones. By contrast, low frequency QIEs were processed faster than high frequency ones by native speakers. This phenomenon may be attributed to semantic satiation which impedes the interpretation of high frequency QIEs. To unravel the puzzle, a further functional MRI experiment on native speakers was conducted. The results revealed that the comparison of high-frequency and low-frequency QIEs promoted significant anterior cingulate activation. Also, the comparison of idiomatic and pseudo-idiomatic constructions exhibited significant activation in the bilateral temporal poles, a region that computes semantics rather than syntactic structure. This result indicated that, for native speakers, processing Chinese idiomatic constructions is a conceptually driven process.

KEYWORDS

Chinese idioms, semantics, syntax, left anterior temporal lobe, anterior cingulate cortex, second language acquisition

1. Introduction

Before the emergence of language, infants refer to several pre-language behaviors, such as pointing, eye contact, and social reference, to communicate their intention. At the onset of word production, the language of infants is primarily composed of one word, which is the pairing of signified (concept) and signifier (sound), with syntax emerging at a later stage. Analyzing the neural substrates of meaning composition, neurolinguistic research has also identified "a system of composition that involves rapidly peaking activity in the left anterior temporal lobe" (Bemis and

Pylkkänen, 2013; Pylkkänen, 2019a,b). This brain region is related to the shared processing between comprehension and production and computes meaning rather than syntactic structure. These findings are contradictory to the traditional componential model, in which a speaker's grammatical knowledge is organized into syntactic, semantic, and phonological components, with all grammar above the word level explainable using highly general rules. Chinese quadrisyllabic idiomatic expressions (QIEs) offer a testing ground for this theoretical debate owing to their high productivity in the modern language. These expressions are to understand the cognitive processing of meaning and structure during reading comprehension, for example, the expression of [*qian-A-wan-B*] "1 k-A-10 k-B" (e.g., 1,000 army 10, 000 horse).

Idiomatic expressions are conventional expressions that have a meaning distinct from the meaning of the constituent words. Idioms involve metaphors (e.g., *love is fire*) and metonymies (e.g., *count heads*); consequently, the precise figure involved can be difficult to determine, such as in *kick the bucket* (Croft and Cruse, 2004). Although some studies have reported that idioms may be lexically flexible (e.g., *button one's lip* → *fasten one's lip*), productive (e.g., *roll out the carpet* → *sweep up the carpet*; Gibbs et al., 1989; McGlone et al., 1994) and undergo syntactic processing (Peterson et al., 2001), other studies have determined idioms to be noncompositional (Swinney and Cutler, 1979) and processed formally, as is the case with single words (Kempler et al., 1999).

Because many Chinese QIEs originate from well-known passages in classic works, they are highly synthetic and concise.[1] By contrast, the syntactic structure of idiomatic expressions in European languages is similar to regular phrase structures, such that they can undergo syntactic operations including passive voicing, insertion, topicalization, and quantification (Cacciari and Glucksberg, 1994; Cacciari, 2014). For instance, the idiom *kick the bucket* and the phrase *kick the ball* share the same grammatical structure; furthermore, most native English speakers would accept the passive version of the idiom *pop the question* (Gibbs et al., 1989). However, Chinese idioms rarely allow change in the syntactic structure (Liu et al., 2019).

The psychological basis of idiom processing has been investigated in depth over the last three decades, with the focus directed toward semantic compositionality as well as the role of frequency in idiom processing. The Principle of Semantic Compositionality, sometimes called "Frege's Principle," is based on the assumption that the whole meaning of a sentence can be reconstructed by means of its component parts (Nunberg et al., 1994; Pelletier, 1994). Compositionality means that the grammar obeys the "rule-to-rule hypothesis," such that the meaning of the whole can be inferred from the meaning of its parts. A challenge for compositionality has been observed in the interpretation of idioms. For instance, in the sentence "Peter *made no bones about* wanting to be promoted," the literal meaning of the italicized phrase is not relevant to the understanding of the sentence, and the meaning of the whole cannot be computed in a compositional sense.

Early theories of idiom comprehension relied on the assumption that a literal interpretation has priority over a figurative meaning, and the assumption was rejected only when the figurative meaning was activated (Bobrow and Bell, 1973). This line of thinking indicates the criticality of semantic transparency in idiom processing.

Taking the role of frequency into account, Cacciari and Tabossi (1988) developed the Configuration Hypothesis, arguing that idioms constitute complex arrangements of single words. An essential claim of this model is that every idiom contains one or more "keys." An idiom key acts as a type of mental signal, evoking the idiomatic configuration holistically for the hearer, which results in the activation of the idiomatic meaning. The hearer first interprets the semantics of incoming word strings based on the literal meaning, but the figurative meaning of the idiom is activated as long as adequate information is collected. Moreover, the key to the rapid retrieval of idiomatic expressions is frequency, with the major difference between idiomatic and literal expressions being that speakers and listeners are familiar with idiomatic expressions, whereas literal expressions are entirely novel. Their semantic judgment paradigm showed that there was no significant difference between decomposable and nondecomposable idioms, suggesting that the role of semantic transparency is insignificant in idiom processing. Contrarily, the high frequency of an idiom accounts for its rapid recognition. Titone and Connine (1999) advocated for a hybrid model of idiom comprehension characterizing idioms both as unitary word configurations and compositional word sequences. Based on this model, both the literal and the figurative meanings were activated. Their results of an eye-tracking experiment demonstrated that reading rates differ as a function of the inherent decompositionality of idioms. It took longer for non-compositional idioms to integrate correct meaning into the idiomatic context given that two meanings were semantically distinct.

The frequency examined by Cacciari and Tabossi (1988) and Titone and Connine (1999) refers to an idiom's "token frequency" rather than its "type frequency." Token frequency reflects the total number of words in the input, and type frequency represents the extent of use or realized productivity of a certain pattern. The productivity of linguistic patterns is reflected through type frequency rather than token frequency (Bybee and Hopper, 2001). This insight is crucial in relation to Chinese idiomatic expression, many of which are highly productive and thus have a high type frequency. For instance, 49 idiomatic expressions have the structure [*i-A-i-B*] "one-A-one-B" (Liu et al., 2019), reflecting a fixed construction and high productivity. Figure 1 illustrates the two idioms with the coordinate form, in which the first and second foot are semantically and structurally equivalent.

The restricted possibilities for the internal morphological and syntactic structures of Chinese QIEs can be analyzed by Construction Grammar (CG) which claims that constructions are the essential building blocks of human language (Fillmore et al., 1988; Goldberg, 1995). Goldberg defined a construction as "the pairing of form and meaning." For example, for *The bigger they come, the harder they fall*, proponents of CG do not reduce this proverb to a mere fixed phrase but identify a template (*The Xer, the Yer*) with "slots" that can be filled with almost any comparative phrase (e.g., *The more you want, the less you get*). Proponents of CG claim that such idiosyncratic templates are frequent in the language and that they can be understood as multiword, partially filled constructions. Construction being the pairing of form and meaning, semantic compositionality is thus derived from the interaction between the meaning of a construction and the slots filled in the construction.

Based on Construction Grammar, Liu et al. (2019) suggested that one of the meaning of the prefab [*i-A-i-B*] "one-A-one-B" is "each and every." For instance, 一草一木 [one-grass-one-tree] means "every tree and bush" and 一舉一動 [one-move-one-action] means "every movement and every action."

1 More than 90% of Chinese idiomatic expressions are quadrisyllabic (Xu, 2006). Zhou (2004) analyzed 17,934 idioms compiled in *Dictionary of Chinese Idioms* and found that 95.57% of idioms collected in the volume are quadrisyllabic.

10.3389/fpsyg.2023.1112611

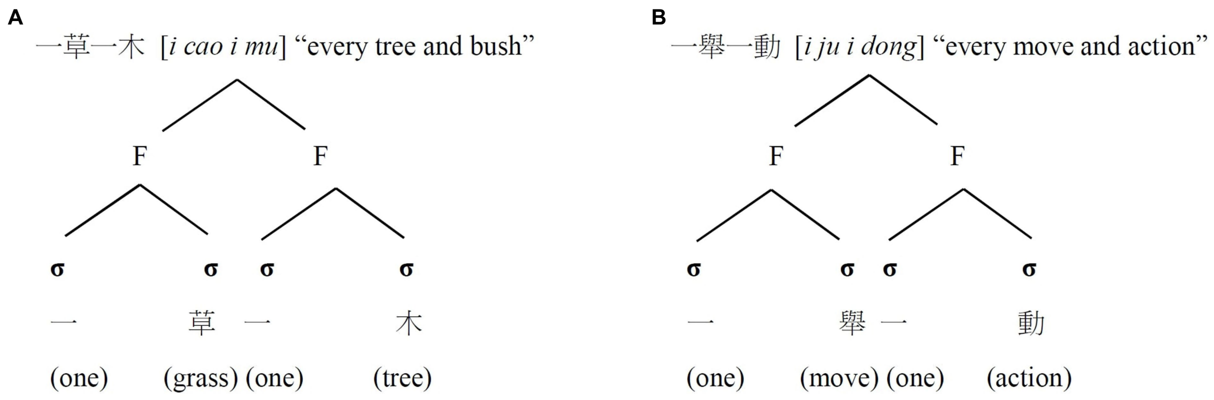


FIGURE 1
Illustration of the internal morphology of two QIEs with the same syntactic structure.

Pursuing the same line of research, Liu and Su (2021) compared the difference between Mandarin speakers and advanced learners of Mandarin in the processing of Chinese idiomatic constructions. In their experiment, 23 Mandarin speakers and 23 advanced learners of the non-Sinosphere were instructed to make semantic judgments in response to a set of QIEs. The results indicated that both the L1 group and the L2 group processed semantically transparent QIEs quicker than semantically opaque ones. The native speakers also processed low-frequency QIEs quicker than high-frequency QIEs, suggesting that semantic satiation plays a role in impeding the recognition of high-frequency idioms. A question arises as to the neural regions involved during the processing of idiomatic constructions and whether neural correlates are dominated by syntax or semantics. Meanwhile, the divergence between native speakers and L2 learners in the processing of idiomatic expressions deserves further investigation.

Arguing for the independence of syntax from semantics, classic linguistic theories place syntax at the core of language theory. Syntax has been recognized as the focal point of language production, with semantics and morphophonology representing interpretations of syntactic structures through a series of transformational rules. Research has addressed this theoretical topic, demonstrating that neither comprehension nor production is syntactic and that the left anterior temporal lobe (ATL) has a distinctly conceptual profile. Westerlund and Pylkkänen (2014) manipulated the specificity of object concepts, and analyzed the brain responses of participants processing combinations with low-specificity nouns (e.g., blue boat) or with high-specificity counterparts (e.g., blue canoe). Frequency and the transition probability between adjective and noun were carefully controlled. Their result revealed that the left ATL responded more strongly (250 ms after noun presentation) for low-specificity combinations than for high-specificity combinations, indicating that the left ATL is associated with the combinatorial aspects of the noun, such as conceptual specificity. Subsequent studies have also identified a system of composition that involves rapidly peaking activity in the left ATL followed by later engagement of the medial prefrontal cortex. Both brain regions are related to the shared processing between comprehension and production and compute meaning rather than syntactic structure (Pylkkänen, 2019a,b). Bemis and Pylkkänen (2013) and Flick et al. (2018) determined that, insofar as the task is to combine meanings, the left ATL can operate without local syntactic combinations. Thus, in the processing of natural language, neural signals are dominated by correlates of semantics rather than syntax.

Two neurolinguistic studies have investigated the processing of Chinese QIEs. Zhang et al. (2013) used event-related potentials (ERPs) to investigate the role of semantic compositionality during the interpretation of idioms. The researchers visually presented 146 QIEs of varying compositionality and nonidiomatic phrases to 18 participants as a semantic judgment task. Their behavioral results demonstrated that, in Mandarin, the meaning of individual characters remained active during the comprehension of idioms. In addition, literal nonidiomatic phrases elicited longer reaction times (RTs) compared with all the Chinese idioms. Consequently, their result demonstrated that the degrees of semantic compositionality may affect the activation of figurative meaning. Investigating the role of the right hemisphere in Chinese idiom processing, Yang et al. (2016) determined that both the left and right hemispheres contribute to the recognition of idiomatic expressions but through different pathways. Both opaque and transparent idioms elicited more activation than nonidioms in the right superior parietal lobule as well as right precuneus. Meanwhile, the activation strength was negatively correlated with the semantic transparency of the idioms. Their result contradicted the graded salience hypothesis (Giora, 1997, 2003), which predicts that the processing of familiar phrases mainly involves the left hemisphere and not the right hemisphere.

In summary, semantic transparency and frequency were both determined to play a role in Chinese idiom processing. However, which neural regions are involved in the processing of idiomatic constructions remains unclear. Do speakers prioritize meaning or syntactic structure? What is the difference between native speakers and L2 learners in the processing of idiomatic expressions? It has been shown that frequency, semantic transparency, familiarity as well as L1-L2 similarity are important factors in L2 idiom processing (Steinel et al., 2007; Garcia et al., 2015). Many Chinese idioms are highly productive and can be analyzed from a constructional perspective. The neural regions involved in the processing of idiomatic constructions can elucidate whether the neural correlates are dominated by syntax or semantics. Moreover, how construction interacts with frequency and the processing difference between native speakers and L2 learners warrant further investigation. The study of Chinese idioms, from a constructional perspective, is crucial on account of the syntactic/semantic parallelism as well as productivity of these idioms, but no fMRI-based research on this topic has been conducted to date.

To bridge the research gaps, the construction [*qian-A-wan-B*] “1 k-A-10 k-B” was selected given that it is composed of 50 idiomatic sequences. Highly productive but varying in token frequency, [*qian-A-wan-B*] "1 k-A-10 k-B" is suitable for investigating the respective role of frequency and construction. In terms of syntax, the A and B of [*qian-A-wan-B*] "1 k-A-10 k-B" can be nearly synonymous nouns, as in [*qian-chou-wan-hen*] (千仇萬恨: [1 K-hate-10 K-resentment] "deep hatred"). They can also be semibound morphemes that form a disyllabic verb, as in [*qian-ding-wan-zhu*] (千叮萬囑: [1 K-urge-10 K-advise] "exhort repeatedly"). The construction [*qian-A-wan-B*] "1 k-A-10 k-B" is symmetrical given that *qian* "1 k" and *wan* "10 k" are numerals and that A and B have equivalent syntactic function. In terms of semantics, *qian* "1 k" and *wan* "10 k" both mean "large in amount." Combined with the above expressions, the construction emphasizes the action to execute on the amount of a noun. Meanwhile, given that the construction is the pairing of form and meaning, the comprehension of idiomatic constructions does not rely on the context in which they occur; instead, the construction itself is the "context" where the native speaker relies on to interpret the meaning.

2. Materials and methods

2.1. Participants

Nineteen university-level Mandarin speakers (age = 24.1 ± 3 years old, age range 20–28 years) and 19 L2 learners of intermediate and advanced levels of Mandarin (age = 24.6 ± 3.4 years old, age range 20–32 years) took part in this study. The L2 participants were all native speakers of the non-Sinosphere.[2] All participants, right-handed, had normal or corrected-to-normal vision and had no language disorder. The study was approved by the Research Ethics Committee of our university, and all the participants provided their informed consent before participating in the experiment.

The revised version of Peabody Picture Vocabulary Test (PPVT) was administered to estimate L2 subjects' receptive lexicon (Sun et al., 2022). The L2 participants' mean score was 83.45, indicating that they had the advanced level of reading proficiency in Mandarin.

2.2. Stimuli

To obtain the frequency of idioms, the Center for Chinese Linguistics (CCL; Beijing University) Corpus containing 4.77 hundred million characters was employed at the first stage. However, seven of the 48 existing constructions were not available in the CCL corpus. We then referred to Google Search to obtain up-to-date frequencies. We acknowledge that frequency information gathered using search engines must be treated carefully; however, such information can reflect the frequency with which certain words and syntactic patterns occur (Meyer et al., 2001). We therefore individually entered the 48 [*qian-A-wan-B*] "1 k-A-10 k-B" idioms into Google Search to obtain the latest frequency information.

The experimental stimuli included the following: (1) 48 [*qian-A-wan-B*] "1 k-A-10 k-B" idiomatic constructions divided by frequency; (2) 16 pseudo [*qian-A-wan-B*] "1 k-A-10 k-B" trials violating the semantic constraint of the construction[3]; and (3) 16 non-constructional idioms, such as 龍飛鳳舞 (*long-fei-feng-wu*, "lively and vigorous in calligraphy") and 虎頭蛇尾 (*hu-tou-she-wei*, "fine start and poor finish"). A summary of the experimental stimuli is presented in Table 1. Because pseudo-idiomatic constructions do not exist in Mandarin, we can only provide word-by-word meanings. A *t* test assessed that a reliable difference in frequency existed between high-frequency and low-frequency idiomatic constructions [$t\ (23) = 4.73$, $p < 0.001$]. The linguistic material is presented in the Appendix.

2.3. Experimental paradigm

The present study used an event-related fMRI paradigm consisting of four runs. Each run had an experimental block and a baseline block. The experimental block contained 12 idiomatic constructions (divided on the basis of frequency) and four pseudo-idiomatic constructions. The baseline block consisted of four idioms that were not construction-based. All idioms were presented randomly. Each trial began with a fixation cross centered on the screen, which was displayed for 1,000 ms. Subsequently, each participant was presented with 80 stimuli. Each stimulus was displayed for 3,000 ms in random order and was followed by its meaning. The participants had at most 5 s to decide whether the displayed meaning was correct or incorrect. Before the formal test, 10 prime-target trials that were not used in the study were administrated for practice. All participants familiarized themselves with the task through this procedure. The experiment lasted approximately 15 min. The pseudo-idiomatic constructions were all conceived, and all answers in this category were thus incorrect. Before getting into the fMRI scanner, participants were reminded to keep still (Figure 2).

2.4. Image acquisition

Images were captured using a Siemens 3 T Prisma scanner with a 20-channel head coil at our university. Participants were in the scanner with their head position fixed. The head coil was placed around the participants' heads. A response box was held in the participants' right hands, and the visual materials were given to them through a goggle. Gradient-echo localizer images were used to decide the position of the functional slices. For functional images, a susceptibility-weighted single-shot echo-planar imaging (EPI) method to measure blood oxygen was used. Functional images were obtained in parallel to the AC–PC plane with interleaved EPI acquisition in the whole brain from bottom to top. The scanning parameters

2 The L2 of native Mandarin participants is English. Some of them also have a basic knowledge in Japanese. The native languages of participants in the L2 group include English (6), Spanish (6), French (3), Dutch (1), Swedish/Dutch (1), Hungarian (1), and Italian (1). The number in the parenthesis indicates the number of native speakers of a certain language.

3 Su (2002), Tsou (2012), and Liu et al. (2019) point out that Chinese idiomatic constructions are characterized by syntactic parallelism in that the slots substituted in the constructions are generally synonyms (e.g., [*qian-cun-wan-luo*] 千村萬落: [1k-village-10k-hamlet] 'thousands of villages and hamlets'), antonyms (e.g., [*yi-chang-yi-duan*] 一長一短: 'merits and demerits') or semi-bound morphemes (e.g., [*qian-hu-wan-huan*] (千呼萬喚: [1k-call-10k-shout] 'after repeated calls'). The slots inserted in pseudo-idiomatic constructions are neither synonyms nor antonyms, and thus violate the syntactic constraints of Chinese idiomatic constructions.

TABLE 1 Examples of trials in the experiment.

Condition	Example	Meaning	Mean frequency	No. of items
Idiomatic constructions				
High Frequency	千軍萬馬 *(qian-jun-wan-ma)*	A huge army	199,231	24
Low Frequency	千生萬死 *(qian-sheng-wan-si)*	Vary dangerous situation	1,764	24
Pseudo-idiomatic constructions	千竹萬酒 *(qian-zhu-wan-jiu)*	[*1k-bamboo-10k-wine*]	0.46	16
Not construction-based idioms	龍飛鳳舞 *(long-fei-feng-wu)*	Lively in calligraphy	15,252,792	16

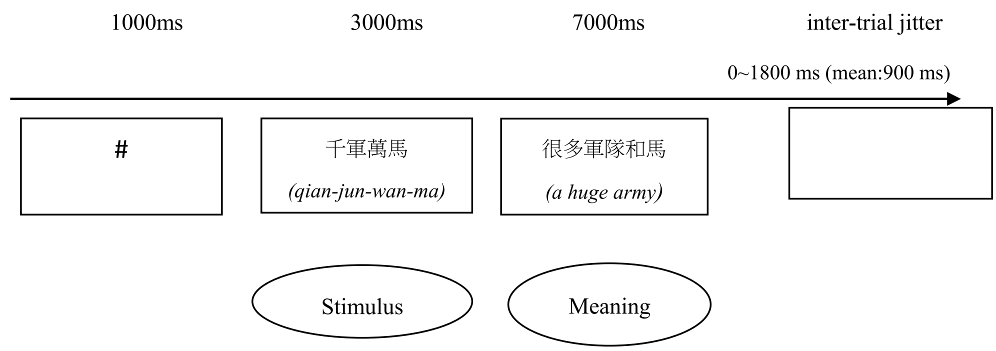


FIGURE 2
Illustration of the experimental procedure.

were as follows: TR = 2,000 ms, TE = 24 ms, flip angle = 90°, matrix size = 64 × 64, field of view = 192, slice thickness = 3.5 mm, and number of slices = 36. All participants completed four functional runs with 143 image volumes for each run (4.8 min/run, total time: 19.2 min). A high-resolution, T1-weighted 3D image was used with the following parameters: TR = 2,000 ms, TE = 2.3 ms, flip angle = 8°, matrix size = 256 × 256, field of view = 240, slice thickness = 0.94 mm, and number of slices = 192.

2.5. Image analysis

Images were preprocessed using Statistical Parametric Mapping software (SPM8, Wellcome Department of Cognitive Neurology, London, United Kingdom). Then, we used the middle slice volume to correct for functional images collected in different slice acquisition time. Functional images were realigned to the first volume in each session *via* affine transformations. No subject had more than 3 mm of movement in any plane. Co-registered images were normalized to the Montreal Neurological Institute space (voxel size: 3 × 3 × 3 mm^3). We used smoothed data for statistical analyses with a 10-mm full-width-at-half-maximum Gaussian kernel. A high-pass filter (128-s cutoff period) was applied to minimize low-frequency artifacts. For whole brain analysis, statistical analysis relied on the general linear model by using event-related analysis. Materials in the three conditions were treated as events and were modeled for the idiomatic meaning using a canonical hemodynamic response function (HRF). There were three event types: idiomatic constructions, pseudo-idiomatic constructions, and non-constructional idioms. Parameter estimates from the canonical HRF contrasts were entered into random-effects models in single-subject models by using single-sample *t* tests across participants in a whole-brain analysis. We compared idiomatic constructions to pseudo-idiomatic constructions. Among idiomatic constructions, we also compared high-frequency and low-frequency idioms. Studies have reported that the left anterior temporal lobe (or ATL, BA 38) is a critical region for compositional processing (Bemis and Pylkkänen, 2013; Pylkkänen, 2019a,b). Considering the involvement of the right hemisphere in idiom processing, we used the bilateral ATL mask from the AAL atlas for the (idiomatic constructions vs. pseudo-idiomatic constructions) contrast. In addition, Benedek et al. (2014) revealed that the dorsal posterior cingulate cortex is associated with figurative language processing. Thus, we used the bilateral midcingulum mask (dorsal part of the cingulate cortex) from the AAL atlas for (high-frequency vs. low-frequency) contrast. All the reported brain regions listed in Table 2 were FDR-corrected ($p < 0.05$) with the use the aforementioned masks as regions of interest (ROI) analysis.

3. Analysis

3.1. Study 1: Behavioral results of native speakers and L2 learners

In the L1 group, the accuracies (mean) for the constructional, pseudo-constructional, and non-constructional conditions were 83, 61, and 90%. In the L2 group, the accuracies (mean) for the constructional, pseudo-constructional, and non-constructional conditions were 68, 55, and 67%. A 2 group (L1, L2) × 3 condition (constructional, pseudo-constructional, and non-constructional) ANOVA on accuracy was performed. Mauchly's test of sphericity indicated that sphericity could not be assumed [$\chi^2(2) = 13.473$, $p = 0.001$]. To deal with violations of sphericity, we used Greenhouse–Geisser estimates of sphericity to correct the degrees of freedom. The effect of construction type was significant [F (1.5, 54.6) = 27.3, $p < 0.001$, $\eta^2 = 0.43$]. Meanwhile, the

TABLE 2 fMRI results in the whole-brain analyses.

Regions	H	BA	voxels	z value	MNI coordinates		
					x	y	z
High-frequency > Low-frequency							
Mid-Cingulate Cortex (dorsal Anterior Cingulate Cortex)	I	32	64	3.29	0	11	38
Mid-Cingulate Cortex (dorsal Posterior Cingulate Cortex)	I	31	102	3.41	0	−31	38
Idiomatic > Pseudo-idiomatic							
Temporal Pole	L	38	27	3.18	−57	5	−5
Temporal Pole	R	38	44	3.81	60	5	−5

The reported areas were FDR-corrected ($p < 0.05$) at the peak level by using masks from the AAL atlas. H, hemisphere; I, interhemispheric; L, left; R, right; and BA, brodmann area.

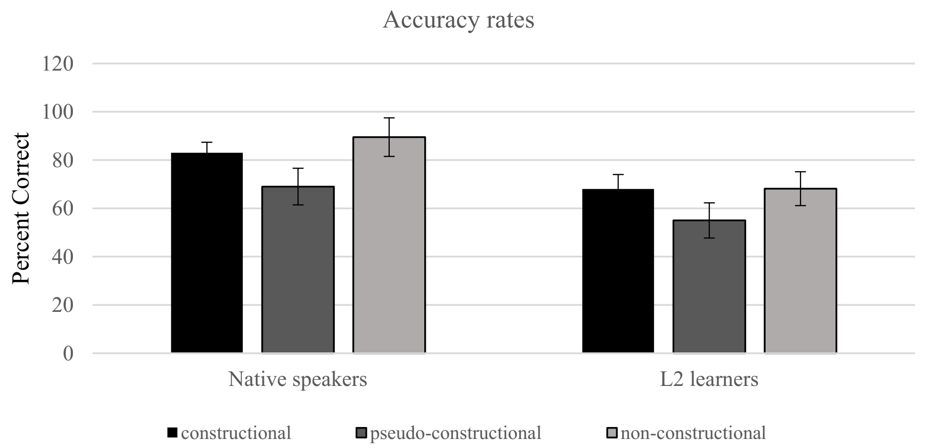


FIGURE 3
Accuracy rates for L1 group and L2 group. The error bars represent SDs.

ANOVA also revealed a significant main effect for language background. The L1 group obtained higher accuracy rates (77.79 vs. 63.71%). Meanwhile, pairwise comparisons showed a significant difference between L1 group and L2 group [F (1, 36) = 62.4, $p < 0.001$, $\eta^2 = 0.63$]. The interaction between construction type and language was also significant [F (1.5, 54.6) = 3.78, $p = 0.04$, $\eta^2 = 0.095$]. It can be observed that different types of construction led to distinct behavioral outcomes, and that the L1 group obtained better results on all types of idioms compared with the L2 group, as can be seen in Figure 3.

Further pairwise comparisons on construction type for each group of subjects were conducted. A Bonferroni correction adjusted the alpha level for the three comparisons to 0.017. Overall, there was no significant difference in accuracy rates between constructional and non-constructional idioms for both groups. Nevertheless, the accuracies between constructional and pseudo-constructional idioms were significant for native speakers ($p = 0.003$) but not significant for L2 learners, indicating that L2 learners were not aware of the agrammaticalness of pseudo-constructional idioms.

To explore how frequency and language background interact, we then conducted a two-way ANOVA to analyze the accuracy rates for constructional idiom (between-subject variable: language background; within-subject variable: frequency). The frequency × language interaction was significant [F (1, 36) = 5.40, $p = 0.026$, $\eta^2 = 0.130$]. Meanwhile, the statistical result also revealed a significant main effect for language background. The L1 group obtained higher accuracy rates (L1: 83.00% vs. L2: 68.42%). Pairwise comparisons also revealed a significant difference between L1 group and L2 group [F (1, 36) = 17.47, $p < 0.001$, $\eta^2 = 0.327$]. The effect of frequency was not significant [frequency: F (1, 36) = 1.667, $p = 0.205$]. Meanwhile, contrary to native speakers, L2 learners gained higher accuracy rates in high-frequency idioms (high-frequency idioms: 69.51% vs. low-frequency idioms: 67.32%).

The mean RTs for the constructional, pseudo-constructional, and non-constructional conditions were 1,442, 1,803, and 1,306 ms in the L1 group. The mean RTs for the constructional, pseudo-constructional, and non-constructional conditions were 2,578, 3,394, and 3,126 ms in the L2 group. A 2 group (L1, L2) × 3 condition (constructional, pseudo-constructional, and non-constructional) ANOVA on RTs was performed. The assumption of sphericity had been violated [$\chi^2(2) = 11.547$, $p = 0.003$], therefore, we used Greenhouse–Geisser estimates of sphericity to correct degrees of freedom. The effect of construction type was significant [F (1.56, 56.21) = 49.92, $p < 0.001$, $\eta^2 = 0.581$]. Meanwhile, the statistical result also revealed a significant main effect for language background. The L1 group took shorter time to respond than the L2 group (L1: 1,258 ms. vs. L2: 2,805 ms.). Pairwise comparisons revealed a significant difference between L1 group and L2 group [F (1, 36) = 95.16, $p < 0.001$, $\eta^2 = 0.726$]. The construction type × language interaction was also significant. [F (1.56, 56.21) = 10.74, $p < 0.001$, $\eta^2 = 0.23$]. Further pairwise comparisons on construction type

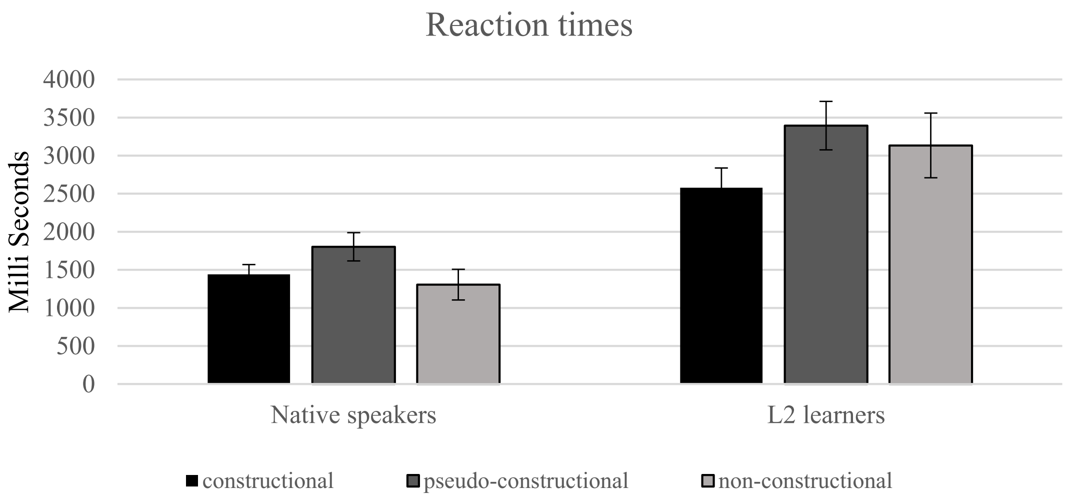


FIGURE 4
RTs for L1 group and L2 group by construction type. The error bars represent SDs.

for each group of subjects were conducted. A Bonferroni correction adjusted the alpha level for the three comparisons to 0.017. For both group of speakers, the difference in RTs between constructional and pseudo-constructional idioms was significant, with the former taking shorter time to respond. Meanwhile, for the L2 group, no significant difference in RTs was observed between non-constructional and pseudo-constructional idioms. For the L1 group, no significant difference in RTs was observed between constructional and non-constructional idioms (see Figure 4).

We further conducted a two-way ANOVA to analyze the RTs for constructional idioms (between-subject variable: language background; within-subject variable: frequency). The frequency × language interaction was significant [F (1, 36) = 5.53, p = 0.024, η^2 = 0.133]. Moreover, the statistical result revealed a significant main effect for language background. The L2 group had slower responses compared with native speakers (native: 1,364 ms. vs. L2 learners: 2,586 ms). Pairwise comparisons revealed a significant difference between L1 group and L2 group [F (1, 36) = 86.82, p < 0.001, η^2 = 0.707]. Nevertheless, the effect of frequency was not significant [frequency: F (1, 36) = 0.634, p = 0.431]. Meanwhile, contrary to native speakers, L2 learners gained slower responses in low-frequency idioms (high-frequency idioms: 2,498 ms vs. low-frequency idioms: 2,675 ms).

3.2. Study 2: fMRI results of native speakers

3.2.1. High-frequency versus low-frequency idiomatic constructions

Our behavioral results demonstrated that native speakers displayed a higher accuracy rate and shorter RTs when processing idiomatic constructions with low-frequency, implying that semantic satiation (Jakobovits, 1962) impedes the interpretation of high-frequency idioms. Semantic satiation is characterized by the attenuation of meaning after uninterrupted repetition of a word (e.g., Severance and Washburn, 1907; Bassett et al., 1919). Similar effects can be observed in cases of sensory habituation, whereby repeated perceptual stimuli fade or are transformed (Warren, 1968; Kohn, 2007). Various studies have further supported the effects of semantic satiation (Balota and Black, 1997; Black, 2001). For example, Balota and Black (1997) and Kuhl and Anderson (2011) observed that, in a decision task on semantic relatedness, weakly-related words may be more satiated following repetition than highly-related words. After the repeated priming (2, 12, or 22 times) of the stimulus, participants were visually presented a word pair and were asked to evaluate its semantic relatedness; when the prime word was repeated 22 times, the relatedness effect was diminished.

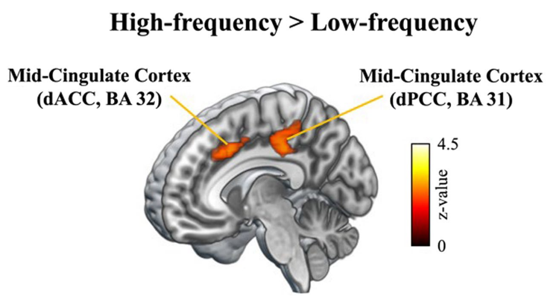


FIGURE 5
Contrast of high-frequency vs. low-frequency idiomatic constructions.

Our fMRI results demonstrated that the comparison of high-frequency and low-frequency QIEs promoted significant mid-cingulate activation (dorsal part of the anterior and posterior cingulate), suggesting that inhibitory control suppresses lasting semantic interference with high-frequency QIEs. Moreover, research has reported greater activation in the anterior cingulate cortex (ACC) in patients with obsessive–compulsive disorder and normalization of the region with treatment-induced symptom reduction (Machlin et al., 1991; Swedo et al., 1992; Perani et al., 1995; Rauch et al., 1998; Fitzgerald et al., 2005; Figure 5).

In terms of the contrast of idiomatic versus pseudo-idiomatic constructions, the participants exhibited greater activation in the bilateral ATL (BA 38), as depicted in Figure 6. It should be noted that more anterior-ventral ATL activation was found with the threshold p < 0.01 uncorrected in a whole brain analysis (also see a review by Patterson et al., 2007).

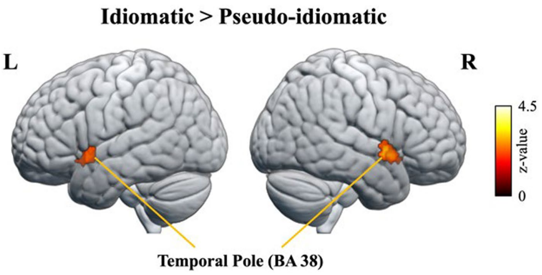


FIGURE 6
Contrast of idiomatic vs. pseudo-idiomatic constructions.

Rogalsky and Hickok (2009) examined task effects in the ATL regions, and found that some of which were activated only for the semantic task, some only for the syntactic task, and some for both tasks. Using rTMS in normal participants, Pobric et al. (2007) observed that the anterior temporal lobes are a critical substrate for semantic representation. Drawing upon recent neuropsychological evidence, and more specifically the striking disorder called semantic dementia (SD), Ralph et al. (2017) noted that bilateral anterior temporal lobe regions might be important for all conceptual domains, given that individuals with SD have semantic impairments across all modalities. In our data, given that the bilateral temporal lobe was activated, we thus infer that these regions compute semantic information.

4. Discussion

The objective of this study was 3-fold. First, we aimed to compare how native speakers and L2 learners processed different types of constructions. Second, given the intriguing behavioral result observed among native speakers, we further analyzed the neural correlates involved in the processing of Chinese idiomatic constructions. Many Chinese idioms are highly productive in the modern language and can be analyzed from a constructional perspective; consequently, the result of the study can be used to elucidate whether their processing is syntactically or conceptually driven. Third, owing to the high productivity of Chinese idiomatic constructions in the modern language, we investigated the role of frequency as well as that of construction to determine how they interact with each other.

Our behavioral results indicated that native speakers and L2 learners displayed different behaviors under the effect of frequency. The L1 group exhibited a higher accuracy rate and shorter RTs when processing idiomatic constructions with low-frequency, implying that semantic satiation (Jakobovits, 1962) impedes the interpretation of high-frequency idioms. Research has demonstrated that semantic satiation slows word associations (Balota and Black, 1997; Black, 2001). Following the assumptions of CG, the construction of idioms is the pairing of form and meaning; consequently, when processing a low-frequency idiom, participants relied on the semantics communicated by the construction to decode its meaning. In other words, intervention of a construction with direct association between form and meaning accelerates its interpretation. Nevertheless, the processing of high-frequency idioms was influenced by semantic satiation, which impeded the participants' recognition. The effect of frequency is different for L2 learners: high-frequency idiomatic constructions had better accuracy rates than low-frequency ones. The result of RTs echoed the result for accuracy rates, with high-frequency idiomatic constructions eliciting shorter RTs than low-frequency idioms for L2 learners. This result indicated that native speakers were aware of the pairing of form and meaning conveyed by the construction, while this linguistic knowledge was absent among L2 learners.

Our fMRI results for native speakers echoed the behavioral results. The comparison of high-frequency and low-frequency QIEs elicited significant anterior cingulate activation, implying that inhibitory control suppresses lasting semantic interference in relation to high-frequency QIEs. Inhibitory control is generally related to the suppression of responses to distracting stimuli (Nigg, 2000; Friedman and Miyake, 2004; Enticott et al., 2006). It has been observed that one of the brain regions heavily involved in attention as well as inhibitory control is the ACC (Posner and Petersen, 1990; Menon et al., 2001; Garavan et al., 2002; Reischies et al., 2005). Subsequent imaging studies have reported increased ACC activation during attention tasks (Devinsky et al., 1995; Cabeza and Nyberg, 1997; Elliott and Dolan, 1998) as well as response monitoring (Taylor et al., 2007). Research has also revealed that the ACC is active among bilinguals, who use this region more efficiently to monitor cognitive conflicts beyond the linguistic domain (Abutalebi et al., 2012).

Regarding the locus of the semantic satiation effect, several ERP studies have demonstrated the role of the N400 during the repetition of a given word. Kounios et al. (2000) measured ERPs in a semantic detection task which involved repeated presentations of primes and related and unrelated words using visual and auditory stimuli. Prime satiation and prime relatedness to the key word showed interaction on the ERP amplitude to key words within the N400 time window. Ströberg et al. (2017) used a 64-channel EEG system to analyze the N400 in a semantic priming task where participants were exposed to primes repeated three or 30 times. It was observed that the N400 was reduced after 30 repetitions (versus three repetitions) for the centrofrontal electrodes. Moreover, after 30 repetitions, explorative source reconstructions suggested reduced activation in wide-spread areas. These above areas and those involved in the N400 overlap, supporting the semantic rather than the perceptual nature of the satiation effect. Our fMRI studies also demonstrated that the ACC, the area dealing with inhibitory control, attention, and response monitoring, was activated during the presentation of high-frequency idiomatic constructions. Greater activation in the ACC may act as the locus of the semantic satiation effect.

Another crucial question probed in this study was the neural basis of syntax and semantics. Our result revealed that, in terms of the contrast of idiomatic versus pseudo-idiomatic constructions, the participants exhibited greater activation in the bilateral ATL (BA 38). This finding is consistent with that of Pallier et al. (2011), namely that the ATL is associated with a constituent size effect in brain activity in the presence of lexicosemantic information, suggesting that this region encodes semantic constituents. The studies of Pylkkänen (2019a,b) have also demonstrated that the ATL computes semantics rather than syntactic structure.

Patterson et al. (2007) argued that semantic generalization depends upon a distinct amodal hub located in the ATL. Crucial to semantic memory, the ATL regions are contiguous to the anterior parts of the medial temporal lobe memory system, which manages the acquisition of new semantic information. Because episodes contribute to the progressive acquisition of new conceptual knowledge, the close proximity of the episodic and semantic systems is logical. Ralph et al. (2017) reported that the ATL semantic region exhibits positive activation for semantic tasks and deactivation for nonsemantic tasks and that individuals with semantic dementia always have bilateral ATL atrophy. Guo et al. (2013) used resting-state fMRI to determine that, in healthy participants, the ATL regions exhibit intrinsic connectivity with modality-specific brain areas. It was observed that, in patients with semantic dementia, the accuracy of comprehension was correlated with the severity of ATL atrophy as well as the degree of reduced functional connectivity of hub–spoke. The aforementioned studies have deepened our understanding of the neural bases of semantic cognition. In the present study, in terms of the contrast in idiomatic versus pseudo-idiomatic constructions, the participants exhibited greater activation in the bilateral ATL (BA 38), a region that computes semantics rather than syntactic structure (Pylkkänen, 2019a,b). This result suggests that the processing of Chinese idiomatic constructions is a conceptually driven process.

Our fMRI results have theoretical implications for linguistic theory. Classic linguistic analyses suggest that languages can be studied independently of their semantic and discourse functions and that meaning is derived from a mental dictionary of words. In the tradition of generative grammar, syntax is regarded as the center of the linguistic system, with all grammar beyond the word level explainable using highly general rules (Chomsky, 2020). Using fMRI to search for the neural mechanisms of this theoretical construct, we determined that the processing of Chinese idiomatic constructions is conceptually driven, which is consistent with the CG perspective. Under this framework, construction is viewed as "the pairing of form and meaning," which is the essential foundation of language construction. Grammar emerges from language use, determined through usage patterns and frequency of use (Bybee and Hopper, 2001).

The present research did not deal with the processing of heritage language (HL) learners, but it has been shown that HL learners are more accurate than L2 learners in linguistic tasks such as sentence conjunction judgment task (Montrul et al., 2008). HL learners are also more target-like than L2 learners in the acquisition of grammatical aspect (Montrul, 2010). The interpretation of Chinese QIEs requires awareness of the conceptual metaphors underlying linguistic expressions as well as understanding of Chinese cultural tradition. Given that HL speakers grow up in a bilingual and bicultural environment, additional research is needed to probe how HL speakers interpret conflicting metaphors. For instance, "wine" has a positive connotation in English (e.g., *Old friends and old wine are best*), while it has a negative connotation in Mandarin (e.g., [jiu-chi-rou-lin] lit. "lakes of wine and forests of meat"; fig. "sumptuous entertainment").

5. Conclusion

Primed using three types of stimuli divided on the basis of frequency and construction type, native Mandarin speakers were recruited to participate in an fMRI experiment. The results revealed that the L1 group processed low-frequency QIEs quicker than high-frequency QIEs. This result may be attributed to semantic satiation which impedes the processing of high-frequency idioms. This study provides an argument in support of CG, namely that the meaning of the construction contributes to the processing of idiomatic expressions with low-frequency.

Our fMRI results also demonstrated that the comparison of high-frequency and low-frequency QIEs elicited significant ACC activation. Furthermore, the contrast of idiomatic and pseudo-idiomatic constructions elicited greater activation in the bilateral ATL, a region that computes semantics rather than syntactic structure. This finding suggests that the processing of Chinese idiomatic constructions is a conceptually driven process.

Our results reveal several avenues that could be explored in future research on foreign language acquisition. For example, the neural correlates of frequency on idiomatic constructions among L2 learners remain unclear. Meanwhile, L2 learners' metalinguistic knowledge and how this knowledge is mapped onto the comprehension of idiomatic constructions warrants further investigation. Finally, while the processing of Chinese idiomatic constructions is a conceptually driven process, further analysis needs to be conducted to explore whether L2 learners exhibit comparable processing mechanism.

Data availability statement

Original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Committee, National Taiwan University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

T-HL took part in data collection, wrote the manuscript, and analyzed the behavioral data. T-HL and T-LC designed the study and provided guidance for data analysis and manuscript writing. C-HL analyzed the imaging data. All authors contributed to the article and approved the submitted version.

Funding

This research is part of the project granted by the Ministry of Science and Technology, Taiwan (Project number: 105-2420-H-002-008-MY2) granted to the T-HL and (Project number: 105-2420-H-002-007-MY2) to the T-LC.

Acknowledgments

The authors would like to thank Lily Iwen Su, Shu-kai Hsieh, Chia-lin Lee, Chiarung Lu, and I-ni Tsai for valuable suggestions during the conception of the project. Also, we thank the Imaging Center for Integrated Body, Mind and Culture Research, National Taiwan University for technical and facility supports.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

10.3389/fpsyg.2023.1112611

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., et al. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cereb. Cortex* 22, 2076–2086. doi: 10.1093/cercor/bhr287

Balota, D. A., and Black, S. (1997). Semantic satiation in healthy young and older adults. *Mem. Cogn.* 25, 190–202. doi: 10.3758/BF03201112

Bassett, M. F., Warne, C. J., Titchener, E. B., and Weld, H. P. (1919). Minor studies from the psychological laboratory of Cornell University: on the lapse of verbal meaning with repetition. *Am. J. Psychol.* 30, 415–418. doi: 10.2307/1413679

Bemis, D. K., and Pylkkänen, L. (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cereb. Cortex* 23, 1859–1873. doi: 10.1093/cercor/bhs170

Benedek, M., Beaty, R., Jauk, E., Koschutnig, K., Fink, A., Silvia, P. J., et al. (2014). Creating metaphors: the neural basis of figurative language production. *NeuroImage* 90, 99–106. doi: 10.1016/j.neuroimage.2013.12.046

Black, S. R. (2001). Semantic satiation and lexical ambiguity resolution. *Am. J. Psychol.* 114, 493–510. doi: 10.2307/1423607

Bobrow, S. A., and Bell, S. M. (1973). On catching on to idiomatic expressions. *Mem. Cogn.* 1, 343–346. doi: 10.3758/BF03198118

Bybee, J., and Hopper, L. (2001). *Frequency and the Emergence of Linguistic Structure*. (Amsterdam: John Benjamins Publishing Company), 123–136.

Cabeza, R., and Nyberg, L. (1997). Imaging cognition: an empirical review of PET studies with Normal subjects. *J. Cogn. Neurosci.* 9, 1–26. doi: 10.1162/jocn.1997.9.1.1

Cacciari, C. (2014). Processing multiword idiomatic strings: many words in one? *Mental Lexicon* 9, 267–293. doi: 10.1075/ml.9.2.05cac

Cacciari, C., and Glucksberg, S. (1994). "Understanding figurative language" in *Handbook of Psycholinguistics*. ed. M. A. Gernsbacher (San Diego, CA, USA: Academic Press), 447–477.

Cacciari, C., and Tabossi, P. (1988). The comprehension of idioms. *J. Mem. Lang.* 27, 668–683. doi: 10.1016/0749-596X(88)90014-9

Chomsky, N. (2020). *Syntactic Structures*. Berlin/Boston: De Gruyter Mouton.

Croft, W., and Cruse, D. A. (2004). Cognitive Linguistics. Cambridge: Cambridge University Press.

Devinsky, O., Morrell, M. J., and Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behavior. *Brain* 118, 279–306. doi: 10.1093/brain/118.1.279

Elliott, R., and Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *NeuroImage* 8, 17–29. doi: 10.1006/nimg.1998.0344

Enticott, P. G., Ogloff, J. R. P., and Bradshaw, J. L. (2006). Associations between laboratory measures of executive inhibitory control and self-reported impulsivity. *Personal. Individ. Differ.* 41, 285–294. doi: 10.1016/j.paid.2006.01.011

Fillmore, C. J., Kay, P., and O'Connor, M. C. (1988). Regularity and Idiomaticity in grammatical constructions: the case of let alone. *Language* 64, 501–538. doi: 10.2307/414531

Fitzgerald, K. D., Welsh, R. C., Gehring, W. J., Abelson, J. L., Himle, J. A., Liberzon, I., et al. (2005). Error-related hyperactivity of the anterior cingulate cortex in obsessive-compulsive disorder. *Biol. Psychiatry* 57, 287–294. doi: 10.1016/j.biopsych.2004.10.038

Flick, G., Oseki, Y., Kaczmarek, A. R., Al Kaabi, M., Marantz, A., and Pylkkänen, L. (2018). Building words and phrases in the left temporal lobe. *Cortex* 106, 213–236. doi: 10.1016/j.cortex.2018.06.004

Friedman, N. P., and Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *J. Exp. Psychol. Gen.* 133, 101–135. doi: 10.1037/0096-3445.133.1.101

Garavan, H., Ross, T. J., Murphy, K., Roche, R., and Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *NeuroImage* 17, 1820–1829. doi: 10.1006/nimg.2002.1326

Garcia, O., Cieślicka, A., and Heredia, R. (2015). *Nonliteral Language Processing and Methodological Considerations*, 117–168.

Gibbs, R. W., Nayak, N. P., and Cutting, C. (1989). How to kick the bucket and not decompose: analyzability and idiom processing. *J. Mem. Lang.* 28, 576–593. doi: 10.1016/0749-596X(89)90014-4

Giora, R. (1997). Understanding Figurative and Literal Language: The Graded Salience Hypothesis. *Cognitive Linguistics* 8, 183–206. doi: 10.1515/cogl.1997.8.3.183

Giora, R. (2003). "Metaphors and Idioms," in *On Our Mind: Salience, Context, and Figurative Language*. ed. R. Giora (Oxford University Press).

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press: Chicago.

Guo, C. C., Gorno-Tempini, M. L., Gesierich, B., Henry, M., Trujillo, A., Shany-Ur, T., et al. (2013). Anterior temporal lobe degeneration produces widespread network-driven dysfunction. *Brain* 136, 2979–2991. doi: 10.1093/brain/awt222

Jakobovits, L. A. (1962). Effects of repeated stimulation on cognitive aspects of behavior: Some experiments on the phenomenon of semantic satiation. Available at: https://escholarship.mcgill.ca/concern/theses/c821gp587 (Accessed January 19, 2023).

Kempler, D., Van Lancker, D., Marchman, V., and Bates, E. (1999). Idiom comprehension in children and adults with unilateral brain damage. *Dev. Neuropsychol.* 15, 327–349. doi: 10.1080/87565649909540753

Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *J. Neurophysiol.* 97, 3155–3164. doi: 10.1152/jn.00086.2007

Kounios, J., Kotz, S. A., and Holcomb, P. J. (2000). On the locus of the semantic satiation effect: evidence from event-related brain potentials. *Mem. Cogn.* 28, 1366–1377. doi: 10.3758/bf03211837

Kuhl, B. A., and Anderson, M. C. (2011). More is not always better: paradoxical effects of repetition on semantic accessibility. *Psychon. Bull. Rev.* 18, 964–972. doi: 10.3758/s13423-011-0110-0

Liu, T.-H., Lü, C.-R., Tsai, I.-N., and Su, L. I.-W. (2019). Idiomaticity and regularity in Chinese four-character idioms: the case of [yi-X-#-Y]. *Tsing Hua J. Chin. Stud.* 49, 683–719. doi: 10.6503/THJCS.201912_49(4).0004

Liu, T.-H., and Su, L. I.-W. (2021). Chinese idioms as constructions: frequency, semantic transparency and their processing. *Lang. Linguist.* 22, 558–592. doi: 10.1075/lali.00091.liu

Machlin, S. R., Harris, G. J., Pearlson, G. D., Hoehn-Saric, R., Jeffery, P., and Camargo, E. E. (1991). Elevated medial-frontal cerebral blood flow in obsessive-compulsive patients: a SPECT study. *Am. J. Psychiatry* 148, 1240–1242. doi: 10.1176/ajp.148.9.1240

McGlone, M. S., Glucksberg, S., and Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Process.* 17, 167–190. doi: 10.1080/01638539409544865

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., and Reiss, A. L. (2001). Error-related brain activation during a go/NoGo response inhibition task. *Hum. Brain Mapp.* 12, 131–143. doi: 10.1002/1097-0193(200103)12:3<131::aid-hbm1010>3.0.co;2-c

Meyer, C., Grabowski, R., Han, T., Mantzouranis, K., and Moses, S. (2001). The World Wide Web as Linguistic Corpus.

Montrul, S. (2010). How similar are adult second language learners and Spanish heritage speakers? Spanish clitics and word order. *Appl. Psycholinguist.* 31, 167–207. doi: 10.1017/S014271640999021X

Montrul, S., Foote, R., and Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: the effects of age and context of acquisition. *Lang. Learn.* 58, 503–553. doi: 10.1111/j.1467-9922.2008.00449.x

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychol. Bull.* 126, 220–246. doi: 10.1037/0033-2909.126.2.220

Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms. *Language* 70, 491–538. doi: 10.1353/lan.1994.0007

Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci.* 108, 2522–2527. doi: 10.1073/pnas.1018711108

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277

Pelletier, F. J. (1994). The principle of semantic compositionality. *Topoi* 13, 11–24. doi: 10.1007/BF00763644

Perani, D., Colombo, C., Bressi, S., Bonfanti, A., Grassi, F., Scarone, S., et al. (1995). [18F] FDG PET study in obsessive-compulsive disorder. A clinical/metabolic correlation study after treatment. *Br. J. Psychiatry* 166, 244–250. doi: 10.1192/bjp.166.2.244

Peterson, R. R., Burgess, C., Dell, G. S., and Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1223–1237. doi: 10.1037/0278-7393.27.5.1223

Pobric, G., Jefferies, E., and Ralph, M. A. L. (2007). Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc. Natl. Acad. Sci.* 104, 20137–20141. doi: 10.1073/pnas.0707383104

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.* 13, 25–42. doi: 10.1146/annurev.ne.13.030190.000325

Pylkkänen, L. (2019a). Neural basis of basic composition: what we have learned from the red–boat studies and their extensions. *Philos. Trans. Roy. Soc. B. Biol. Sci.* 375:20190299. doi: 10.1098/rstb.2019.0299

Pylkkänen, L. (2019b). The neural basis of combinatory syntax and semantics. *Science* 366, 62–66. doi: 10.1126/science.aax0050

Ralph, M. A. L., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150

Rauch, S. L., Dougherty, D. D., Shin, L. M., Alpert, N. M., Manzo, P., Leahy, L., et al. (1998). Neural correlates of factor-analyzed OCD symptom dimensions: a PET study. *CNS Spectr.* 3, 37–43. doi: 10.1017/S1092852900006167

Reischies, F. M., Neuhaus, A. H., Hansen, M. L., Mientus, S., Mulert, C., and Gallinat, J. (2005). Electrophysiological and neuropsychological analysis of a delirious state: the role of the anterior cingulate gyrus. *Psychiatry Res.* 138, 171–181. doi: 10.1016/j.pscychresns.2004.06.008

Rogalsky, C., and Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cereb. Cortex* 19, 786–796. doi: 10.1093/cercor/bhn126

Severance, E., and Washburn, M. F. (1907). Minor studies from the psychological laboratory of Vassar college: the loss of associative power in words after long fixation. *Am. J. Psychol.* 18, 182–186. doi: 10.2307/1412411

Steinel, M. P., Hulstijn, J. H., and Steinel, W. (2007). SECOND LANGUAGE IDIOM LEARNING IN a PAIRED-ASSOCIATE PARADIGM: effects of direction of Learning, direction of testing, idiom Imageability, and idiom transparency. *Stud. Second. Lang. Acquis.* 29, 449–484. doi: 10.1017/S0272263107070271

Ströberg, K., Andersen, L. M., and Wiens, S. (2017). Electrocortical N400 effects of semantic satiation. *Front. Psychol.* 8:2117. doi: 10.3389/fpsyg.2017.02117

Su, L. I. (2002). Why a construction—that is the question! *Concent. Stud. English Literat. Linguist.* 28, 27–42.

Sun, X., Zhang, K., Marks, R. A., Nickerson, N., Eggleston, R. L., Yu, C.-L., et al. (2022). What's in a word? Cross-linguistic influences on Spanish-English and Chinese-English bilingual children's word reading development. *Child Dev.* 93, 84–100. doi: 10.1111/cdev.13666

Swedo, S. E., Pietrini, P., Leonard, H. L., Schapiro, M. B., Rettew, D. C., Goldberger, E. L., et al. (1992). Cerebral glucose metabolism in childhood-onset obsessive-compulsive disorder. *JAMA Psychiatry* 49, 690–694. doi: 10.1001/archpsyc.1992.01820090018003

Swinney, D. A., and Cutler, A. (1979). The access and processing of idiomatic expressions. *J. Verbal Learn. Verbal Behav.* 18, 523–534. doi: 10.1016/S0022-5371(79)90284-6

Taylor, S. F., Stern, E. R., and Gehring, W. J. (2007). Neural systems for error monitoring: recent findings and theoretical perspectives. *Neuroscientist* 13, 160–172. doi: 10.1177/1073858406298184

Titone, D. A., and Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *J. Pragmat.* 31, 1655–1674. doi: 10.1016/S0378-2166(99)00008-9

Tsou, B. K. (2012). "Idiomaticity and classical traditions in some east Asian languages." in *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation*. Bali, Indonesia: Faculty of Computer Science, Universitas Indonesia), 39–55. Available at: https://aclanthology.org/Y12-1003 (Accessed January 19, 2023).

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin* 70, 261–270. doi: 10.1037/h0026275

Westerlund, M., and Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia* 57, 59–70. doi: 10.1016/j.neuropsychologia.2014.03.001

Xu, S. (2006). Xianglin yu buzu: chengyu xingcheng de renzhi yanjiu zhi yi. *Sichuan Waiyu Xueyuan Xuebao* 22, 107–111.

Yang, J., Li, P., Fang, X., Shu, H., Liu, Y., and Chen, L. (2016). Hemispheric involvement in the processing of Chinese idioms: an fMRI study. *Neuropsychologia* 87, 12–24. doi: 10.1016/j.neuropsychologia.2016.04.029

Zhang, H., Yang, Y., Gu, J., and Ji, F. (2013). ERP correlates of compositionality in Chinese idiom comprehension. *J. Neurolinguistics* 26, 89–112. doi: 10.1016/j.jneuroling.2012.05.002

Zhou, J. (2004). Cihuixue cidianxue yanjiu. Beijing: Shangwu Yinshuguan.

10.3389/fpsyg.2023.1112611

Appendix

Idiomatic constructions used for targets.

High frequency idiomatic constructions		Low frequency idiomatic constructions	
千變萬化 Qian-bian-wan-hua	Constant permutations	千仇萬恨 Qian-chou-wan-hen	Deep hatred
千辛萬苦 Qian-xin-wan-ku	Much hardship	千歡萬喜 Qian-huan-wan-xi	Extremely happy
千年萬載 Qian-nian-wan-zai	A very long time	千刀萬剁 Qian-dao-wan-duo	A 1,000 cuts and myriad pieces
千軍萬馬 Qian-jun-wan-ma	A huge army	千變萬狀 Qian-bian-wan-zhuang	Have much variety
千山萬水 Qian-shan-wan-shui	A long and arduous journey	千端萬緒 Qian-duan-wan-xu	With many thoughts in mind
千叮萬囑 Qian-ding-wan-zhu	Exhort repeatedly	千轉萬變 Qian-zhuan-wan-bian	Constant permutations
千差萬別 Qian-cha-wan-bie	Completely different	千嬌萬態 Qian-jiao-wan-tai	Beautiful appearance and figure
千家萬戶 Qian-jia-wan-hu	Every family	千村萬落 Qian-cun-wan-luo	Many villages
千難萬險 Qian-nan-wan-xian	Many hazards and difficulties	千依萬順 Qian-yi-wan-shun	Always obedient
千真萬確 Qian-zhen-wan-que	Absolutely true	千變萬態 Qian-bian-wan-tai	Constant permutations
千恩萬謝 Qian-en-wan-xie	Thank again and again	千思萬慮 Qian-si-wan-lü	Think or consider repeatedly
千態萬狀 Qian-tai-wan-zhuang	Have much variety	千支萬派 Qian-zhi-wan-pai	Many (philosophical, martial, etc.) sects
千峰萬壑 Qian-feng-wan-huo	Many mountains and valleys	千兵萬馬 Qian-bing-wan-ma	A huge army
千秋萬世 Qian-qiu-wan-shi	A long, long time	千回萬轉 Qian-hui-wan-zhuan	Go through ups and downs
千門萬戶 Qian-men-wan-hu	A big house or lots of inhabitants	千章萬句 Qian-zhang-wan-ju	Many phrases and articles
千呼萬喚 Qian-hu-wan-huan	Call repeatedly	千倉萬箱 Qian-cang-wan-xiang	Massive storage of food
千山萬壑 Qian-shan-wan-huo	Many mountains and valleys	千推萬阻 Qian-tui-wan-zu	Do everything to decline
千條萬縷 Qian-tiao-wan-lü	Many threads	千緒萬端 Qian-xu-wan-duan	Tangled thoughts
千乘萬騎 Qian-sheng-wan-ji	Lots of carriages and cavalry	千生萬死 Qian-sheng-wan-si	Very dangerous situation
千言萬語 Qian-yan-wan-yu	Have many words to say	千齡萬代 Qian-ling-wan-dai	Generation after generation
千刀萬剮 Qian-dao-wan-gua	A 1,000 cuts and myriad pieces	千匯萬狀 Qian-hui-wan-zhuang	Have much variety
千頭萬緒 Qian-tou-wan-xu	So many thoughts in one's mind	千言萬說 Qian-yan-wan-shuo	Have many words to say
千絲萬縷 Qian-tiao-wan-lü	Tangled connections	千妥萬當 Qian-tuo-wang-dang	Very appropriate
千秋萬歲 Qian-qiu-wan-sui	A long, long time	千岩萬穀 Qian-yan-wan-gu	A group of mountain ranges

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 22 March 2023
DOI 10.3389/fpsyg.2023.1134229

Check for updates

OPEN ACCESS

EDITED BY
Linjun Zhang,
Peking University,
China

REVIEWED BY
Xingsan Chai,
Beijing Language and Culture University,
China
Jie Dong,
Xi'an International Studies University,
China
Tao Zeng,
Hunan University,
China

*CORRESPONDENCE
Lijuan Feng
✉ nkgalina@163.com

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 30 December 2022
ACCEPTED 24 February 2023
PUBLISHED 22 March 2023

CITATION
Feng L and Jiang N (2023) Prediction error cost exists in the reading processing of Chinese native speakers and advanced Chinese L2 learners.
Front. Psychol. 14:1134229.
doi: 10.3389/fpsyg.2023.1134229

Prediction error cost exists in the reading processing of Chinese native speakers and advanced Chinese L2 learners

Lijuan Feng[1]* and Nan Jiang[2]

[1]School of Chinese Language and Literature, University of Chinese Academy of Social Sciences, Beijing, China, [2]School of Languages, Literatures, and Cultures, University of Maryland, College Park, MD, United States

This study applies the paradigm of self-paced reading to examine the Context Predictability Effect in the processing of Chinese and detect whether there is a prediction error cost. Context constraint strength (constraining and neutral) and word predictability (predictable and unpredictable) were strictly manipulated. The statistical results suggest that: (1) There is a Context Predictability Effect for Chinese native speakers in reading processing, which is consistent with most previous studies; (2) There is also a Context Predictability Effect for advanced Chinese L2 learners; (3) Both Chinese native speakers and Chinese L2 learners have a prediction error cost in reading processing, a finding different from those of much previous research. (4) Chinese L2 learners are significantly slower than Chinese native speakers when they conduct predictive reading processing. This paper is very enlightening in that it identifies the existence of a prediction error cost in Chinese L2 processing by means of behavioral experiments, providing evidence for the hypothesis of Lexical Prediction. In a strongly predictive setting, when encountering a plausible but unpredictable word, the brain must expend extra effort to suppress, revise, or reanalyze the material, and this may account for the prediction error cost.

KEYWORDS

context predictability effect, prediction error cost, self-paced reading, lexical prediction, graded prediction

1. Introduction

1.1. Context predictability effect in L1 processing

Context predictability or context constraint is an important factor affecting the reading processing of words. It refers to the probability that readers predict the target word based on the previous information. The more constrained the context, the fewer words that meet the conditions, and the greater the possibility of the target word being guessed; the weaker the constraint, the more words meet the conditions, the smaller probability of the target word being guessed. This is the Context Predictability Effect (Li et al., 2022).

There is a stable contextual predictive effect in L1 reading processing in both Western languages and Chinese (Zhu, 1991; Bai et al., 2011; Roland et al., 2012; Staub, 2015; Kwon et al., 2017; Chow et al., 2018; Zhang et al., 2019, 2023; Zhao, 2022). In eye movement studies, the constraint of the context can affect the target word fixation time: the more constrained the context is, the shorter the fixation time will be; the less constrained the context, the longer the fixation time (Rayner and Well, 1996; Rayner et al., 2004, 2005, 2006). Rayner and Well (1996) divided contexts into three types, high-constraint, medium-constraint, and low-constraint, and investigated the influence of context constraint on eye movement in reading. It was found that the fixation time of participants in the low-constraint context was longer than that in the high-constraint and medium-constraint contexts. Rayner et al. (2005) used a Chinese corpus as the research object and copied the research process of Rayner and Well (1996), and the results were highly consistent with Rayner and Well (1996). In addition, word skipping is more likely to occur in high- contexts than in low-constraint contexts (Rayner et al., 2005; Bai et al., 2008; Guo, 2012; Zhang, 2015; Li, 2016). This phenomenon also exists among child readers. Children under the high-constraint context had greater skipping rates and spent less time reading (Zhao, 2020).

There are also many studies on the interaction between context predictability and other factors, as follows: Liu et al. (2020) explored how context predictability affects Chinese vocabulary processing in reading by observing the interaction between context predictability, whole word frequency, and Chinese character frequency. The results showed that context predictability, whole word frequency, and first character frequency affect vocabulary processing relatively independently; contextual predictability directly affects the processing of the second character within words. Liu et al. (2019) divided children's reading skills into high and low groups and used the boundary paradigm, which changes with fixation, to investigate the impact of context predictability on children's parafoveal processing. The results showed that children with high reading skills had earlier use of context predictability than children with low reading skills. Word frequency and context predictability also have an impact on the reading of the elderly (Wang et al., 2012), who adopt more cautious reading strategies when the difficulty of reading content is too high or too low. Under the right conditions, older adults rely more heavily on contextual predictive information than younger adults (Zhao, 2009).

In summary, the context predictability effect is mainly manifested in the promotion of target word processing by high-constraint contexts; that is, in high-constraint contexts, readers process target words faster and with lower processing difficulty (Wang, 2017; Zhao et al., 2021). Conversely, if an unpredictable word appears instead of a predictable word in a high-constraint context, will the reader's processing be disturbed? In other words, will there be a prediction error cost? This is another issue we are concerned about.

1.2. Prediction error cost in L1 processing

At present, researchers have not yet reached a consensus on whether the prediction error cost exists. There are two main views. One view is that there is a prediction error cost (Coulson and Van Petten, 2002; Moreno et al., 2002; Wicha et al., 2004; DeLong et al., 2005, 2007, 2011, 2012; Van Berkum et al., 2005; Federmeier et al., 2007). This view holds that in a high-constraint context, the specific vocabulary with the highest cloze score is activated first, and when the reader encounters a word other than the target word in the reading process, it will violate the previous prediction, resulting in processing interference. This is the Lexical Prediction view. Another view is that there is no prediction error cost (Luke and Christianson, 2016; Frisson et al., 2017; Zhao et al., 2021; Zhao, 2022). This view holds that the reader does not predict a specific word in reading processing, but activates a series of words in parallel to varying degrees. This is the Graded Prediction view.

Luke and Christianson (2016) used a method of large-scale survey to explore the role of context in reading processing. The study did not find a prediction error cost and believed that the high-constraint context activates not a whole word, but more semantic and morphosyntactic information related to it, thus supporting the Graded Prediction view. Frisson et al. (2017) used the method of controlled experimental design for the first time to provide evidence for the conclusion of Luke and Christianson (2016). The experiment adopts a design of 2 (context constraint strength: constraining and neutral) * 2 (word predictability: predictable and unpredictable). The context constraint strength is one variable and this variable has two levels, which are high-constraint context and low-constraint context; another variable is lexical predictability, which also has two levels, predictable words, and unpredictable words. The predictable words here are the specific words that have the greatest possibility of being predicted in the high-prediction context we discussed above, and the unpredictable words are other words that cannot be predicted in advance but reasonable semantically. The innovation of this study is to judge whether there is a prediction error cost by comparing the processing time of unpredictable words in high-constraint contexts and low-constraint (neutral) contexts. The basic logic is: if there is a prediction error effect, then the processing difficulty of unpredictable words in a high-constraint context must be greater than that in a low-constraint context, and correspondingly more processing time will be spent. The experiment did not find a prediction error cost, supporting the view of Graded Prediction.

We believe that one of the reasons why the prediction error cost was not found may be that the constraint of the high-constraint context is not enough, Frisson et al. (2017) is only 70.2%, Zhao et al. (2021) improved this problem by increasing the constraint in the high-constraint context, but still found no prediction error effect. The key point is that the eye movement data of first fixation time and skipping rate in Zhao et al. (2021) reflect the situation of early parafoveal processing, while the results of ERP research on prediction error cost indicated that the interference on the processing of unpredictable words in a high-constraint context is likely to occur in the later stage (Moreno et al., 2002; Federmeier, 2007; Federmeier et al., 2007; Delong et al., 2011; DeLong et al., 2012).

1.3. Context predictability effect in L2 processing

On whether there is a context predictability effect in second language reading processing, researchers have not yet reached a consensus. Some studies have shown that even at an advanced level, L2 learners cannot predict information during reading processing, at least not to the same extent as native speakers (Kaan et al., 2010; Grüter and Rohde, 2013; Kaan, 2014). The difference in grammatical processing between L2 learners and native speakers is mainly affected by the following factors: incomplete acquisition of the target grammar, cognitive limitation of the target language, interference from the grammar and processing system of first language, etc. Other studies suggest that L2 learners can make predictions just like native speakers (Clahsen and Felser, 2006; Chambers and Cooke, 2009; Dussias et al., 2013). In Spanish grammar, there is a distinction between feminine and masculine nouns, and the modifiers should be consistent in gender with the nouns. Dussias et al. (2013) took advantage of this characteristic of Spanish grammar to investigate whether L2 learners can activate nouns through the gender marker of the modifiers. The results showed that advanced L2 learners of English showed the same context predictability effect as native speakers.

1.4. Prediction error cost in L2 processing

We only found one article on prediction error cost in L2 processing. Zirnstein et al. (2018) used ERP experiments to prove that L2 learners can not only predict upcoming words through context but also generate prediction error cost like native speakers. The performance of L2 learners in processing unpredictable words is related to their mastery of native language regulations. L2 learners with a better grasp of native language regulations have larger frontal positivity when processing unpredictable words, and this effect is attenuated by cognitive control, especially inhibitory control ability. Inhibition control ability appears to mediate the difficulty readers incur.

1.5. Current study

We took Chinese native speakers and advanced Chinese L2 learners as research objects, used the method of self-paced reading, and replicated the experimental design of Frisson et al. (2017). The following are research questions: 1. Are there context predictability effect and prediction error cost in the reading processing for Chinese native speakers? 2. Are there context predictability effect and prediction error cost in the reading processing for advanced Chinese L2 learners? 3. Is there any difference in this regard between Chinese native speakers and advanced Chinese L2 learners? Our experimental design has made the following improvements: 1. We improved the constraint strength of high-constraint contexts based on Frisson et al. (2017); 2. We used the method of self-paced reading, because this method is closest to natural reading (Mitchell and Green, 1978, p. 610); 3. We examined the potential differences of prediction mechanism between Chinese native speakers and Chinese L2 learners. Doing so can provide theoretical support and guidance for the teaching of Chinese as a second language.

2. Experiment 1

2.1. Method

2.1.1. Participants

Twenty six participants (10 males) were recruited from a university in Hebei, China, aged 18–34 ($M = 20.30$, SD = 3.08). These participants are all Chinese speaking, have undergone at least some undergraduate education, and have normal vision or corrected vision.

2.1.2. Design

Experiment 1 adopted a two-factor within-subjects design, which is 2 (context constraint strength: constraining context and neutral context) * 2 (word predictability: predictable word and unpredictable word). As shown in Table 1.

2.1.3. Materials

We selected 106 contexts from Chinese textbooks at junior and intermediate level, including 53 high-constraint and 53 low-constraint contexts. Each context contains one predictable word and one unpredictable word. For example: "大卫送给妈妈的生日礼物/卡片让妈妈特别感动 (David's birthday gift/card to her mother especially touched her)" is a high-constraint context. In this high-constraint context, "礼物 (gift)" is a predictable word, and "卡片 (card)" is an unpredictable word; "我看到这个礼物/卡片的时候非常惊讶 (When I saw this gift/card, I was very surprised)" is a low-constraint context that also contains the words "礼物 (gift)" and "卡片 (card)." To avoid the interference of the packing effect at the end of the sentence, the position of the target word is set in the middle of each sentence.

First, we removed the target words from 53 high-constraint contexts and asked 31 native Chinese speakers to fill in the blanks with the first word that comes to their mind. In previous studies, the entire sentence was presented at once. But in the self-paced experiment, the vocabulary is presented sequentially, and the participants can only predict through the context before the target word. To match the real reading process as much as possible, we only presented the content before the blanks. After getting the cloze data, we selected the word given by the largest number of people as the predictable word. The ratio of the number of people giving the target word to the total number of people is called the constraint strength (cloze value). We deleted the high-constraint contexts with a constraint strength below 60%, and the average constraint strength reached 82.4%. We finally retained 37 high-constraint contexts, plus 37 low-constraint contexts, for a total of 74 contexts. Secondly, the 74 contexts were divided into 148 sentences, which were divided into two groups by means of Latin squares. Then, 29 and 41 native Chinese speakers were invited to judge the semantic rationality of the sentences, using the Likert scale. Sentences with an average value above 4 were selected, and a total of 128 target sentences were obtained.

The target sentences were divided into 4 groups. The constraint strength of constraining context with predictable word (CP) is 82.4%, that of constraining context with unpredictable word (CU) is 3.7%, and that of neutral context with predictable word (NP) is 2.4%, while neutral context with unpredictable word (NU) is 2.3%. The results of

TABLE 1 Sample items in Experiment 1.

Conditions	Example
Constraining context-Predictable word (CP)	大卫送给妈妈的生日礼物让妈妈特别感动。 David's birthday gift to her mother especially touched her.
Constraining context-Unpredictable word (CU)	大卫送给妈妈的生日卡片让妈妈特别感动。 David's birthday card to her mother especially touched her.
Neutral context- Predictable word (NP)	我看到这个礼物的时候非常惊讶。 When I saw this gift, I was very surprised.
Neutral context- Unpredictable word (NU)	我看到这个卡片的时候非常惊讶。 When I saw this card, I was very surprised.

The target word is underlined.

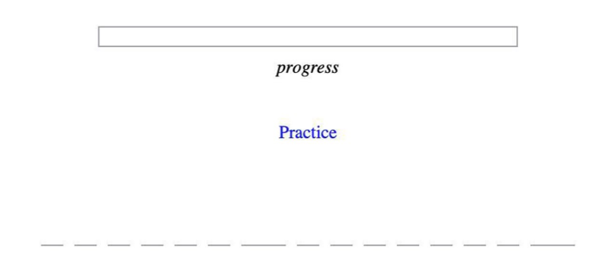


FIGURE 1
Practice page of the formal experiment.

paired-sample *t*-tests showed that constraint strength of CP was significantly higher than that of CU, $t(31) = -33.114$, $p < 0.001$, and NP, $t(31) = -35.065$, $p < 0.001$; there was no significant difference between CU and NU, $t(31) = 1.603$, $p = 0.119$. We balanced character frequency and number of strokes of predictable and unpredictable words from the first character to the fifth character, first character frequency: $t(31) = 0.051$, $p = 0.960$, first character stroke number: $t(31) = -0.651$, $p = 0.520$; second character frequency: $t(31) = 1.006$, $p = 0.322$, second character number of strokes: $t(31) = 0.896$, $p = 0.377$; third character frequency: $t(31) = 0.424$, $p = 0.674$, third character number of strokes: $t(31) = 0.056$, $p = 0.956$; fourth character frequency: $t(31) = -0.303$, $p = 0.764$, fourth character number of strokes: $t(31) = 0.823$, $p = 0.417$; and fifth character frequency: $t(31) = -1.657$, $p = 0.108$, fifth character number of strokes: $t(31) = -0.113$, $p = 0.911$. Character frequency statistics are based on SUBTLEX-CH (Cai and Brysbaert, 2010), and the unit is times/million. In the self-paced reading experiment, we also added 49 semantically unreasonable sentences as distractors.

2.1.4. Procedure

We set up the self-paced reading experiment at.[1] The system automatically divided the experimental materials into two groups according to the Latin square and randomly assigned them to the participants. Using the online experiment method can make participants more relaxed and get closer to the state of natural reading. Before the experiment, participants were required to prepare a computer in an environment with normal network speed. After accessing the webpage, participants saw a brief description of the experiment and needed to fill in information such as mother tongue, age, years of learning Chinese, gender, and email address first; then, they were required to finish the experiment as quickly and as well as possible. The first page of the formal experiment is as shown in Figure 1.

The participant should press the space bar of the computer to start reading. Sentences are presented in units of Chinese characters, from left to right; participants press the space bar once to make the next Chinese character appear. After they finish reading a sentence, participants are required to answer a multiple-choice question. There are 6 practice sentences, and the target materials start from the seventh sentence. Each participant needs to read 113 sentences (64 target sentences, 49 distractor items), and the experiment takes 15–20 min in total. The system records the reaction time for each Chinese character and the correct rate of answers.

We did not present word by word for the following reasons: First, Chinese words vary in length. In Chinese, although two-syllable words are dominant, there are still many monosyllabic words, such as "是 (is)," and three-syllable words, such as "有时候 (sometimes)," and even words with more syllables. Therefore, the reaction time for given words is affected by their number of characters. Second, word-by-word presentation is not conducive to the analysis of the spillover effect. We analyzed the reaction time for three more Chinese characters after the target word; if stimuli were presented word by word, the spillover effect was not easy to measure. Third, character-by-character presentation is the closest to the typesetting of authentic reading materials in Chinese.

2.2. Results

The correct rate of the answers was above 90%, which proved that the participants had read the sentences carefully. We recorded the reaction time for the target word, which contained two characters, and the three characters following it. We deleted the data according to the following criteria: (1) if the correct rate of a participant is less than 60%, delete all the data for that participant; (2) data whose reaction time is shorter than 80 ms but greater than 1,500 ms; (3) data beyond ±3 standard deviations. A total of 492 data points were deleted, accounting for 4.6% of the total data.

We constructed linear mixed models of the self-paced reading data, using the lme4 package (Bates et al., 2015) in R version 4.2.2. We mainly examine the responses in five regions: the first character of the target word, the second character of the target word, and the third, fourth and fifth characters after that. The extra three characters after the target word are examined because there is a spillover effect in the self-paced experiment.

There was a significant difference in reaction time from the first character to the fifth character between the conditions of CP and NP: first character: $t(223) = -3.967$, $p < 0.01$; second character: $t(347) = -4.859$, $p < 0.01$; third character: $t(348) = -5.128$, $p < 0.01$; fourth character: $t(345) = -5.300$, $p < 0.01$; fifth character: $t(347) = -5.609$, $p < 0.01$, and CP < NP (see Figure 2). This shows that processing time for predictable words in high-constraint contexts is significantly shorter than that in low-constraint contexts. There was also a significant difference in reaction time from the first character to the fifth character between the conditions of CP and CU, first character: $t(223) = -3.788$, $p < 0.01$; second: $t(346) = -3.832$, $p < 0.01$; third: $t(350) = -4.222$, $p < 0.01$; fourth: $t(345) = -4.753$, $p < 0.01$; fifth: $t(345) = -5.112$, $p < 0.01$, and CP < CU (see Figure 2). This suggests that in high-constraint contexts, processing time for predictable words is significantly shorter than that for unpredictable words. The above two points confirm the existence of the context predictability effect in Chinese native speakers' reading processing.

There was a significant difference in reaction time from the first character to the fifth character between the conditions of CU and NU, first character: $t(221) = 3.145$, $p < 0.05$; second character: $t(349) = 3.439$, $p < 0.05$; third character: $t(349) = 4.014$, $p < 0.05$; fourth character: $t(344) = 4.448$, $p < 0.01$; fifth character: $t(344) = 5.007$, $p < 0.01$, and CU > NU (see Figure 2). The processing time of unpredictable words in the high-constraint context is significantly longer than that in

1 https://spellout.net/ibexfarm

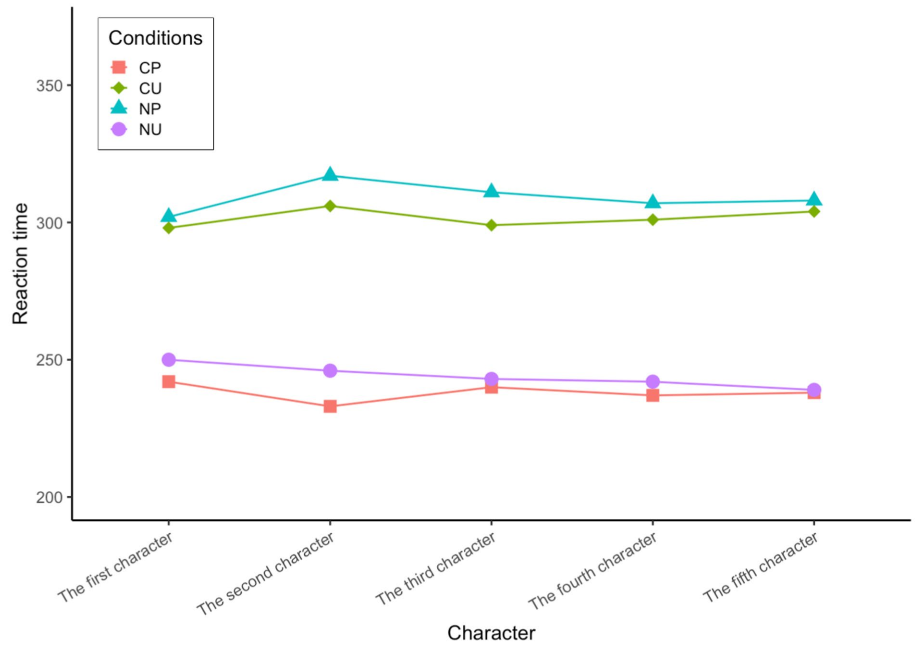


FIGURE 2
Reaction time of Chinese native speakers in four conditions.

TABLE 2 The average reaction time (in milliseconds) and standard error of Chinese L1 speakers.

Conditions	The first character	The second character	The third character	The fourth character	The fifth character
CP	242 (132.70)	233 (122.19)	240 (135.01)	237 (128.37)	238 (134.01)
NP	302 (71.64)	317 (94.64)	311 (71.34)	307 (71.09)	308 (70.94)
CU	298 (63.71)	306 (87.54)	299 (68.21)	301 (66.56)	304 (70.22)
NU	250 (147.87)	246 (140.97)	243 (133.88)	242 (133.99)	239 (133.12)

low-constraint context, indicating that there is a prediction error cost. Descriptive statistics are presented in Table 2.

3. Experiment 2

3.1. Method

3.1.1. Participants

A total of 19 Chinese L2 learners (9 males), aged 18–34 ($M = 23.15$, $SD = 5.39$), were recruited from the fourth grade (of five grades) of the Chinese summer program of a university in the United States. All of them were English native speakers. The participants had been learning Chinese for more than 3 years and can be assessed as advanced-level learners. They were able to read the experimental material with no difficulty, for two reasons: 1. The experimental materials were selected from Chinese teaching textbooks at junior and intermediate level; 2. We asked two Chinese L2 learners from the same group with participants to finish a pre-test to ensure that no unfamiliar words appeared.

3.1.2. Design

The experimental design was identical to that in Experiment 1.

3.1.3. Materials

The experimental materials were identical to those in Experiment 1.

3.1.4. Procedure

The experimental procedure was identical to that in Experiment 1.

3.2. Results

The correct rate of the questions was above 85%, which proved that the participants read the sentences carefully. We recorded the reaction time of the target word, which contained two characters, and the three characters following it. We deleted data according to the following criteria: (1) Data whose reaction time was shorter than 80 ms or greater than 1,500 ms; (3) Data beyond ±3 standard deviations. A total of 226 data points were deleted, accounting for 7.4% of the total data.

We constructed linear mixed models of the self-paced reading data, using the lme4 package (Bates et al., 2015) in R version 4.2.2 (R Core Team, 2022). We mainly examine the responses of five regions: the first character of the target word, the second character of the target word, and the third, fourth and fifth characters after that.

There was a significant difference in the reaction time from the first character to the fifth character between the conditions of CP and NP, first character: $t(392) = -6.440$, $p < 0.01$; second character: $t(606) = -6.388$, $p < 0.01$; third character: $t(581) = -6.524$, $p < 0.01$; fourth character: $t(572) = -5.933$, $p < 0.01$; fifth character: $t(584) = -6.612$, $p < 0.01$, and CP < NP (see Figure 3). This shows that the processing time of predictable words in high-constraint contexts is significantly shorter than that in low-constraint contexts. There was also a significant difference in the reaction time from the first to the fifth character between the conditions of CP and CU, the first character: $t(389) = -6.152$, $p < 0.01$; the second character: $t(588) = -7.539$, $p < 0.01$; the third character: $t(587) = -4.676$, $p < 0.01$; the fourth character: $t(600) = -5.383$, $p < 0.01$; the fifth character: $t(580) = -6.311$, $p < 0.01$, and CP < CU (see Figure 3). This shows that in high-constraint contexts, the processing time of predictable words is significantly shorter than that of unpredictable words. The above two points confirm the existence of a context predictability effect in Chinese L2 learners' reading processing.

There was also a significant difference in the reaction time for the first, second, fourth and fifth characters between the conditions of CU and NU, the first character: $t(386) = 5.912$, $p < 0.01$; the second character: $t(608) = 22.712$, $p < 0.01$; the third character: $t(585) = 3.664$, $p < 0.01$; the fourth character: $t(596) = 5.902$, $p < 0.01$; the fifth character: $t(575) = 6.070$, $p < 0.01$, and CU > NU (see Figure 3). The processing time of unpredictable words in high-constraint context is significantly longer than that in low-constraint context, indicating that there is a prediction error cost. Descriptive statistics are as follows (see Table 3)

4. Comparison between two experimental results

We compared data from Chinese native speakers and Chinese L2 learners. There was a significant difference in the reaction time for all the conditions. For the first character, CP: $t(218) = 20.891$, $p < 0.01$; CU: $t(420) = 20.837$, $p < 0.01$; NP: $t(424) = 20.937$, $p < 0.01$; NU: $t(214) = 20.633$, $p < 0.01$. For the second character, CP: $t(2412) = 18.408$, $p < 0.01$; CU: $t(604) = 19.808$, $p < 0.01$; NP: $t(624) = 17.672$, $p < 0.01$; NU: $t(2411) = 19.819$, $p < 0.01$. For the third character, CP: $t(2408) = 23.043$, $p < 0.01$; CU: $t(602) = 20.642$, $p < 0.01$; NP: $t(594) = 21.790$, $p < 0.01$; NU: $t(2404) = 24.046$, $p < 0.01$. For the fourth character, CP: $t(2413) = 23.393$, $p < 0.01$; CU: $t(608) = 21.214$, $p < 0.01$; NP: $t(581) = 21.528$, $p < 0.01$; NU: $t(2402) = 22.597$, $p < 0.01$. This

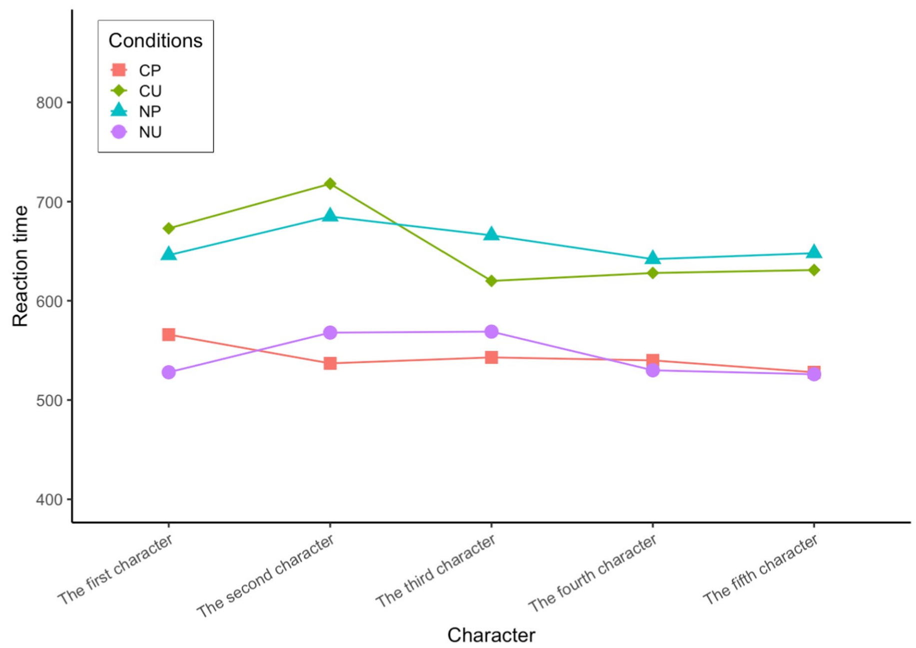


FIGURE 3
Reaction time of Chinese L2 Learners in four conditions.

TABLE 3 The average reaction time (in milliseconds) and standard error of Chinese L2 learners.

Conditions	The first character	The second character	The third character	The fourth character	The fifth character
CP	566 (195.88)	537 (205.49)	543 (210.18)	540 (203.24)	528 (196.48)
NP	646 (182.27)	685 (174.89)	666 (153.96)	642 (167.26)	648 (183.65)
CU	673 (199.38)	718 (198.23)	620 (158.43)	628 (157.89)	631 (148.79)
NU	528 (183.28)	568 (199.04)	569 (208.89)	530 (201.54)	526 (189.09)

indicates that advanced Chinese L2 learners were significantly slower than Chinese native speakers when they did predictive processing. In a word, even though advanced Chinese L2 learners can establish the same prediction mechanism as Chinese native speakers, there is still a big gap in terms of speed of predictive reading between them.

5. Discussion

In response to the questions raised above, our results are: There is a context predictability effect in the reading processing of Chinese L1, which is consistent with the current academic consensus. There is a prediction error cost in reading processing for Chinese L1 speakers. The ERP result of Federmeier et al. (2007), DeLong et al. (2011, 2012), Chou et al. (2014) was proved by our behavioral experiments. There is a context predictability effect in the reading processing of advanced Chinese L2 learners. This is consistent with the findings of Clahsen and Felser (2006), Chambers and Cooke (2009), and Dussias et al. (2013). There is a prediction error cost in reading processing for advanced Chinese L2 learners. The ERP result of Zirnstein et al. (2018) is supported by our behavioral experiments. These results suggest that Chinese L2 learners can establish the same prediction mechanism as Chinese native speakers in reading processing.

The brain seems to cope with the speed and complexity of language processing by "thinking ahead," that is, generating information about what might come and preparing to process it in advance at multiple levels, and this predictive processing brings benefit; at the same time, it will also bring cost (Federmeier, 2007). The benefit is manifested in the facilitation of the context predictability effect in L1 and L2 processing; that is, the processing time of predictable words is shorter than that of unpredictable words in high-constraint contexts (CP < CU) and that of predictable words in a high-constraint context is shorter than that in a low-constraint context (CP < NP); the cost is manifested in that the brain takes longer to process unpredictable words, because early activated target words are replaced, thus causing processing interference (CU > NU).

We can provide explanations for the above conclusions from two perspectives. One is Lexical Prediction. This theory holds that in the high-constraint context, target word activation is an "all or nothing" serial approach (DeLong et al., 2014; Kwon et al., 2017), so that when what is activated is not the target word, the brain will enter a "nothing" mode, that is, it needs to use more cognitive space to deal with emergencies, resulting in a prediction error cost.

The second explanation is Frontal LP (Late Positivity) Effects. Federmeier et al. (2007) observed the ERP data of the participants in the two intervals of 300–500 ms and 500–900 ms when they processed the unpredictable word in strongly constraining context and found the increased Frontal Positivity from 500 to 900 ms post-stimulus-onset, namely frontal LP effects. This discovery is very critical. On the one hand, it provides a possible explanation for the existence of prediction error cost, and on the other hand, it provides a new methodology for the research on contextual constraints. Delong et al. (2011) proved the existence of frontal LP effects, and further proposed that this effect is likely to occurs in parallel with the N400 of semantic activation and integration.

The reason why it is called processing "cost" is that when encountering an unpredictable component, the brain may need to invest more "extra" processing to override, revise, inhibit, or reanalyze the accruing contextual representation (Kutas et al., 2011; DeLong et al., 2012; Chou et al., 2014). For learners, in the short term, it may be "cost," in the long run, it seems that readers will benefit from such a process (Kutas et al., 2011).

In addition, cognitive control ability also plays an important role in reading processing, especially inhibitory control. When readers encounter processing difficulties, meaning that their predictions are denied, they can play a mediating role. The cost of L2 reading is higher for L2 learners if they have poorer inhibitory control, and the same pattern of effect was found for L1 speakers (Zirnstein et al., 2018). Therefore, in second language teaching and learning, consciously training learners' cognitive control, especially inhibitory control ability, is an effective way to improve reading ability.

6. Conclusion

This paper proves the existence of context predictability effect and prediction error cost in the reading processing of Chinese native speakers and advanced Chinese L2 learners. At an advanced level, Chinese L2 learners can build the same prediction mechanism as Chinese native speakers, but the learners are significantly slower than the native speakers in predictive riding processing. Teachers can take advantage of this to train Chinese L2 learners to consciously use predictive strategies, thereby improving reading ability and speed. The experimental results can be explained by Lexical Prediction and Frontal LP Effects.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LF is responsible for the specific content of the paper, including the writing of the literature review, the implementation of experiments, and data statistics, etc. NJ guides the paper in terms of theory and methods. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bai, X. J., Cao, Y. X., Gu, J. J., Guo, Z. Y., and Y, G. L. (2011). Effect of p predictability and space on Chinese reading: an eye movement study. *J. Psychol. Sci.* 34, 1282–1288. doi: 10.16719/j.cnki.1671-6981.2011.06.021

Bai, X. J., Liu, L. P., and Yan, G. L. (2008). An eye movement study on word-skipping in sentence Reading. *J. Psychol. Sci.* 31, 1045–1048. doi: 10.16719/j.cnki.1671-6981.2008.05.005

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H., Singmann, H., et al. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1, 2014.

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

Chambers, C. G., and Cooke, H. (2009). Lexical competition during second-language listening: sentence context, but not proficiency, constrains interference from the native lexicon. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1029–1040. doi: 10.1037/a0015901

Chou, C. J., Huang, H. W., Lee, C. L., and Lee, C. Y. (2014). Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *J. Neurolinguistics* 31, 42–54. doi: 10.1016/j.jneuroling.2014.06.003

Chow, W. Y., Lau, E., Wang, S., and Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Lang. Cogn. Neurosci.* 33, 803–828. doi: 10.1080/23273798.2018.1427878

Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/S0142716406060024

Coulson, S., and Van Petten, C. (2002). Conceptual integration and metaphor: an event-related potential study. *Mem. Cogn.* 30, 958–968. doi: 10.3758/BF03195780

DeLong, K. A., Groppe, D. M., Urbach, T. P., and Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain Lang.* 121, 226–239. doi: 10.1016/j.bandl.2012.02.006

DeLong, K. A., Troyer, M., and Kutas, M. (2014). Pre-processing in sentence comprehension: sensitivity to likely upcoming meaning and structure. *Lang. Linguis. Compass* 8, 631–645. doi: 10.1111/lnc3.12093

Delong, K. A., Urbach, T. P., Groppe, D. M., and Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology* 48, 1203–1207. doi: 10.1111/j.1469-8986.2011.01199.x

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504

DeLong, K. A., Urbach, T. P., and Kutas, M. (2007). A cost to mispredicting: effects of sentential constraint violations. Poster presented at the 20th Annual CUNY Conference on Human Sentence Processing La Jolla, CA.

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., and Gerfen, C. (2013). When gender and looking go hand in hand: grammatical gender processing in L2 Spanish. *Stud. Second. Lang. Acquis.* 35, 353–387. doi: 10.1017/S0272263112000915

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., and Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84. doi: 10.1016/j.brainres.2006.06.101

Frisson, S., Harvey, D. R., and Staub, A. (2017). No prediction error cost in reading: evidence from eye movements. *J. Mem. Lang.* 95, 200–214. doi: 10.1016/j.jml.2017.04.007

Grüter, T., and Rohde, H. (2013). L2 processing is affected by RAGE: evidence from reference resolution. in 12th conference on Generative Approaches to Second Language Acquisition (GASLA).

Guo, X. F. (2012). How the word frequency-predictability affect the fixation duration and land position in Chinese reading. Master's thesis, Tianjin: Tianjin Normal University.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: what is different? *Linguis. Approach. Bilingualism* 4, 257–282. doi: 10.1075/lab.4.2.05kaa

Kaan, E., Dallas, A. C., and Wijnen, F. (2010). "Syntactic predictions in second-language sentence processing," in *Structure preserved. Festschrift in the honor of Jan Koster*. eds. J.-W. Zwart and M. de Vries (Amsterdam: John Benjamins), 207–213.

Kutas, M., DeLong, K. A., and Smith, N. J. (2011). *A look around at what lies ahead: Prediction and predictability in language processing*. Oxford: Oxford Academic press.

Kwon, N., Sturt, P., and Liu, P. (2017). Predicting semantic features in Chinese: evidence from ERPs. *Cognition* 166, 433–446. doi: 10.1016/j.cognition.2017.06.010

Li, Y. (2016). A study on word skipping during Chinese sentence Reading: Effects of word length and predictability. Master's thesis, Changsha: Hunan Normal University.

Li, L., Zhao, S. N., Zhang, L. J., and Wang, J. X. (2022). Understanding mechanisms of prediction error cost in Chinese reading for older adults. *Adv. Psychol. Sci.* 30, 1–14. doi: 10.3724/SP.J.1042.2022.00001

Liu, Z. F., Tong, W., Zhang, Z. J., and Zhao, Y. J. (2020). Predictability impacts word and character processing in Chinese reading: evidence from eye movements. *Acta Psychol. Sin.* 52, 1031–1047. doi: 10.3724/SP.J.1041.2020.01031

Liu, N. N., Wang, X., Liu, Z. F., Wang, Y. S., and Yan, G. L. (2019). The effect of contextual predictability on Parafoveal process for highly- and low-skilled developing readers during Chinese Reading. *J. Psychol. Sci.* 4, 848–853. doi: 10.16719/j.cnki.1671-6981.20190412

Luke, S. G., and Christianson, K. (2016). Limits on lexical prediction during reading. *Cogn. Psychol.* 88, 22–60. doi: 10.1016/j.cogpsych.2016.06.002

Mitchell, D. C., and Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Q. J. Exp. Psychol.* 30, 609–636. doi: 10.1080/14640747808400689

Moreno, E. M., Federmeier, K. D., and Kutas, M. (2002). Switching languages, switching palabras (words): an electrophysiological study of code switching. *Brain Lang.* 80, 188–207. doi: 10.1006/brln.2001.2588

Rayner, K., Ashby, J., Pollatsek, A., and Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: implications for the EZ reader model. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 720–732. doi: 10.1037/0096-1523.30.4.720

Rayner, K., Li, X., Juhasz, B. J., and Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychon. Bull. Rev.* 12, 1089–1093. doi: 10.3758/BF03206448

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., and Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychol. Aging* 21, 448–465. doi: 10.1037/0882-7974.21.3.448

Rayner, K., and Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: a further examination. *Psychon. Bull. Rev.* 3, 504–509. doi: 10.3758/BF03214555

Roland, D., Yun, H., Koenig, J. P., and Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition* 122, 267–279. doi: 10.1016/j.cognition.2011.11.011

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: critical review and theoretical interpretation. *Lang. Linguis. Compass* 9, 311–327. doi: 10.1111/lnc3.12151

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467. doi: 10.1037/0278-7393.31.3.443

Wang, X. (2017). *The contextual facilitation effect for developing readers' lexical processing in natural Reading: The evidence from eye movements, Master's thesis*, Tianjin: Tianjin Normal University.

Wang, L. H., Bai, X. J., Yan, G. L., and Wu, J. (2012). The role of word frequency and word predictability in reading of elderly people. *Chin. J. Gerontol.* 16, 3503–3507. doi: 10.3969/j.issn.1005-9202.2012.16.073

Wicha, N. Y., Moreno, E. M., and Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cogn. Neurosci.* 16, 1272–1288. doi: 10.1162/0898929041920487

Zhang, M. M. (2015). The mechanism of word skipping in Chinese Reading: An eye movement study. Doctoral dissertation, Tianjin: Tianjin Normal University.

Zhang, W., Chow, W. Y., Liang, B., and Wang, S. (2019). Robust effects of predictability across experimental contexts: evidence from event-related potentials. *Neuropsychologia* 134:107229. doi: 10.1016/j.neuropsychologia.2019.107229

Zhang, W., Dong, J., Duan, X., Zhang, Y., Gao, X., Zhen, A., et al. (2023). Prediction of semantic features is modulated by global prediction reliability: evidence from the N400 effect. *J. Neurolinguistics* 65:101109. doi: 10.1016/j.jneuroling.2022.101109

Zhao, S. N. (2009). Effect of word predictability on lexical processing in young and older Chinese readers: An eye movement study, Master's thesis, Tianjin: Tianjin Normal University.

Zhao, S. W. (2020). Interaction effects of contextual predictiveness and parafoveal anticipation in children's sentence reading, Master's thesis, Tianjin: Tianjin Normal University.

Zhao, S. N. (2022). Mechanism of the predictive effect of context in sentence reading, Doctoral dissertation, Tianjin: Tianjin Normal University.

Zhao, S. N., Li, L., Zhang, L. J., and Wang, J. X. (2021). The influence of constrictive context on the processing of unpredictable word in Chinese Reading. *Stud. Psychol. Behav.* 19, 736–742.

Zhu, X. P. (1991). The effect of Chinese sentence context on word recognition. *Acta Psychol. Sin.* 23, 35–42.

Zirnstein, M., van Hell, J. G., and Kroll, J. F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition* 176, 87–106. doi: 10.1016/j.cognition.2018.03.001

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 04 April 2023
DOI 10.3389/fpsyg.2023.1131913

OPEN ACCESS

EDITED BY
Linjun Zhang,
Peking University,
China

REVIEWED BY
Yu Li,
United International College,
China
Yufang Ruan,
United Nations Educational,
Scientific and Cultural Organization,
France

*CORRESPONDENCE
Ying Xiong
✉ xiongying22@sjtu.edu.cn

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 26 December 2022
ACCEPTED 06 March 2023
PUBLISHED 04 April 2023

CITATION
Yang L, Xiong Y and Chen Q (2023) The role of linguistic and cognitive skills in reading Chinese as a second language: A path analysis modeling approach.
Front. Psychol. 14:1131913.
doi: 10.3389/fpsyg.2023.1131913

The role of linguistic and cognitive skills in reading Chinese as a second language: A path analysis modeling approach

Lei Yang[1], Ying Xiong[2]* and Qi Chen[1]

[1]Department of Foreign Language, Xi'an Jiao Tong University, Xi'an, China, [2]School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China

This study examined the role of basic linguistic skills (vocabulary, syntax, orthography, and morphological awareness), basic cognitive skills (working memory), and higher-order cognitive skills (inference making and reading monitoring) in reading Chinese as a second language (L2). A total of 252 international students from Pakistan, Indonesia, Malaysia, and Laos were recruited, and a range of measures including a Chinese reading comprehension test (HSK level 3), four linguistic knowledge tests on Chinese lexical, syntactic, and orthographic knowledge as well as morphological awareness, a reading span test, an inference making task, and an inconsistency detection test. The results of hierarchical multiple regressions showed that the measured linguistic skills and cognitive skills explained 80% of the variances in L2 Chinese reading, among which morphological awareness made the largest contribution. The path analysis revealed that linguistic skills and working memory contributed indirectly to reading comprehension *via* inference making and comprehension monitoring, while the two higher-order cognitive skills made direct contributions. Overall, this study demonstrates that inference making and comprehension monitoring contributed directly to reading comprehension, while linguistic skills and working memory functioned indirectly *via* the higher-order cognitive skills It also highlights the importance of morphological awareness in a hierarchical model of L2 Chinese reading.

KEYWORDS

L2 Chinese reading, linguistic skills, cognitive skills, morphological awareness, path analysis

1. Introduction

Reading is a fundamental part of second language (L2) learning. International students learning in China also take Chinese reading as a way of cultural exploration. However, the remarkable differences between the Chinese language, ideography, and their native language may result in great difficulties in comprehension. In addition, complex cognitive processing underpins reading comprehension. It requires readers to understand not only the meanings of individual words but also ideas across sentences and larger stretches of texts. What factors may influence Chinese reading comprehension among international students in China? This study intends to explore a range of factors involved in Chinese reading comprehension and reveal the complex relations between language and cognitive skills. The findings of this study may provide insights into Chinese teaching and learning.

2. Literature review

2.1. Theories of reading comprehension

Studies of reading have been conducted at both word and sentence levels. At the word level, reading concerns how readers automatically map printed forms of words to the semantic meaning of words and initiate orthographic decoding, phonological encoding, and semantic integration (Frost et al., 2008). At the sentence level, the crux of the matter is how words are related to their mental representations. According to the reader's situation model (Van Dijk and Kintsch, 1983), readers need to integrate related words into a comprehensible conceptual structure to achieve sentence level comprehension as their goal is beyond the mapping between forms and meanings. Therefore, readers need to construct a series of closely related conceptual units and establish a hierarchy of themes as their mental representations rather than a loose set of concepts (Fesel et al., 2015; Kim and Clariana, 2015, 2017).

Taking reading as the construction of situation models rather than form-to-meaning mapping does not diminish the importance of word knowledge in reading. On the contrary, it is vital for readers to integrate word meanings into relevant conceptual units. Reading System Framework (RSF, Perfetti and Stafura, 2014) encompasses both linguistic knowledge and cognitive skills, providing a framework that explains word-to-text comprehension processes. In RSF theory, three key hypotheses are proposed:

1. Three types of knowledge sources are employed in reading ranging from basic linguistic knowledge to general knowledge.
2. The processes of reading include decoding, word identification, meaning retrieval, constituent building (sentence parsing), inferencing, and comprehension monitoring.
3. These cognitive processes occur within an interactive processing system that has limited resources in terms of attention, memory, and control.

According to RSF theory, reading comprehension relies on basic cognitive function, fundamental linguistic knowledge, and higher-order cognitive functions. There is a hierarchical relationship between these language and cognitive skills. In terms of hierarchical relations, inference making is considered as a higher-order cognitive skill directly related to reading comprehension, as higher-order cognitive skills are capable of integrating multiple information across discourse to form a coherent whole (Strasser and Del Río, 2013; Kim, 2016). Comprehension monitoring is another higher-order cognitive skill that involves reflecting on and regulating an individual's understanding of the incoming information. Readers use it to evaluate whether their understanding of the newly received information is coherent with those from the previous portion of the discourse and to resolve any discrepancies. In addition to the skills mentioned above, working memory, referring to the basic cognitive resource for processing general information, was also considered an important contributor to reading comprehension (Kim, 2016). Working memory defined as the capacity to store and manipulates information is the key to how efficiently a reader can hold, process, and recall information in mind (Seigneuric et al., 2000).

Reading comprehension is supported by cognitive skills, which, in turn, are supported by foundational language skills. Foundational language skills construct initial and literal propositions based on the words and phrases of the discourse. Foundational language skills are necessary for the construction of initial propositions by assigning meaning to individual words and sentences. But in many cases, the initial propositions are not complete and sometimes contradictory. Higher-order cognitive skills are necessary to integrate information across sentences, detect contradictions, and construct coherent mental representations (Kim, 2016, 2017; Oakhill, 2020).

In sum, reading comprehension is conceptualized as having three levels: the basic cognitive skill, foundational language skills, and higher-order cognitive skills. These language and cognitive skills have hierarchical relationships, which can be characterized as having direct and/or indirect routes towards reading comprehension. According to the RSF theory, cognitive skills (i.e., inference making and comprehension monitoring) are higher-order skills, directly related to reading comprehension. As foundational language skills (e.g., syntactic, lexical, morphological, and orthographic knowledge) are necessary to attach meaning to individual words and sentences, they should be placed at a lower level, contributing directly or indirectly to reading comprehension. The basic cognitive skill (i.e., working memory) that manipulates and stores the information in mind is placed at the same level as the foundational language skills.

2.2. Factors in reading comprehension

2.2.1. The impact of foundational linguistic skills

Successful reading comprehension includes two phases: a foundation phase and a construction phase. Linguistic knowledge as a foundational skill for reading encompasses vocabulary, morphological awareness, syntactic knowledge, and orthographic knowledge to facilitate further information integration. In this section, we will discuss these different types of linguistic knowledge that are essential to reading comprehension.

Developing a strong vocabulary is at the core of reading comprehension. In order to make propositions, readers must be able to attach meaning to most words in a text, assisting the integration and construction of text meaning. Vocabulary is strongly correlated with reading comprehension (Cain et al., 2004; Seigneuric and Ehrlich, 2005). In previous studies, vocabulary was a strong predictor of reading comprehension among native Chinese speakers (Leong and Ho, 2008) as well as L2 Chinese learners (Zhang and Koda, 2018; Zhang et al., 2021).

Despite that vocabulary is foundational, in the absence of syntactic knowledge, the text cannot be understood. Syntactic knowledge provides readers with necessary rules to understand sentence structures, identify syntactic categories of words, and process text information at discourse level. A good understanding of syntactic structures facilitates the recognition of word meaning as well as sentence and discourse comprehension. Syntactic knowledge is crucial for Chinese reading due to the unique features of Chinese such as the absence of an inflectional system, more flexible word order, and more extensive use of connectives (Li and Thompson, 1981). Chinese readers rely on their syntactical knowledge to seek information about tense, number, and their semantic relations (Lin, 2006). Besides, word order is more flexible in Chinese. There are two main types of structure in Chinese: subject-verb and topic-prominent. The basic word order in Chinese is SVO. For example, in 我爱吃面条 (I love to eat noodles), 我 (I) is the subject, 爱吃 (love to eat) are two verbs serving as the predicate, and 面条(noodles) is a noun serving as the object. In the above sentence, the object 面条(noodles) can be moved to the beginning of the sentence to form a topic-comment sentence (面条，我爱吃), which means "As for noodle, I love to eat (it)." The topic can be extended to several subsequent sentences once it has been established (Chik et al., 2012). Therefore, Chinese readers often come across sentences without subjects. Another important characteristic of Chinese syntax is the extensive use of connectives indicating cause, time, and contrast. It is thus important to keep track of the logic and semantic relations between words and phrases within a sentence and also across sentences with the help of syntactic knowledge of word order and connectives. Therefore, Chinese readers rely on their syntactic knowledge to accomplish reading comprehension and information processing. Syntactic knowledge was found to be an important contributor to the comprehension of sentences and texts (Chik et al., 2012; Tong et al., 2014; Lo et al., 2016).

However, whether syntactic knowledge may exert an impact on Chinese reading *via* cognitive skills has not been examined in depth. Hung's (2021) recent study represents such an attempt to find the intertwined relationships between syntactic skills and cognitive skills. She found an indirect effect of executive function (as measured by working memory tasks and inhibition tasks) on Chinese Grade 2 and 3 students' reading comprehension through syntactic awareness (as measured by a word order correction task). However, no indirect effects of executive function were found on reading through inference-making, refuting her original hypothesis that executive function can support inference making to achieve passage-level comprehension. The question that remains to be answered is whether the intertwined relationships between language and cognitive skills remain the same among L2 Chinese readers.

Orthography also plays an important role in Chinese reading. Chinese has a morphosyllabic orthography that consists of characters, radicals, and strokes, and each character represents both a morpheme and a syllable (Shu et al., 2000). A character is composed of radicals, which in turn is composed of strokes. In Chinese characters, about 80% or so are phonetic-semantic (Shu et al., 2003), containing both semantic and pronunciation information. Semantic radicals reveal a character's meaning, while phonetic radicals reveal a character's sound. Chinese radicals at particular positions (top, bottom, left or right) determine or pertain to the sound or meaning of the character. The orthographic knowledge of Chinese can greatly promote beginner-level L2 readers of Chinese since a significant part of modern Chinese characters is regarded as semantic–phonetic compounds that contain both semantic and phonetic components (Shu et al., 2003). In Chinese, faster mapping occurs from orthographic information to semantic information than from orthographic information to phonological information (Yang et al., 2006). Empirical evidence has shown that for L1 Chinese children, Chinese reading comprehension was greatly influenced by their orthographic knowledge (Cheung et al., 2007; Yeung et al., 2016). A further question to be resolved is whether this is also the case for L2 Chinese readers.

Apart from the factors mentioned above, morphological awareness also plays a significant role in reading comprehension because it can help readers generate important clues about meaning in alphabetic languages like English and Spanish (Carlisle, 2000) as well as in the non-alphabetic language like Chinese (Packard, 2000). Morphological awareness is defined as the ability to reflect and draw on the componential units of words that carry meanings such as derivational and inflectional affixes in English (Carlisle, 2000; Kieffer et al., 2013). It has been well-established that morphological awareness is a crucial skill for achieving high-quality reading comprehension in alphabetic language such as English (Carlisle, 2000; Nagy et al., 2006; Tong et al., 2011; Kieffer et al., 2013). The samples of these studies mainly constituted English-speaking children from elementary grades. It has been noticed that there was a lack of variance in the certain aspect of morphological awareness (e.g., inflectional knowledge) among English-speaking children from more advanced grades because the foundational linguistic skill such as morphological awareness became well-developed as children grew up (Kuo and Anderson, 2006). It has also been found that morphological awareness can contribute to reading comprehension both directly and indirectly *via* a range of other foundational language skills such as lexical skill (Kieffer et al., 2013) and phonological skill (Nagy et al., 2006). However, less is known about how morphological awareness interacts with cognitive skills.

A major difference between Chinese and alphabetic languages is that in Chinese, there is a lack of correspondence between the orthographic units and phonemes as seen in alphabetic languages. The orthographic unit needs to be recognized as a whole before activating the phonology (Florit and Cain, 2011). As an analytic language, 75% of Chinese words contain two or three morphemes that rely heavily on morphological knowledge to access semantic clues and make inference about words' meaning (Chen et al., 2009). For example, 上下车 (get on the car/get off the car) in Chinese is formed by three morphemes or characters, 上 (get on), 下 (get off), and 车(car). Readers can make inferences about word meaning based on the three morphemes. Therefore, morphological knowledge is essential for Chinese reading comprehension.

Most studies on the role of morphological awareness in Chinese reading focus on native Chinese speakers. It was found that morphological awareness directly contributed to reading comprehension and indirectly predicted reading comprehension through word reading for Chinese kindergarteners (Pan and Lin, 2023), Grade 1 to 2 Chinese pupils (Cheng et al., 2017), Grades 3 and 4 children (Qiao et al., 2021). However, Zhang's (2014) study on Chinese second graders found that the effect of morphological awareness on reading comprehension was mediated by lexical inference ability and vocabulary knowledge and that morphological awareness had no significant contribution to reading comprehension after controlling for the two mediating factors. It should be noticed that morphological awareness was measured differently in these studies. Zhang (2014) adopted a morpheme recognition and a morpheme discrimination task, while others employed either a compound production task (Cheng et al., 2017) or a morphological construction task (Qiao et al., 2021; Pan and Lin, 2023).

Only a small number of recent studies have examined morphological awareness in L2 Chinese reading comprehension. A recent study on L1 English learners of L2 Chinese showed that morphological awareness (as measured by morpheme segmentation and morpheme discrimination) directly contributed to reading comprehension, while lexical inference directly contributed to reading and mediated the effect of morphological awareness on reading comprehension (Zhang et al., 2021). Another study on Chinese as a heritage language adult learners found that morphological awareness (as measured by structural awareness and functional awareness) directly contributed to reading comprehension (in the form of lexical inference and passage comprehension) and mediated the effect of vocabulary knowledge on reading comprehension (Zhang and Koda, 2018).

These studies showed that although morphological awareness is essential in explaining Chinese reading performance, the route it takes to affect comprehension varies when different skills are added to the model. Until recently, few studies integrate a full range of cognitive skills such as monitoring and working memory into a model of Chinese reading. We now turn to see why these cognitive skills are essential in understanding Chinese reading comprehension.

2.2.2. Basic cognitive skill: Working memory

Working memory, as the basic cognitive function, is related to a set of processes occurring during reading comprehension, such as inhibition, deployment of attention, information update, and so on. It has a significant impact on the development of reading comprehension capacity (Perfetti and Stafura, 2014). Working memory is more than a passive repository otherwise it would be reduced to a mere container of the incoming information. It is a cognitive system serving as a threshold that can control and manipulate the information going into the permanent memory. To be more precise, working memory is a cognitive system limited in capacity, which can affect higher-order cognitive activities such as reading comprehension, language production, inference making, and problems solving (Oberauer et al., 2003).

Working memory not only integrates information but also inhibits irrelevant information in order to release the limited processing capacity that can be used for drawing complex inferences. According to the Structure Building Framework (Gernsbacher, 1990), the goal of comprehension is the building of a cohesive mental representation, or structure, which relies on three processes, laying a foundation, mapping, and shifting. When there is an inconsistency in the information flow, working memory will suppress the irrelevant information and shift it to another subset. If there the information is consistent, working memory will be reinforced.

Working memory inhibitory mechanisms consist of two important skills: blocking information flow and deleting irrelevant information (Hasher and Zacks, 1988). These skills ensure that irrelevant information is deleted and unable to enter working memory so that only relevant information gets processed. Readers with stronger working memory can better suppress the information (Yeari and van den Broek, 2011). The stronger the suppression is, the more efficient working memory becomes, and the better reading comprehension is achieved. The inhibitory mechanism can promote working memory and contribute to the construction of a coherent mental representation (Perfetti and Stafura, 2014). Chinese reading comprehension consumes more cognitive resources than alphabetical reading (Diamond, 2013). Recent studies on Chinese reading have shown that working memory supports reading comprehension (Weng et al., 2016; Liu and Wu, 2022; Yeung, 2022). Readers with poor working memory are not likely to perform well in reading. However, it is still unclear how working memory would influence L2 Chinese reading in a hierarchical model of comprehension.

2.2.3. Higher-order cognitive skills

Inference making and comprehension monitoring were the two higher-order cognitive skills investigated in this study. Inference making is crucial for proficient reading and pertains to higher-order cognitive skills (Singer et al., 1994). To construct mental representations of meaningful events in a text, readers need to establish relations and draw inferences based on the context and background (Currie and Cain, 2015). Readers make inferences about local and global coherence to fill in details not explicitly stated in the text (Graesser et al., 1994). As a result, inference making skill ensures that information in the text is integrated with information from other sentences and world knowledge, filling in the information gaps and promoting local and global coherence. Readers' inference making skills predict their reading comprehension in both English (Cain et al., 2001) and Chinese (Zhao et al., 2022). However, less is known about the role that inference making plays in a hierarchical model of the linguistic-cognitive skills for L2 Chinese reading comprehension.

Comprehension monitoring skills will be employed during the unfolding of text and the integration of new information (Zhao et al., 2022). Readers with strong reading comprehension skills are more likely to identify contradictory information and monitor their comprehension more effectively and incorporate it into a situation model (Cain, 2009). As readers monitor their understanding of words and sentences, they evaluate their propositions and integrate their knowledge of word usage and word knowledge (Kinnunen et al., 1998). Readers with strong comprehension skills are more likely to engage in strategic processing, discover connections between text elements, and draw on prior knowledge to promote understanding. In contrast, readers with weaker comprehension skills are less likely to achieve consistent monitoring of their comprehension and often fail to suppress or repair inconsistencies within a text (Kinnunen et al., 1998). A study on English-speaking children has shown that comprehension monitoring is an important variable contributing to reading comprehension after controlling other aspects (decoding, vocabulary, and working memory) of language comprehension (Yeomans-Maldonado, 2017). Zhao et al. (2022) found direct relation of comprehension monitoring to Chinese reading comprehension. The case of L2 Chinese readers is less discussed.

2.2.4. Higher-order skills as mediators

According to RSF theory, the higher-order cognitive skills (e.g., monitoring and inference making) might mediate the effects of the foundational language skills on reading comprehension. Empirical evidence has shown, for example, that the skills for generating different types of inferences are heavily dependent on different types of vocabulary knowledge (Oakhill et al., 2015). Another study on English-speaking children found that the effect of vocabulary knowledge on reading comprehension was mediated by inference making skill (Silva and Cain, 2015).

Similar to the results of inference making studies, there has been evidence showing that foundational language skills underpin the comprehension monitoring skill (Kim, 2015; Zargar et al., 2020). Zargar et al.'s (2020) eye-tracking study on English-speaking children found that the gaze duration (an indicator of comprehension monitoring in the form of comprehension evaluation) was dependent on vocabulary knowledge. Kim's (2015) study on Korean-speaking children found that vocabulary and syntactic knowledge were directly related to listening comprehension and also indirectly predicted listening comprehension *via* comprehension monitoring. It also found that listening comprehension and word reading mediated the effects of all language skills and cognitive skills on reading comprehension.

The importance of syntactic knowledge for comprehension monitoring lies in the fact that it allows readers to better understand the grammatical relations between words and sentences and detect syntactic errors.

2.3. Purpose of this study

Despite the fact that previous studies have found evidence that supports the significant contributions of basic cognitive skills, foundational language skills and higher-order cognitive skills to Chinese reading comprehension, the hierarchical relationships among these skills have not been examined in a single synthesized model. Several previous studies have started to construct such an integrated model of reading comprehension. For example, Kim's (2017) model of native English children's reading comprehension covered working memory, vocabulary, grammatical knowledge, inference, comprehension monitoring, word reading, and listening comprehension. Pan and Lin's (2023) model of Chinese children's reading comprehension covered nonverbal IQ, phonological awareness, morphological awareness, orthographic knowledge, vocabulary knowledge, word reading, and listening comprehension. These two models both drew on the two important factors in the simple view of reading, word reading and listening comprehension (Hoover and Gough, 1990). However, these two skills can be further divided into componential skills. Treating them as a whole may blur the hierarchical relationships among the componential language and cognitive skills in reading. Zhao et al.'s (2022) model shared a similar attempt with the current study. She explored the direct and indirect effects of language skills (vocabulary, syntactic knowledge, and orthographic knowledge) and cognitive skills (inference making and comprehension monitoring) on Chinese Grade 3 pupils' reading comprehension. However, she did not include key factors such as morphological awareness and working memory. Also, previous studies are mainly interested in native speakers of English or Chinese, few studies provide insights into the L2 Chinese learners' reading performance.

To extend our understanding of the significance of the basic cognitive skill (working memory), foundational language skills (syntactic knowledge, lexical knowledge, orthographic knowledge, and morphological awareness), and higher-order cognitive skills (inference making and comprehension monitoring) in explaining Chinese reading comprehension, these skills were examined in one model in the current study. More importantly, the research also examined the direct and indirect contributions of each variable to gain the hierarchical relations among language skills, basic cognitive skill, and higher-order cognitive skills with respect to reading comprehension in L2 Chinese learners. Specifically, the study sought to answer two research questions:

1. How do these three facets (the basic cognitive skill, foundational language skills, higher-order cognitive skills) make their contributions to the L2 Chinese reading comprehension?
2. What are the direct and/or indirect routes each skill took to affect L2 Chinese reading comprehension?

Regarding the first research question, we hypothesized that Chinese reading comprehension would be influenced by language and cognitive skills, and each cognitive and language skill had a unique contribution to Chinese reading comprehension. Multiple regressions were used to determine each variable's contribution.

Regarding the second question, based on previous research, language skills were intertwined with cognitive skills in reading comprehension, for example, the association between syntactic knowledge and executive function (Hung, 2021), the relationship between vocabulary and inferencing (Oakhill et al., 2015; Silva and Cain, 2015), and the relationship between vocabulary and monitoring (Zargar et al., 2020). The higher-order cognitive skills may mediate the relationship between language skills and reading comprehension. Therefore, we used a path analysis method to determine how well our proposed model fits current data and examine the significance of each direct and indirect path in the model. We hypothesized that the basic cognitive skill and foundational language skills contribute to reading comprehension *via* higher-order cognitive skills which function as mediators.

3. Method

3.1. Participants

A total of 252 first-year college students ranging from 18 to 21 years old were enrolled from four different universities in China. One student dropped out, so 251 students took the tests (131 males, 120 females, $M_{age} = 19.73$ years, $SD = 1.28$). They were international students from Pakistan, Indonesia, Malaysia, and Laos. Their mother tongues belong to alphabetic languages different from Chinese. All of them passed the HSK2-level language exam before arriving in China.

All participants learned HSK3-level Chinese curriculum including reading as a core subject, not as a separate one. Chinese curriculums focus on the teaching and learning of texts from HSK3-level Standard Course. At the end of HSK3-level Chinese curriculum, participants have grasped 600 Chinese characters with basic proficiency in Chinese reading. According to their teachers, all participants exhibited no hearing impairments or language impairments. Participant permission was obtained for the study.

3.2. Measures

3.2.1. Working memory test

Working memory tests are a battery of tasks that measures people's capacity to temporarily store and process various types of information. These tasks can be categorized into simple/complex linguistic tasks and simple/complex non-linguistic tasks. Reading span tasks and operation span tasks are widely used in empirical studies. The present study adopted a linguistic working memory task. Adapted from Kim's (2016) study, a reading span task was carried out on E-Prime 2.0. It contained 30 experimental trials and 5 practice trials. In each trial, participants saw a sentence of 7 to 10 Chinese characters followed by a Chinese word in bold type (e.g., 小王，请帮我开门，谢谢。北京． *Xiao Wang, open the door for me please. Beijing.*). All the trial sentences were selected from HSK3-level syllabuses. The participants were tested in two dimensions. The first dimension was the capacity to store linguistic information. After each sentence was presented, participants were required to identify whether a new word on the screen matched the last word of the trial sentence. The second dimension was comprehension. A statement about the trial sentence was presented and the participants were asked to judge whether the statement was correct or not (yes/no response). The task consisted of a total of 15 items. Each item is given 2 marks. In order to get 2 marks, participants had to recall new words on the screen and respond to yes/no questions. The maximum score for the task was 30. Cronbach's alpha for this working memory test was $\alpha = 0.91$.

3.2.2. Chinese reading comprehension test

Reading comprehension capacity can be tested at sentence and discourse level (Tunmer and Chapman, 2012). This study administered a Chinese reading comprehension test adapted from the HSK3-level texts. There were 20 multiple choice questions for them to answer. Each item was given one mark. At sentence level, the participants were presented with 10 sentences, each with a blank. They were required to select appropriate answers from the choices provided. For example:

1. A：你有什么（ ）What () do you have?	A. 关心 care
B：我喜欢体育 I love doing sports.	B. 差 left
2. A：我的飞机票呢？怎么（ ）找不到了？	C. 爱好 hobby
Where is my plane ticket? How () I still can't find it?	D. 节目 TV programs
B：是不是和报纸放一起了？	E. 还 come

At discourse level, the participants were presented with 10 paragraphs, each with one or two questions. They were required to select appropriate answers from the choices provided. For example:

中国在道别的时候经常会说“再见”。“再见”是一个很有意思的词语。“再见”表示“再一次见面”，所以人们离开时说“再见”，其实也是希望以后再见面。

In China, we often say "goodbye" when we get parted from each other. "Goodbye" is a very interesting word. Its literal meaning is "see you again". Therefore, when people say "goodbye", they actually hope to see you again.

“再见”出现在什么时候？When does "goodbye" appear?

A 关灯 shut off the light B 见面 meet C 离开 leave

The maximum score for the task was 20. The reliability of the comprehension test was 0.81.

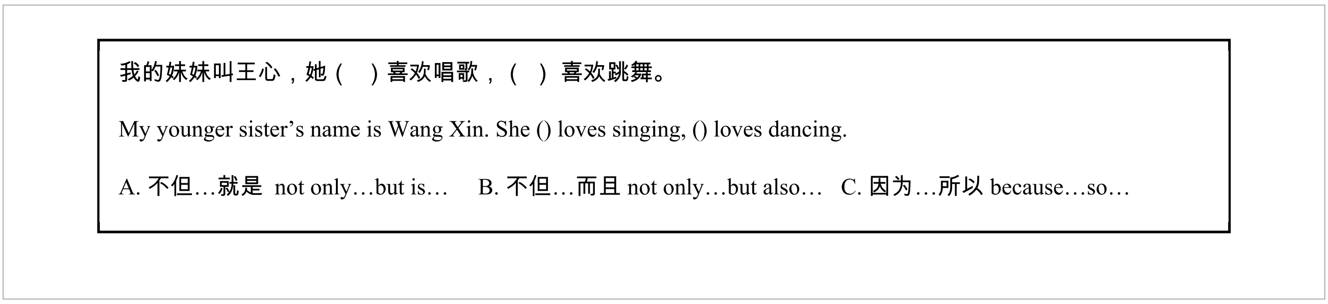
3.2.3. Linguistic knowledge tests

The linguistic knowledge of participants was measured in terms of four aspects: syntactic knowledge, lexical knowledge, orthographic knowledge, and morphological awareness. Totally 60 test items (15 items for each type of knowledge and each item one mark) were selected or adapted from HSK3-level examinations and the HSK3-level syllabus. The maximum score for the task was 60.

Syntactic knowledge was assessed by using a cloze task (Cronbach's alpha at 0.87). There were 15 items, and each item was given 1 mark. The total score was 15 marks. The highly frequent use of conjunctions is a prominent feature of Chinese syntax and the knowledge about conjunctions can be regarded as an indicator of syntactic knowledge (Li and Thompson, 1981). The participants were presented with sentences missing conjunction words and required to select appropriate answers.

Lexical knowledge was measured by two tasks. The first task was to choose synonyms for underlined words in a sentence or paragraph. The second task was to choose appropriate words to complete a sentence or a paragraph. Altogether, there were 15 items in the two tasks, with 7 items for choosing synonyms and 8 items for choosing appropriate words and each item was given 1 mark. The total score was 15 marks. The reliability of this vocabulary test was 0.89.

Orthographic knowledge was assessed using a Chinese character decision task containing 5 real Chinese characters, 5 non-characters, and 5 pseudo-characters. The design of the non-characters and pseudo-characters followed a fundamental principle that possible constituents (radicals and strokes) of a Chinese character should be placed in their legal positions (Liu et al., 2017). There were 15 items,



and each judgement was given 1 mark. The total score was 15 marks. The reliability of the orthographic knowledge test was 0.91.

The morphological awareness test was adapted from that of McBride-Chang et al. (2003). The morphological awareness in the study was composed of a test on homophones and a test on compounds. The homophone production test (Cronbach's alpha at 0.83) is intended to measure participants' knowledge of homophones. For example, participants were asked to produce a word (e.g., 时间 time) using a Chinese character时shí and then produce words (e.g., 石头stone、食堂canteen、实验experiment) using its homophones such as 石shí, 食shí, 实shí. Two solutions were required for each target character. Each correct solution was given a score of 0.5. The total number of test items was 10. The compounding semantic morpheme test (Cronbach's alpha at 0.79) required participants to produce a compound word based on a specific description of a concept. An example question would be: "If a type of oil is made of sesame, then it is called sesame oil (芝麻油). If a type of oil is made of peanuts, what should we call it?" The answer should be peanuts oil (花生油). Similarly, each correct answer was given 1 mark. There is no point for participants who spelled incorrectly. There were 10 items. The total score was 15 marks.

3.2.4. Higher-order cognitive skill tests

The inference making task and the comprehension monitoring task were developed to test whether the participants derive coherent information and detect inconsistencies in stories. Both tasks were administered on Eprime 2.0. The inference making task was adapted from Muijselaar (2018). It included 10 stories at the difficulty level appropriate for HSK3-level readers. Each story was followed by one question. After reading the stories, questions were presented on the screen. Participants were given one mark if they answer one question correctly. The total score was 10. In inference making test, the reliability of the inferencing was 0.83.

The monitoring task was adapted from Kim (2015). It still included 10 stories at the difficulty level appropriate for HSK3-level readers, with 7 information inconsistencies stories, and 3 normal stories. Each story was presented on the screen and followed by a true-or-false question. Each question was given 1 mark, and the total score was 10. The reliability of the monitoring task was 0.87.

3.3. Data analysis strategy

The data were analyzed using three major statistical approaches. Correlation analysis was conducted to investigate the relationship between the variables. Multiple regression methods were used to determine the unique contribution of each cognitive and language skill to reading comprehension. Correlation and hierarchical multiple regression were conducted by SPSS 28.0. Path analysis was conducted by Mplus 8.3 (Muthén and Muthén, 2017), which is used for determining how well our proposed model fit the available data, as well as evaluating the significance of each direct and indirect path.

4. Results

4.1. Descriptive analysis and preliminary analysis

The results of descriptive statistics and correlation analysis of the variables were shown in Table 1. Table 1 included the mean and *SD*, skewness, kurtosis as well as the bivariate correlations among the measures. Skewness and kurtosis of all variables are within an acceptable range, demonstrating normal distribution. Correlation analyses used the Pearson correlation test. L2 Chinese reading grade was weakly to strongly correlated with fundamental skills (orthographic knowledge, vocabulary, syntactic knowledge, morphological knowledge), basic cognitive skills (working memory) and high-order cognitive skills (inference making and monitoring) in the Pearson correlation test ($0.16 \leq rs \leq 0.64$, $ps < 0.01$). Higher-order cognitive skills (inference making and monitoring) were strongly related to L2 Chinese reading ($0.63 \leq rs \leq 0.64$); There was a weak to strong relationship between basic cognitive skills (working memory), fundamental skills (orthographic knowledge, vocabulary, syntactic knowledge, morphological knowledge) and high-order skills (inference making and monitoring) ($0.21 \leq rs \leq 0.66$).

4.2. Hierarchical multiple regression

A hierarchical multiple regression was performed to figure out the contributions of these variables to Chinese reading comprehension. Orthography, vocabulary, syntactic knowledge, morphology, working memory, inferencing and monitoring were independent variables; L2 Chinese reading was the dependent variable. In multiple regressions, the unstandardized coefficient *B*, standardized coefficient β, standard error (*SE*), ΔR^2, and R^2 are presented in Table 2. Standardized coefficient β was adopted to interpret the contributions of each variable (Jordan et al., 2013). All variables made statistically significant contributions to Chinese reading comprehension. Orthographic knowledge ($\beta = 0.01$, $p < 0.05$), lexical knowledge ($\beta = 0.05$, $p < 0.05$), syntactic knowledge ($\beta = 0.05$, $p < 0.01$), morphological awareness ($\beta = 0.40$, $p < 0.01$), working memory ($\beta = 0.22$, $p < 0.01$), inference making ($\beta = 0.23$, $p < 0.01$), and comprehension monitoring ($\beta = 0.21$,

TABLE 1 Descriptive statistics and bivariate correlations among variables.

	M (SD)	Orthographic knowledge	Vocabulary	Syntactic knowledge	Morphological knowledge	Working memory	Inference making	Monitoring	Chinese reading
Orthographic knowledge	10.64 (1.77)	1							
Vocabulary	9.85 (1.64)	0.36**	1						
Syntactic knowledge	10.31 (1.53)	0.24**	0.15*	1					
Morphological knowledge	10.03(2.21)	0.69**	0.42**	0.34**	1				
Working memory	25.41 (1.83)	0.54**	0.54*	0.48**	0.61**	1			
Inference making	5.62 (1.71)	0.45**	0.21**	0.18**	0.66**	0.38**	1		
Monitoring	5.39 (1.89)	0.46**	0.56**	0.50**	0.59**	0.53**	0.30**	1	
Chinese reading		0.16**	0.24**	0.26**	0.64**	0.66**	0.63**	0.64**	1
	15.84 (1.78)								
Skewness		0.65	−0.57	−0.79	−0.77	−1.7	0.41	−0.7	0.54
Kurtosis		0.54	0.08	0.19	0.21	0.94	0.55	0.23	0.32

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

$p < 0.01$) had a significant influence on Chinese reading comprehension. Overall, the basic cognitive skills, foundational language skills and higher-order cognitive skills explained 80% (F (7, 243) =72.89; p <0.01) of the variances in L2 Chinese reading comprehension.

Another purpose of the analysis was to determine the amount of unique variance in L2 Chinese reading that could be explained by each independent variable, respectively. In order to achieve the purpose, hierarchical multiple regressions were conducted. In the first step of each regression, all other variables were entered except for one, which was entered in the second step to determine the unique variance explained by that particular variable. Statistical analysis revealed that each of these variables contributed to the L2 Chinese reading after accounting for all other variables: orthographic knowledge 1% ($p < 0.01$), vocabulary 4% ($p < 0.01$), syntactic knowledge 5% ($p < 0.01$), morphological knowledge 23% ($p < 0.01$), working memory 17% ($p < 0.01$), inference making 16% ($p < 0.01$) and comprehension monitoring 14% ($p < 0.01$). According to Cohen (1988) f^2's definition of effect sizes, small effect sizes as 0.02, medium effect sizes as 0.15, and large effect sizes as 0.35. The effect sizes of morphological skills (Cohen f^2 =0.85), working memory (Cohen f^2 =0.55), inference making (Cohen f^2 =0.5) and monitoring (Cohen f^2 =0.4) are large. The effect sizes of vocabulary (Cohen f^2 =0.2) and syntactic knowledge (Cohen f^2 =0.25) are medium. The effect size of the orthography skill (Cohen f^2 =0.02) was small.

4.3. A path analysis

The path analysis was carried out using Mplus 8.3 (Muthén and Muthén, 2017) and the full information maximum likelihood estimator. Model fits were evaluated by the following indices: the χ2 statistic, comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). It was suggested that chi-square test should be used to compare model fits and the model was rejected if $p < 0.05$ (Hu and Bentler, 1999). A good fit of the model is indicated by the χ^2 statistic to degrees of freedom ratio (χ^2/df) equal to or less than 3 (Mclver and Carmines, 1981), TLI and CFI equal to or greater than 0.95 (Kline, 2005). A cutoff value close to 0.08 for SRMR and 0.06 for RMSEA are needed to report a good fit (Hu and Bentler, 1999). RMSEA values below 0.08 and SRMR equal to or less than 0.05 indicate an excellent model fit (Kline, 2005). The model had a good fit shown in Table 3.

All the statistically significant paths and their standardized path coefficients are reported in the model, as shown in Figure 1 and Table 4. First, we focused on the direct relationship between three facets and L2 Chinese reading comprehension. The path coefficients (β) in the structural model ranging from 0 to 0.1, 0.11 to 0.30, 0.31 to 0.50, and > 0.50 are indicative of weak, modest, moderate, and strong effect sizes (Hair and Alamer, 2022). Statistically significant evidence was found for the direct effect of cognitive skills, inference making (β = 0.46, $p < 0.01$, moderate effect sizes) and comprehension monitoring (β = 0.41, $p < 0.01$, moderate effect sizes), on L2 Chinese reading. Inference making was predicted by orthographic knowledge (β = 0.04, p < 0.05, modest effect sizes), vocabulary (β = 0.23, $p < 0.01$, modest effect sizes), syntactic knowledge (β = 0.29, $p < 0.01$, modest effect sizes), morphological knowledge (β = 0.51, $p < 0.01$, strong

TABLE 2 Hierarchical multiple regression of the model predictors on reading comprehension.

Measures	L2 Chinese Reading Comprehension				
	B	*SE*	β	ΔR^2	R^2
Orthography	0.01	0.03	0.01*	0.01	0.8
Vocabulary	0.06	0.03	0.05*	0.04	
Syntax	0.06	0.03	0.05**	0.05	
Morphology	0.33	0.04	0.40**	0.23	
Working memory	0.23	0.05	0.22**	0.17	
Inferencing	0.25	0.05	0.23**	0.16	
Monitoring	0.23	0.05	0.21**	0.14	

*$p < 0.05$; **$p < 0.01$.

TABLE 3 A path analysis model fit.

Model fit indices	χ^2	*Df*	*P*	CFI	TLI	RMSEA	SRMR
	6.818	6	0.338	0.998	0.995	0.023	0.026

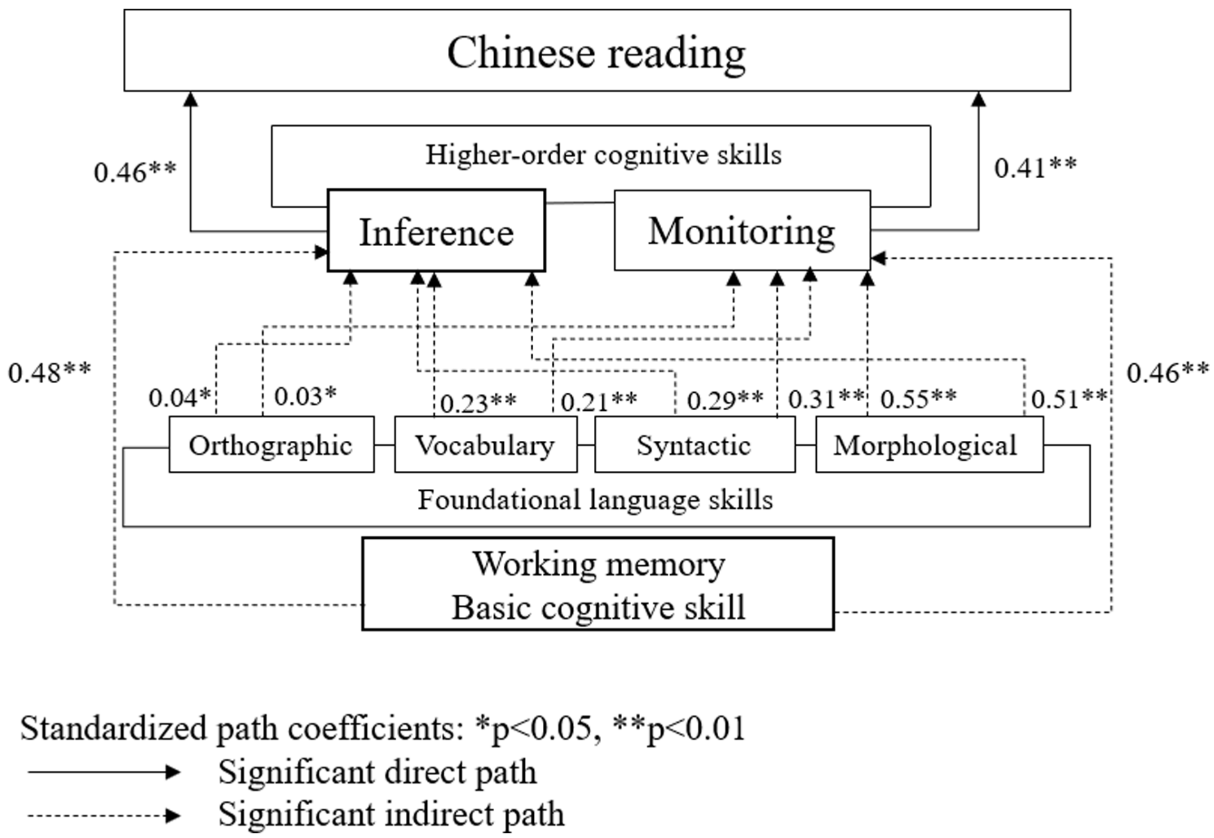


FIGURE 1
Standardized structural regression weights for the direct and indirect effects of the predictors in the L2 Chinese reading model.

effect sizes) and working memory (β = 0.48, $p < 0.01$, moderate effect sizes). Monitoring was predicted by orthographic knowledge (β = 0.03, $p < 0.05$, modest effect sizes), vocabulary (β = 0.21, $p < 0.01$, modest effect sizes), syntactic knowledge (β = 0.31, $p < 0.01$, moderate effect sizes), morphological knowledge (β = 0.55, $p < 0.01$, strong effect sizes) and working memory (β = 0.46, $p < 0.01$, moderate effect sizes).

Second, we focused on the indirect relationship between three facets and L2 Chinese reading comprehension. Orthographic knowledge, vocabulary, syntactic knowledge, morphological

TABLE 4 Direct, indirect, and total effect estimates of the model predictors on L2 Chinese reading.

	Predictors	Total effect	Direct effect	Indirect effect
Chinese reading	Orthographic knowledge	0.03	—	0.03
	Vocabulary	0.19	—	0.19
	Syntactic knowledge	0.26	—	0.26
	Morphological awareness	0.46	—	0.46
	Working memory	0.41	—	0.41
	Inference making	0.46	0.46	—
	Comprehension monitoring	0.41	0.41	—

knowledge and working memory had indirect effects on Chinese reading *via* two routes, inference making or comprehension monitoring. Inference making was predicted by Orthographic knowledge (ab = 0.018, 95% CI [0.012, 0.093]; weak effect sizes), vocabulary (ab = 0.106, 95% CI [0.019, 0.103]; modest effect sizes), syntactic knowledge (ab = 0.133, 95% CI [0.023, 0.119]; modest effect sizes), morphological knowledge (ab = 0.235, 95% CI [0.105, 0.218]; modest effect sizes) and working memory (ab = 0.221, 95% CI [0.093, 0.220]; modest effect sizes). Comprehension monitoring was predicted by orthographic knowledge (ab = 0.073, 95% CI [0.014, 0.099]; weak effect sizes), vocabulary (ab = 0.086, 95% CI [0.021, 0.112]; weak effect sizes), syntactic knowledge (ab = 0.127, 95% CI [0.020, 0.116]; modest effect sizes), morphological knowledge (ab = 0.226, 95% CI [0.102, 0.211]; modest effect sizes) and working memory (ab = 0.189, 95% CI [0.087, 0.191]; modest effect sizes). In general, the effect sizes of orthographic knowledge and vocabulary in L2 Chinese reading are modest, even though they are statistically significant. Therefore, the results of the path analysis highlight the indirect effects of working memory, morphological knowledge and syntactic knowledge on L2 Chinese reading comprehension *via* inference making and comprehension monitoring.

5. Discussion

For the first question, how do these three facets (the basic cognitive, foundational language skills, and higher-order cognitive skills) make contributions to L2 Chinese reading comprehension? This study used hierarchical multiple regressions to examine the unique contributions of the three facets to L2 reading in Chinese. The results showed that altogether these factors explained 80% of the variances in L2 Chinese reading. The results are partly supported by evidence from Zhao et al. (2022) model. We extended prior research findings by (a) including morphological awareness and working memory in the multiple regressions model; (b) using data from L2 Chinese learners.

The present model explained substantial variances (80%) in L2 Chinese reading comprehension. Morphological knowledge explained the largest variance, that is 23% of the variance in Chinese reading comprehension, compared with other linguistic skills (i.e., vocabulary 4%, syntactic knowledge 5%, orthographic knowledge 1%). The importance of morphological awareness was found in L1 English children (Carlisle, 2000; Nagy et al., 2006; Tong et al., 2011; Kieffer et al., 2013), L1 Chinese children (Cheng et al., 2017; Qiao et al., 2021; Pan and Lin, 2023), heritage learners of Chinese (Zhang and Koda, 2018). There is a lack of correspondence between the orthographic units and phonemes in Chinese, which is different from alphabetic languages. This requires native Chinese speakers to rely on morphological knowledge, especially L1 Chinese children (e.g., Cheng et al., 2017). They need to develop a sensibility of how Chinese words represent meaning at the morphemic level because Chinese has a logography that contains a large number of homophones and polysemes. These structures are challenging for L1 Chinese learners and can interfere with comprehension (Leong and Ho, 2008). L1 Chinese readers with stronger morphological skills have quicker access to the lexicon and are able to extract meaning more easily from the text. Especially when they read words they do not already know, a morphological analysis may assist them in making an educated guess about the semantic and syntactic category of an unknown word. In the current study, the importance of morphological awareness was extended to L2 Chinese learners. The participants were college-level international students from alphabetic language backgrounds with well-developed cognitive skills and ample reading experience in their native language, but the results showed that they still relied on morphological knowledge. However, it should be noticed that morphological awareness influenced L2 Chinese reading indirectly through cognitive skills, which is an interesting finding we will now turn to.

For the second question, what are the direct or indirect roles each facet plays in the process of L2 Chinese reading comprehension?

The results indicated that that working memory was directly related to inference making and monitoring and indirectly contributed to L2 Chinese reading comprehension *via* the two higher-order cognitive skills. These cognitive skills are cooperating with each other when supporting reading comprehension. Working memory holds and retrieves relevant information in readers' mind when processing incoming language input. At the same time, working memory inhibits guessing errors when readers encounter contradictory information (Van Reybroeck and De Rom, 2020). For example, when Chinese readers encounter an orthographic neighbor, which replaced strokes with an orthographically close Chinese character (such as 友 (friends) v.s. 发(hair)), poor readers may fail to monitor and inhibit Chinese lexical orthographic neighbors in lexical competition, consequently causing a negative influence on the Chinese reading comprehension. An inference in reading comprehension is the process of filling in information gaps by manipulating the clues stored in working memory. According to Yuill and Oakhill (1991), a lack of adequate inferencing occurs when one is not aware of the fact that inferencing is necessary, when one lacks the background knowledge, or when one's working memory is inadequate.

This finding extended the existing research findings about the role of working memory in Chinese reading (Weng et al., 2016; Liu and Wu, 2022; Yeung, 2022), showing a more complex relationships among working memory, inference making, comprehension monitoring and L2 Chinese reading comprehension.

A notable result in the current study was that inferencing and monitoring mediated the relationship between morphological awareness and reading comprehension. This is in line with the findings by Ke and Koda (2017) who have observed that L1 English learners of Chinese who were more sensitive to the morphological structure of multi-character words were more successful in inferring the meanings of unfamiliar words.

For native speakers of alphabetic language such as English, it has been found that morphological awareness can contribute to reading comprehension directly or indirectly *via* a range of other foundational language skills such as vocabulary (Kieffer et al., 2013) and phonological awareness (Nagy et al., 2006). In our study, morphological awareness and other language skills turned out to be parallel factors whose effects on reading comprehension was mediated by higher-order cognitive skills. Previous studies (e.g., Nagy et al., 2006; Kieffer et al., 2013) did not consider the hierarchal relationships among cognitive skills and language skills. The variations among these findings could be explained by the specific component skills under investigation in different models.

Our findings are also slightly different from a recent study on L1 English learners of L2 Chinese (Zhang et al., 2021). They found that morphological awareness directly contributed to reading comprehension, while lexical inference directly contributed to reading and mediated the effect of morphological awareness on reading comprehension. In our case, we only found indirect contribution of morphological awareness to L2 Chinese reading through the two higher-order cognitive skills (i.e., inference making and monitoring). The discrepancy may be explained by different measurements of morphological awareness. The morphological awareness in the current study was composed of a test on homophones and a test on compounds, while Zhang et al. (2021) adopted a morpheme segmentation and a morpheme discrimination task. In addition, Zhang et al. (2021) only include a narrow range of reading-related skills (morphological awareness, vocabulary, and lexical inference), which may not be able to unveil a full picture, as shown in this study, of the hierarchal relationships among varying language and cognitive skills in reading comprehension.

In addition to morphological knowledge, our findings showed that inference making and monitoring skills mediated the relationship between syntactic knowledge and L2 Chinese reading comprehension. Chinese differs from alphabetic languages in several ways in terms of the syntactic features such as the absence of an inflectional system, more flexible word order, and more extensive use of connectives (Li and Thompson, 1981). The mediating role of higher-order cognitive skills may be explained by the possibility that L2 Chinese readers need to consume more cognitive resources during inference making and monitoring to comprehend Chinese sentences. Hung's (2021) recent study also found the intertwined relationship between syntactic skills and cognitive skills and found an indirect effect of executive function (as measured by working memory tasks and inhibition tasks) on Chinese Grade 2 and 3 students' reading comprehension through syntactic awareness (as measured by a word order correction task). It should be noticed that executive function covered some aspects of working memory and monitoring. In Hung's study, indirect effects of executive function on reading through inference-making was not significant, refuting her original hypothesis that executive function can support inference making to achieve passage-level comprehension. However, in our model, working memory, inference making, and monitoring were measured independently and entered the model as three variables. We found that higher-order cognitive skills (i.e., inference making and monitoring) completely mediated the effect of the basic cognitive skill (i.e., working memory) on reading comprehension. Syntactic skill made no direct contribution to reading comprehension, but it made indirect contributions through both inference making and monitoring, which is in line with Zhao et al.'s (2022) findings from L1 Chinese readers. By adding more micro-level componential skills into a more holistic model of reading comprehension, we extended Hung's findings and showed a clear hierarchical relationship between higher-order cognitive skills and the basic cognitive skill as well as the hierarchical relationship between higher-order cognitive skills and foundational language skills in reading comprehension.

Regarding inference making and monitoring, we found that monitoring and inference making made direct and moderate contributions to L2 Chinese reading comprehension, supporting the theory that developing a text representation with coherence requires evaluation and integration of initial propositions (Perfetti et al., 2005). We extend previous findings to a different population with L2 Chinese readers, demonstrating that higher-order cognitive skills are essential for both L2 Chinese readers, L1 Chinese readers (Zhao et al., 2022), and L1 English readers (Cain et al., 2004).

6. Limitation and future research

This study is limited in the following aspects. First, we did not include other factors that have been proven to be crucial to reading comprehension such as phonological awareness (Veenendaal et al., 2016; Lin and Zhang, 2021) and reading strategies (Yapp et al., 2021). Future research can explore more factors to establish a more comprehensive model of L2 Chinese reading. Second, the participants in this study have heterogeneous language backgrounds. Although their mother tongues belong to the broad alphabetic language family, the heterogeneity of their first language has not been controlled in the current study. First language may exert great influence on reading comprehension. For example, coherence relations have different forms across languages and first language may entail misuse of cohesive devices in another language (Wetzel et al., 2022). Future studies can take language background as a key variable in developing a more refined model of Chinese reading.

7. Conclusion

Reading is a complex process involving complex interactions between linguistic and cognitive skills. The results of this study showed that cognitive skills and foundational language skills can influence Chinese reading through direct and indirect pathways. In addition, there is a clear hierarchical relationship between the contributions of cognitive skills and linguistic skills. Higher-order cognitive skills completely mediated the contributions of foundational language skills and the basic cognitive skill. Also, a systematic understanding of reading comprehension needs to consider both general language-related issues and special features of a specific language. Chinese morphosyllabic orthography is different from many alphabetic languages and entails a large body of homophones. This uniqueness makes morphological awareness exceptionally important in Chinese reading comprehension. Future studies can further explore the role of other components of the linguistic-cognitive skills such as the central executive and the phonological loop of working memory in Chinese reading comprehension.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Xi'an Jiao Tong University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LY was responsible for conducting the research, writing and revising the manuscript. YX was responsible for writing and revising the manuscript. LY and YX responded to all the reviewers' comments. QC helped in editing the final version of the manuscript. First authorship goes to LY. Second authorship goes to YX. Third authorship goes to QC. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Ministry of Education, People's Republic of China (grant no. 18YJA740063).

Acknowledgments

The authors thank the many students who participated in this study. The authors also thank two reviewers for their patience and helpful comments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Cain, K. (2009). Making sense of text: skills that support text comprehension and its development. *Perspect. Lang. Liter.* 35, 11–14.

Cain, K., Oakhill, J. V., Barnes, M. A., and Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Mem. Cogn.* 29, 850–859. doi: 10.3758/BF03196414

Cain, K., Oakhill, J., and Bryant, P. (2004). Children's Reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills. *J. Educ. Psychol.* 96, 31–42. doi: 10.1037/0022-0663.96.1.31

Carlisle, J. F. (2000). Awareness of the structure and leaning of morphologically complex words: impact on reading. *Read. Writ.* 12, 169–190. doi: 10.1023/A:1008131926604

Chen, X., Hao, M., Geva, E., Zhu, J., and Shu, H. (2009). The role of compound awareness in Chinese children's vocabulary acquisition and character reading. *Read. Writ.* 22, 615–631. doi: 10.1007/s11145-008-9127-9

Cheng, Y., Zhang, J., Li, H., Wu, X., Liu, H., Dong, Q., et al. (2017). Growth of compounding awareness predicts Reading comprehension in young Chinese students: a longitudinal study from grade 1 to grade 2. *Read. Res. Q.* 52, 91–104. doi: 10.1002/rrq.155

Cheung, H., Chan, M., and Chong, K. (2007). Use of orthographic knowledge in Reading by Chinese-English bi-scriptal children. *Lang. Learn.* 57, 469–505. doi: 10.1111/j.1467-9922.2007.00423.x

Chik, P. P., Ho, C. S., Yeung, P., Wong, Y., Chan, D. W., Chung, K. K., et al. (2012). Contribution of discourse and morphosyntax skills to reading comprehension in Chinese dyslexic and typically developing children. *Ann. Dyslexia* 62, 1–18. doi: 10.1007/s11881-010-0045-6

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbraum Associates.

Currie, N. K., and Cain, K. (2015). Children's inference generation: the role of vocabulary and working memory. *J. Exp. Child Psychol.* 137, 57–75. doi: 10.1016/j.jecp.2015.03.005

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Fesel, S. S., Segers, E., Clariana, R. B., and Verhoeven, L. (2015). Quality of children's knowledge representations in digital text comprehension: evidence from pathfinder networks. *Comput. Hum. Behav.* 48, 135–146. doi: 10.1016/j.chb.2015.01.014

Florit, E., and Cain, K. (2011). The simple view of Reading: is it valid for different types of alphabetic orthographies? *Educ. Psychol. Rev.* 23, 553–576. doi: 10.1007/s10648-011-9175-6

Frost, S. J., Sandak, R., Mencl, W. E., Landi, N., Moore, D., Porta, G. D., et al. (2008). "Neurobiological and behavioral studies of skilled and impaired word Reading" in *Single-word reading: Behavioral and biological perspectives*. eds. E. L. Grigorenko and A. J. Naples (Mahwah: Erlbaum), 355–376.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hove: Psychology Press.

Graesser, A. C., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychol. Rev.* 101, 371–395. doi: 10.1037/0033-295X.101.3.371

Hair, J., and Alamer, A. (2022). Partial least squares structural equation modeling (PLS-SEM) in second language and education research: guidelines using an applied example. *Res. Methods Appl. Linguist.* 1:100027. doi: 10.1016/j.rmal.2022.100027

Hasher, L., and Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. *Psychol. Learn. Motiv.* 22, 193–225. doi: 10.1016/S0079-7421(08)60041-9

Hoover, W. A., and Gough, P. B. (1990). The simple view of reading. *Read. Writ.* 2, 127–160. doi: 10.1007/BF00401799

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118

Hung, C. O. (2021). The role of executive function in reading comprehension among beginning readers. *Br. J. Educ. Psychol.* 91, 600–616. doi: 10.1111/bjep.12382

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., and Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *J. Exp. Child Psychol.* 116, 45–58.

Ke, S. E., and Koda, K. (2017). Contributions of morphological awareness to adult L2 Chinese word meaning inferencing. *Mod. Lang. J.* 101, 742–755. doi: 10.1111/modl.12428

Kieffer, M. J., Biancarosa, G., and Mancilla-Martinez, J. (2013). Roles of morphological awareness in the reading comprehension of Spanish-speaking language minority learners: exploring partial mediation by vocabulary and reading fluency. *Appl. Psycholinguist.* 34, 697–725. doi: 10.1017/S0142716411000920

Kim, Y.-S. (2015). Language and cognitive predictors of text comprehension: evidence from multivariate analysis. *Child Dev.* 86, 128–144. doi: 10.1111/cdev.12293

Kim, Y. S. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *J. Exp. Child Psychol.* 141, 101–120. doi: 10.1016/j.jecp.2015.08.003

Kim, Y.-S. G. (2017). Why the simple view of Reading is not simplistic: unpacking component skills of Reading using a direct and indirect effect model of Reading (DIER). *Sci. Stud. Read.* 21, 310–333. doi: 10.1080/10888438.2017.1291643

Kim, K., and Clariana, R. B. (2015). Knowledge structure measures of Reader's situation models across languages: translation engenders richer structure. *Tech. Know Learn.* 20, 249–268. doi: 10.1007/s10758-015-9246-8

Kim, K., and Clariana, R. B. (2017). Text signals influence second language expository text comprehension: knowledge structure analysis. *Educ. Tech. Res. Dev.* 65, 909–930. doi: 10.1007/s11423-016-9494-x

Kinnunen, R., Vauras, M., and Niemi, P. (1998). Comprehension monitoring in beginning readers. *Sci. Stud. Read.* 2, 353–375. doi: 10.1207/s1532799xssr0204_4

Kline, R. B. (2005). *Principles and practice of structural equation modeling. 2nd*. New York, NY: Guilford Press.

Kuo, L., and Anderson, R. C. (2006). Morphological awareness and learning to read: a cross-language perspective. *Educ. Psychol.* 41, 161–180. doi: 10.1207/s15326985ep4103_3

Leong, C. K., and Ho, M. K. (2008). The role of lexical knowledge and related linguistic components in typical and poor language comprehenders of Chinese. *Read. Writ.* 21, 559–586. doi: 10.1007/s11145-007-9113-7

Li, C. N., and Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Lin, Y.-A. (2006). Against the deficit in computational grammatical complexity hypothesis: a corpus-based study. *Concent. Stud. Linguist.* 32, 59–70.

Lin, J., and Zhang, H. (2021). Cross-linguistic influence of phonological awareness and phonological recoding skills in Chinese reading acquisition among early adolescent students. *J. Gene. Psycho.* 1–22. doi: 10.1080/00221309.2021.1922345

Liu, Y., Georgiou, G. K., Zhang, Y., Li, H., Liu, H., Song, S., et al. (2017). Contribution of cognitive and linguistic skills to word-reading accuracy and fluency in Chinese. *Int. J. Educ. Res.* 82, 75–90. doi: 10.1016/j.ijer.2016.12.005

Liu, C.-Y., and Wu, C.-J. (2022). Effects of working memory and relevant knowledge on reading texts and infographics. *Read. Writ.* doi: 10.1007/s11145-022-10368-1

Lo, L., Ho, C. S., Wong, Y., Chan, D. W., and Chung, K. K. (2016). Understanding the microstructure and macrostructure of passages among Chinese elementary school children. *J. Psycholinguist. Res.* 45, 1287–1300. doi: 10.1007/s10936-015-9402-2

McBride-Chang, C., Shu, H., Zhou, A., Wat, C. P., and Wagner, R. K. (2003). Morphological awareness uniquely predicts young Children's Chinese character recognition. *J. Educ. Psychol.* 95, 743–751. doi: 10.1037/0022-0663.95.4.743

Mclver, J. P., and Carmines, E. G. (1981). *Unidimensional Scaling*. Beverly Hills, CA: Sage Publications.

Muijselaar, M. M. L. (2018). The dimensionality of inference making: are local and global inferences distinguishable? *Sci. Stud. Read.* 22, 117–136. doi: 10.1080/10888438.2017.1371179

Muthén, L. K., and Muthén, B. O. (2017). *Mplus User's Guide*. *8th*. Los Angeles, CA: Muthén & Muthén.

Nagy, W., Berninger, V. W., and Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *J. Educ. Psychol.* 98, 134–147. doi: 10.1037/0022-0663.98.1.134

Oakhill, J. (2020). Four decades of research into Children's Reading comprehension: a personal review. *Discourse Process.* 57, 402–419. doi: 10.1080/0163853X.2020.1740875

Oakhill, J., Cain, K., and McCarthy, D. (2015). "Inference processing in children: the contributions of depth and breadth of vocabulary knowledge" in *Inferences during Reading*. eds. E. J. O'Brien, A. E. Cook and R. F. Lorch (Cambridge: Cambridge University Press), 140–159.

Oberauer, K., Wendland, M., and Kliegl, R. (2003). Age differences in working memory— the roles of storage and selective access. *Mem. Cogn.* 31, 563–569. doi: 10.3758/BF03196097

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge, UK: Cambridge University Press.

Pan, D. J., and Lin, D. (2023). Cognitive–linguistic skills explain Chinese Reading comprehension within and beyond the simple view of Reading in Hong Kong kindergarteners. *Lang. Learn.* 73, 126–160. doi: 10.1111/lang.12515

Perfetti, C. A., Landi, N., and Oakhill, J. (2005). "The acquisition of reading comprehension skill," in *The Science of Reading: A Handbook*. eds. M. J. Snowling and C. Hulme (Malden, MA: Blackwell Publishing), 227–247.

Perfetti, C., and Stafura, J. (2014). Word knowledge in a theory of Reading comprehension. *Sci. Stud. Read.* 18, 22–37. doi: 10.1080/10888438.2013.827687

Qiao, S., Liu, Y., and Yeung, S. S. (2021). Mechanisms by which morphological awareness contributes to Chinese and English reading comprehension. *J. Res. Read.* 44, 189–209. doi: 10.1111/1467-9817.12345

Seigneuric, A., and Ehrlich, M.-F. (2005). Contribution of working memory capacity to Children's Reading comprehension: a longitudinal investigation. *Read. Writ.* 18, 617–656. doi: 10.1007/s11145-005-2038-0

Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., and Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Read. Writ.* 13, 81–103. doi: 10.1023/A:1008088230941

Shu, H., Anderson, R. C., and Wu, N. (2000). Phonetic awareness: knowledge of orthography–phonology relationships in the character acquisition of Chinese children. *J. Educ. Psychol.* 92, 56–62. doi: 10.1037/0022-0663.92.1.56

Shu, H., Chen, X., Anderson, R. C., Wu, N., and Xuan, Y. (2003). Properties of school Chinese: implications for learning to read. *Child Dev.* 74, 27–47. doi: 10.1111/1467-8624.00519

Silva, M., and Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *J. Educ. Psychol.* 107, 321–331. doi: 10.1037/a0037769

Singer, M., Graesser, A. C., and Trabasso, T. (1994). Minimal or global inference during Reading. *J. Mem. Lang.* 33, 421–441. doi: 10.1006/jmla.1994.1020

Strasser, K., and Del Río, F. (2013). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Read. Res. Q.* 49, 169–187. doi: 10.1002/rrq.68

Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., and Parrila, R. (2011). Morphological awareness: a key to understanding poor reading comprehension in English. *J. Educ. Psychol.* 103, 523–534. doi: 10.1037/a0023495

Tong, X., Tong, X., Shu, H., Chan, S., and McBride-Chang, C. (2014). Discourse-level reading comprehension in Chinese children: what is the role of syntactic awareness?: syntactic awareness in discourse-level reading comprehension. *J. Res. Read.* 37, S48–S70. doi: 10.1111/1467-9817.12016

Tunmer, W. E., and Chapman, J. W. (2012). The simple view of Reading redux: vocabulary knowledge and the independent components hypothesis. *J. Learn. Disabil.* 45, 453–466. doi: 10.1177/0022219411432685

Van Dijk, T., and Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Van Reybroeck, M., and De Rom, M. (2020). Children with dyslexia show an inhibition domain-specific deficit in reading. *Read. Writ.* 33, 907–933. doi: 10.1007/s11145-019-09986-z

Veenendaal, N. J., Groen, M. A., and Verhoeven, L. (2016). The contribution of segmental and suprasegmental phonology to reading comprehension. *Read. Res. Q.* 51, 55–66. doi: 10.1002/rrq.127

Weng, X., Li, G., and Li, R. (2016). Mediating effects of working memory in the relation between rapid automatized naming and Chinese Reading comprehension. *J. Psycholinguist. Res.* 45, 945–959. doi: 10.1007/s10936-015-9385-z

Wetzel, M., Crible, L., and Zufferey, S. (2022). Processing clause-internal discourse relations in a second language: a case study of specifications in German and French. *JSLS* 5, 206–234. doi: 10.1075/jsls.21032.wet

Yang, J., Zevin, J., Shu, H., McCandliss, B. D., and Li, P. (2006). A "triangle model" of Chinese Reading available at: https://www.semanticscholar.org/paper/A-%22Triangle-Model%22-of-Chinese-Reading-Li-McCandliss/35fa6ac1c098158ee46d6332673b51f48ca3fac2 (accessed December 26, 2022).

Yapp, D. J., de Graaff, R., and van den Bergh, H. (2021). Improving second language reading comprehension through reading strategies: a meta-analysis of L2 reading strategy interventions. *JSLS* 4, 154–192. doi: 10.1075/jsls.19013.yap

Yeari, M., and van den Broek, P. (2011). A cognitive account of discourse understanding and discourse interpretation: the landscape model of reading. *Discourse Stud.* 13, 635–643. doi: 10.1177/1461445611412748

Yeomans-Maldonado, G. (2017). Development of comprehension monitoring in beginner readers. *Read. Writ. Interdiscip. J.* 30, 2039–2067. doi: 10.1007/s11145-017-9765-x

Yeung, P. (2022). A model of Chinese reading comprehension: the role of cognition and motivation. *Learn. Individ. Differ.* 99:102202. doi: 10.1016/j.lindif.2022.102202

Yeung, P., Ho, C. S., Chan, D. W., and Chung, K. K. (2016). Orthographic skills important to Chinese literacy development: the role of radical representation and orthographic memory of radicals. *Read. Writ.* 29, 1935–1958. doi: 10.1007/s11145-016-9659-3

Yuill, N., and Oakhill, J. (1991). *Children's Problems in Text Comprehension: An Experimental Investigation*. New York: Cambridge University Press.

Zargar, E., Adams, A., and Connor, C. M. (2020). The relations between Children's comprehension monitoring and their Reading comprehension and vocabulary knowledge: an eye-movement study. *Read. Writ.* 33, 511–545. doi: 10.1007/s11145-019-09966-3

Zhang, H. (2014). Morphological awareness in literacy Acquisition of Chinese Second Graders: a path analysis. *J. Psycholinguist. Res.* 45, 103–119. doi: 10.1007/s10936-014-9327-1

Zhang, H., and Koda, K. (2018). Vocabulary knowledge and morphological awareness in Chinese as a heritage language (CHL) reading comprehension ability. *Read. Writ.* 31, 53–74. doi: 10.1007/s11145-017-9773-x

Zhang, H., Zhang, X., Li, M., and Zhang, Y. (2021). Morphological awareness in L2 Chinese Reading comprehension: testing of mediating routes. *Front. Psychol.* 12:736933. doi: 10.3389/fpsyg.2021.736933

Zhao, A., Guo, Y., and Dinnesen, M. S. (2022). The direct and indirect effects of language and cognitive skills on Chinese reading comprehension. *Read. Writ.* 35, 539–564. doi: 10.1007/s11145-021-10192-z

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 19 April 2023
DOI 10.3389/fpsyg.2023.1143062

Check for updates

OPEN ACCESS

EDITED BY
Linjun Zhang,
Peking University, China

REVIEWED BY
Ileana Ratiu,
Arizona State University, United States
Jingjing Zhao,
Shaanxi Normal University, China

*CORRESPONDENCE
Luyao Chen
✉ luyaochen@bnu.edu.cn
Liping Feng
✉ fengliping@bnu.edu.cn

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

RECEIVED 12 January 2023
ACCEPTED 30 March 2023
PUBLISHED 19 April 2023

CITATION
Chen L, Yang M, Gao F, Fang Z, Wang P and Feng L (2023) Mandarin Chinese L1 and L2 complex sentence reading reveals a consistent electrophysiological pattern of highly interactive syntactic and semantic processing: An ERP study.
Front. Psychol. 14:1143062.
doi: 10.3389/fpsyg.2023.1143062

Mandarin Chinese L1 and L2 complex sentence reading reveals a consistent electrophysiological pattern of highly interactive syntactic and semantic processing: An ERP study

Luyao Chen[1,2]*, Mingchuan Yang[1], Fei Gao[3,4], Zhengyuan Fang[1], Peng Wang[5,6,7] and Liping Feng[1]*

[1]Max Planck Partner Group, School of International Chinese Language Education, Beijing Normal University, Beijing, China, [2]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [3]Institute of Modern Languages and Linguistics, Fudan University, Shanghai, China, [4]Centre for Cognitive and Brain Sciences, University of Macau, Macao, Macao SAR, China, [5]Methods and Development Group (MEG and Cortical Networks), Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [6]Institute of Psychology, University of Greifswald, Greifswald, Germany, [7]Institute of Psychology, University of Regensburg, Regensburg, Germany

Introduction: A hallmark of the human language faculty is processing complex hierarchical syntactic structures across languages. However, for Mandarin Chinese, a language typically dependent on semantic combinations and free of morphosyntactic information, the relationship between syntactic and semantic processing during Chinese complex sentence reading is unclear. From the neuropsychological perspective of bilingual studies, whether second language (L2) learners can develop a consistent pattern of target language (i.e., L2) comprehension regarding the interplay of syntactic and semantic processing, especially when their first language (L1) and L2 are typologically distinct, remains to be determined. In this study, Chinese complex sentences with center-embedded relative clauses were generated. By utilizing the high-time-resolution technique of event-related potentials (ERPs), this study aimed to investigate the processing relationships between syntactic and semantic information during Chinese complex sentence reading in both Chinese L1 speakers and highly proficient L2 learners from South Korea.

Methods: Normal, semantically violated (SEM), and double-violated (containing both semantic and syntactic violations, SEM + SYN) conditions were set with regard to the nonadjacent dependencies of the Chinese complex sentence, and participants were required to judge whether the sentences they read were acceptable.

Results: The ERP results showed that sentences with "SEM + SYN" did not elicit early left anterior negativity (ELAN), a component assumed to signal initial syntactic processing, but evoked larger components in the N400 and P600 windows than those of the "SEM" condition, thus exhibiting a biphasic waveform pattern consistent for both groups and in line with previous studies using simpler Chinese syntactic structures. The only difference between the L1 and L2 groups was that L2 learners presented later latencies of the corresponding ERP components.

Discussion: Taken together, these results do not support the temporal and functional priorities of syntactic processing as identified in morphologically rich languages (e.g., German) and converge on the notion that even for Chinese complex sentence reading, syntactic and semantic processing are highly interactive. This is consistent across L1 speakers and high-proficiency L2 learners with typologically different language backgrounds.

KEYWORDS

Chinese complex sentence, syntactic information, semantic information, ELAN, block effect, ERP

1. Introduction

The relationship between syntactic and semantic information during language comprehension has received considerable attention in psycholinguistic academia in recent decades. Whether and to what extent such a relationship is modulated by factors such as language typological differences and language proficiency are not well understood and await specification. Specifically, two research gaps were indentified in the development of research questions of the current study: (1) The relationship between syntactic and semantic processing during Chinese complex sentence reading is unclear; (2) Whether second language (L2) learners can develop a consistent pattern of the target language (i.e., L2) comprehension regarding the interplay of syntactic and semantic processing remains to be determined. Therefore, in *Introduction*, after setting the general background of the ERP (event-related potential) studies in morphologically rich languages and in Mandarin Chinese (1.1 and 1.2), we highlighted the importance of using complex sentences as experimental materials (1.3), which was then followed by the introduction of L2 settings (1.4). At last, we introduced the development of experimental design (esp., the double-violation paradigm) (1.5), and specified the research aims and expectations of the present study (1.6).

1.1. A syntax-first model in morphologically rich languages

A prominent cognitive model of auditory language comprehension (Friederici, 2002, 2011, 2017) holds that syntactic information precedes and then interacts with semantics, and this early syntactic information mainly deals with syntactic categories of words. In particular, word category information has been recognized as the foundation of syntactic structure building (Lenneberg, 1967; Chomsky, 1995; Hornstein and Nunes, 2008; Hornstein, 2009; Hornstein and Pietroski, 2009; Adger, 2013; Miyagawa et al., 2013; Epstein et al., 2014; Hoshi, 2018, 2019), inspiring a series of studies on syntactic processing that adopt various word category information (Bemis and Pylkkänen, 2011; Miyagawa et al., 2013; Fujita, 2014; Goucha and Friederici, 2015; Friederici, 2017; Goucha et al., 2017; Zaccarella et al., 2017; Chen et al., 2019). Using the event-related potentials (ERPs) technique of high temporal resolution, previous studies have investigated the interplay between syntactic category and semantic information processing in morphologically rich languages (especially German). The evidence converged to demonstrate a priority of syntactic processing over semantic processing, supporting a "syntax-first model." For instance, in the double-violation/combined violation paradigm (e.g., Hahne and Jescheniak, 2001; Hahne and Friederici, 2002) during German sentence comprehension, the simultaneous violation of syntactic and semantic information elicited pronounced early left anterior negativity (ELAN) in the time window of 120–200 ms (see Neville et al., 1991, Friederici et al., 1993 for more information on ELAN; see Friederici and Weissenborn, 2007 for a systematic review), a negative component assumed to signal initial syntactic processing, while N400 was absent, a classic negative component reflecting semantic violations in the time window of 300–500 ms (see Kutas and Hillyard, 1980, 1983, and Friederici et al., 1993 for more information on N400; see Kutas and Federmeier, 2011 for an overview). ELAN could therefore index a temporal priority of syntactic category processing. The absence of N400 suggested an inhibition on the following semantic processing from the failure of syntactic encoding, which was defined as a "block effect" (Yu and Zhang, 2008). Experiments on French (Isel et al., 2007), English (Yamada and Neville, 2007), and Dutch (Hagoort et al., 2003) reported similar electrophysiological patterns regarding the time course of syntactic processing in sentence comprehension. At a two-word-phrase level, syntactic violation in German could elicit an ELAN-like early syntactic negativity (Maran et al., 2022). Therefore, syntactic processing in such languages was primarily evidenced to be both temporally and functionally prior to semantic processing. Nevertheless, emerging evidence from Indo-European languages began to conflict with the syntax-first account. For example, in a recent ERP study of French sentence reading where uninflected nouns and verbs were swapped (Fromont et al., 2020), ELAN did not appear in syntactic category violations. These findings suggested that inflectional cues might trigger and thus speed up syntactic processes (see also Dikker et al., 2009).

1.2. The interplay of syntactic and semantic processing in Mandarin Chinese

Syntactic category information is primarily marked by morphological inflections in these morphologically rich languages. For instance, "*gegessen*" (eaten) is the past participle of "*essen*" (eat) in the German material "*Das Brot wurde gegessen*" (The bread was eaten) in Hahne and Friederici's study (2002). These morphological cues might facilitate access to word category information and lead to syntactic priority (Friederici, 2017). However, there is no such inflection in morpho-syllabic languages such as Mandarin Chinese (hereafter, Chinese) (DeFrancis, 1989; Shen, 2016; Gao et al., 2022), where word order and functional words constitute the principle grammatical operations (Zhu, 1982). Existing studies have therefore examined whether the syntax-first model fits Chinese processing as well by using various syntactic structures. Ye et al. (2006) first approached this issue in the Chinese BA ("把")[1] structure (i.e., subject + BA + object + VP). Their results revealed that double-violations

1 BA structure is constructed by a subject (NP1) followed by the preposition BA and the object (i.e., NP2), and finally a VP. The preposition BA forces the default word order in Chinese [i.e., subject–verb–object (SVO)] into subject–BA–object–verb to emphasize the meaning of disposal. For example, 设计师把布料裁了 [literal glosses: the stylist BA the cloth cut le (*functional word*); translation: the stylist cut the cloth] (Jin, 1997; Ye et al., 2006; Zhang et al., 2010).

elicited ELAN and a continuous negative wave in the time window of 250–400 ms. This pattern was similar to the findings of morphologically rich languages. However, this study could not rule out the possibility that N400 was included in the continuous negative component due to the existence of numerous homonyms in the Chinese lexicon (Zhang et al., 2010). Yu and Zhang (2008) also examined this issue in the Chinese BA structure with more careful control and included semantic violations and double violations. The reason for not including syntactic violations was twofold. First, Chinese word category violations are always accompanied by semantic violations such that there should be no pure syntactic violations in Chinese. Second, to examine whether syntactic processing would manifest a temporal and functional priority in sentence comprehension, the presence of ELAN and "block effect" in the double-violation condition is sufficient to draw a conclusion. Interestingly, their results demonstrated a biphasic N400-P600 pattern, where double violations were associated with significantly greater N400 and P600 compared to semantic violations. In light of the null results that both ELAN and the block effect did not appear, the authors concluded that Chinese syntactic processing might not exhibit a priority over semantic operations during sentence comprehension, and the two types of processing should be highly interactive. These results were replicated in a study focusing on BA construction and subject-verb-object (SVO) structures by Zhang et al. (2010). In addition, Wang S. et al. (2013) and Wang et al. (2015) manipulated the transitivity of verbs in Chinese BA/BEI ("被")[2] and NP1 + VP + NP2 structures in light of a double-violation paradigm. Their results revealed that double violations and pure semantic violation elicited comparable N400 and late positivities, again not supporting the application of the syntax-first model to Chinese sentence comprehension. Two additional studies re-examined this issue in Chinese passive BEI (Zeng et al., 2020) and Qing ("请")[3] structures (Yang et al., 2021). Zeng et al. (2020) obtained similar results as Zhang et al. (2010), Wang S. et al. (2013), and Wang et al. (2015), while Yang et al. (2021) also observed an interaction between syntactic category and semantic processing in an early time window of 100–300 ms. A very recent study using intracranial high-density electrocorticography found that syntactic and semantic processing in Chinese showed spatial–temporal separations (Zhu et al., 2022). Only the local syntactic violation, not the syntactic category violation, elicited an ELAN-like component in the early time window. However, it is not clear whether this inconsistent pattern resulted from the different types of syntactic violations or from the differences in local preferences (see below).

2 BEI structure [subject (NP1)+BEI+object (NP2)+VP] in Chinese is used as the passive voice structure, which is similar to the passive voice structure in English [subject (NP1)+be + VP-ed+by + object (NP2)]. E.g., 自行车被小偷偷走了 [literal glosses: the bike BEI a thief stolen away le (*functional word*); translation: the bike was stolen away by a thief] (Wang S. et al., 2013; Wang et al., 2015).

3 Chinese QING structure consists of the following items in sequence: QING + NP1+VP+NP2, which is similar to "ask + NP1+to + VP+NP2" in English. Qing structure is frequently used to express imperative mood in Chinese, and it sounds polite. E.g., 请你打扫这个房间 [literal glosses: QING you clean up this room; translation: please clean up this room] (Yang et al., 2021).

1.3. Local versus long-distance dependencies: Limitations on syntactically simple sentences

Although the aforementioned studies were in favor of the null primacy of syntactic information of distinguishing structures in Chinese simple sentences, it remains unclear how syntactic and semantic processing interact with each other when reading relatively complex sentences. One significant feature of the human language faculty is that language parsers can comprehend complex hierarchical sentences containing center-embedded relative clauses (Makuuchi et al., 2009, 2013; Friederici, 2017). For example, in the complex sentence "The dog the cat chased barked," the relative clause "the cat chased" is embedded between the subject "the dog" and the verb "barked," which requires processing on the long-distance/nonadjacent dependency. This hierarchical syntactic structure reflects the complexity of human capacities of sequence processing (Petkov and ten Cate, 2020; Wilson et al., 2020). In contrast, simple sentence processing implicates mental operations on local dependency. For example, in the double-violation condition "警察交战骗局…" ("police fought the fraud") (Wang et al., 2015), the verb "交战" (fight) and the object "骗局" (fraud) are locally adjacent. Such local dependencies depict the adjacent collocation between word categories, which could be confounded with the more cognitive-general effect of "local preference" in syntactic processing. Specifically, local preference originates from the limitation of human cognitive resources, such that language users tend to integrate syntactically local or adjacent information as early as possible (Gibson, 1998). As such, the ERP components associated with local syntactic violations (e.g., ELAN) could indicate both word category violations and local preference violations. Moreover, a local syntactic violation may also be mixed with a violation of template matching, resulting in an unclear detection of the violation *via* the failure to build up a grammatical phrase (e.g., "the in," a determiner cannot be combined with a preposition to form a determiner phrase) or a mismatch with an *a priori* template (e.g., "the in" does not match the "determiner noun" template) (Friederici, 2017; Goucha et al., 2017). Therefore, the existence of local syntactic relations in simple sentences fails to provide optimal material for examining the interplay between syntactic and semantic processing. Examining this issue in complex sentences might overcome these limits and substantially advance our understanding of the interaction between these types of information as well as the human language faculty (Hauser et al., 2002).

In particular, syntactic complexity could be measured by integration and storage costs, instead of the mere sentence length (Gibson, 1998; Chesi and Moro, 2015). In specific, integration cost is concerned with the process of integrating syntactic categories, while storage cost is qualified by the number of the involved categories. Both integration and storage cost are impacted by locality. As such, sentences containing center-embedded subject relative clauses manifest great complexity and processing difficulties in light of their integration and storage cost, which also highlights the crucial role of word category information. Chinese complex sentences containing center-embedded subject relative clauses, such as "警察抓了偷电脑的小偷" ("Police caught the thief who stole a computer"), in which "小偷" is non-adjacently dependent with the main verb "抓" as its object and with the verb "偷" in the relative clause as the subject, implicate even higher complexities and more processing difficulties

(Hsiao and Gibson, 2003; Chen et al., 2008; Chen and Ning, 2008; Zhang and Yang, 2010; Zhou et al., 2010; Liu, 2011; Liu et al., 2011; Wang, 2011; He et al., 2012; Feng and Wang, 2013; Wang and Bing, 2013; Yan, 2018; Xu et al., 2019; Zhou et al., 2020). The current study therefore adopted these complex sentences with subject relative clauses center-embedded as experimental materials to highlight the interactive relationships between syntactic category information and semantics in Chinese sentence reading.

1.4. The interplay of syntactic and semantic processing in L2 learners

In addition to the language typological distinctions in the interplay between syntactic and semantic information, as manifested by the distinct patterns between morphologically rich Indo-European languages and morpho-syllabic languages, language proficiency might be another critical factor that affects the relationships between syntactic and semantic processes during language comprehension (Kotz, 2009). Zhang et al. (2010) concluded that language experience might affect the interplay between syntactic and semantic processes. A recent review (Niharika and Prema Rao, 2020) also proposed that language typological differences are closely associated with the language-specific brain correlates underlying syntactic processing. However, it is largely unknown to what extent language experience or background modulates this pattern in second language (L2) settings. For example, given adequate language exposure, do L2 learners employ the processing strategies from their first language (L1) or, alternatively, do they resemble native speakers of the target language? This issue would be more intriguing if learners' L1 and L2 manifested marked differences regarding linguistic typology.

More importantly, since the ability to encode syntactic information during complex sentence comprehension constitutes a crucial part of the human language capacity, it is critical to know whether and how L2 learners can acquire the native-like strategies of discerning syntactic and semantic information at the neuroscientific level. Specifically, the unified competition model (MacWhinney, 1997, 2005) proposes that L2 learners employ the cognitive resources and processing strategies from their L1 to deal with L1-L2 shared structures. In addition, given presumably adequate L2 proficiency, learners could develop native-like sensitivity to syntactic information and eventually achieve native-like attainment (following the convergence hypothesis, e.g., Steinhauer et al., 2009). Fromont et al. (2020) compared the ERP responses to L2 French sentences containing syntactic-category violations of native English speakers (with an intermediate level in L2 French) with those of native French speakers. Interestingly, L1 speakers and L2 learners showed differing electrophysiological patterns such that the L2 group only manifested an N400 effect in syntactic violations, while the L1 group displayed a biphasic N400-P600 effect.

However, few studies have examined whether L2 learners could develop the processing patterns of the target language regarding the interplay of syntactic and semantic information across typologically distinct languages. Moreover, there have been few investigations into the interplay between syntactic and semantic information among highly proficient L2 learners. By using comparatively simple structures (i.e., BEI structure), Yang (2012) found results supportive of an interactive model over a syntax-first model in Chinese comprehension for both Chinese L1 speakers and Chinese L2 learners (German L1 speakers). However, this study failed to make a direct comparison between the two groups. Additionally, it remains unclear whether this pattern extends to complex hierarchical sentences. Going beyond this, our study aimed to include high-proficiency Chinese L2 learners whose native language was Korean to compare their neurocognitive patterns with those of Chinese native speakers during complex sentence reading. As an agglutinative language, Korean morphology includes abundant word form changes to mark syntactic features whose diversity lies between German and Chinese (Zhang et al., 2011). For instance, Korean relative clauses are usually led by verb form changes, while Chinese relies on the functional word "的" (de, meaning: of) (Zhao, 2011). Specifically, employing Chinese L2 learners whose native languages are morphologically rich could effectively eliminate interference from their L1 processing strategies (e.g., semantic analysis preference). Furthermore, the inclusion of L2 learners of high proficiency could rule out the confounding effect of language incompetence. Through these manipulations, we aimed to provide critical insights into the interplay between syntactic and semantic information during Chinese complex sentence comprehension, thus drawing on a comparison of L1 versus L2.

1.5. Development of the experimental design

With regard to the experimental design, double-violation can be realized by altering either the sentential context (e.g., Hahne and Friederici, 2002; Frisch et al., 2004; Ye et al., 2006; Zhang et al., 2010; Wang X. et al., 2013) or the critical word (Yu and Zhang, 2008; Zhang et al., 2010, 2013; Wang S. et al., 2013; Lu, 2015; Wang et al., 2015; Yang et al., 2015; Li, 2016; Fromont et al., 2020). Sentential context alternation might elicit imbalanced ERP effects on the same critical word. For instance, Hahne and Friederici (2002) kept the critical word (e.g., *gegessen*, "eaten") unchanged across three violation conditions (i.e., double violations, semantic violation, and syntactic violation). The critical word was embedded in a "Verb + Preposition + Noun" structure in the double-violation condition but in a "Verb + Noun" structure in the other two conditions. Ye et al. (2006) placed the semantic violation after the "BA + Noun" construction, while syntactic violation and double violations appeared immediately after the preposition "BA." These contextual asymmetries might result in the so-called "spillover effect" on ERP signals, which would contaminate the ERP data locked to the critical words (Steinhauer and Drury, 2012; Zhang et al., 2013; Fromont et al., 2020). As the current study focused on the complex structure of Chinese sentences, we needed to keep the sentential contexts unchanged across conditions. To better evaluate the ERPs on critical words and avoid the spillover effect, this study manipulated violations by altering critical words.

The critical word alteration should take the processing differences of different word categories (e.g., Ns and *Vs*) into account. In particular, related Chinese studies have attempted to change the critical words from verbs to nouns in double violations (e.g., Yu and Zhang, 2008; Zhang et al., 2010, 2013; Lu, 2015; Yang et al., 2015; Li, 2016), which might involve the confounding effect of word category processing. For example, Ma et al. (2007) reported that Chinese nouns could elicit greater N400 amplitudes than Chinese verbs. In that case, we could not rule out the possibility that the greater negativities observed in these studies could stem from word category processing *per se* rather than semantic and syntactic violations. To resolve this confounder, Yu and Zhang (2008) performed a *post hoc* analysis comparing the negativity distribution of double violations on the scalp with that of word category processing, which revealed that ERP results were not contaminated by the word category processing difference. However, the scalp distributions of word category processing were retrieved from previous studies, which might not match the case of their critical words. Alternatively, some studies have included control sentences containing the critical words (Steinhauer and Drury, 2012; Fromont et al., 2020) as a reference to better evaluate the effects of syntactic and semantic violations. For example, Fromont et al. (2020) manipulated the prior context of the critical words while altering the critical words during French sentence comprehension. Unfortunately, this manipulation does not fit Chinese complex sentences, in which it is difficult to ensure that the experimental and control sentences are comparable. It is even more difficult to change the prior context of the critical words while keeping the structure unchanged in Chinese given the language typological difference between French and Chinese.

To resolve these limitations, the current study tried to eliminate the word category difference by two approaches. First, we altered the critical nouns in double violations into verbs instead of the other way around. In this case, the observed greater negativities elicited by double violations during the time-window of N400 than the other conditions (if any) would be more convincing for the existence of semantic processing. Additionally, we included a single-word reading session in which participants read identical critical words as the sentence comprehension session. By subtracting the ERP signals locked to the single words from those locked to the critical words in the experimental sentences, we hoped to better eliminate the confounding effects of word category processing differences and other physical properties of no interest.

1.6. Research aims and expectations

By tentatively developing the critical word alternation design by subtracting the ERP signals of processing single words *per se* from those of the same critical words, the current study aimed to examine the interplay between syntactic and semantic processing among Chinese native speakers and highly proficient L2 learners from a distant linguistic background when reading Chinese complex sentences containing center-embedded relative clauses, which should be highly dependent on word-category-based syntactic processes in linguistic theories. In particular, we are interested in whether Chinese complex sentences manifest a syntax-first pattern, which would be shown by the ELAN and the block effect. Crucially, we wanted to investigate whether Chinese L2 learners would acquire syntactic and semantic processing strategies at a relatively higher proficiency with regard to the interplay of syntactic and semantic processing during the ELAN, N400, and P600 time windows. If highly proficient Chinese L2 learners could exhibit a similar ERP pattern to that of L1 speakers, we could suggest that L2 proficiency might fill the gap in language typological differences to tune L2 reading performance.

2. Methods

2.1. Participants

A total of 42 adults were recruited, including 21 Mandarin-Chinese native speakers (6 males, 23.48 ± 2.91 years) and 21 highly proficient Chinese L2 learners whose native language was Korean (7 males, 23.29 ± 3.51 years). The sample size was consistent with existing related ERP studies (e.g., Yang, 2012; Wang S. et al., 2013; Wang et al., 2015). Specifically, the Chinese L2 learners from South Korea were year-four college students or postgraduates majoring in Chinese, all of whom passed the HSK (i.e., Hanyu Shuiping Kaoshi, a standardized Chinese proficiency test, ranging from bands 1 to 6) with band 5 or above and were verified as highly proficient learners by their instructors. All participants were right-handed with normal or corrected-to-normal vision and reported no reading difficulty. They all signed the consent form prior to the experiment and received a monetary reward afterward. This study was approved by the Ethics Committee of Beijing Normal University. Data from one L1 speaker and one L2 learner were excluded from the analyses due to excessive ERP artifacts, such as blinks.

Given the long-distance dependency between critical words and other related words in the experimental materials and the word-by-word presentation format in the current study, we needed to exclude the confounding effect from working memory capacity differences between the two groups. The working memory capacity measure was adapted from the automated operation span task by Unsworth et al. (2005). In this task, participants were asked to first judge the correctness of a hybrid math operation [e.g., (8–2) × 3 = 18?] and then to memorize a random English letter following this operation. As the number of operations increased, the participants needed to remember more letters. At the end of each trial, they were required to recall all the letters in each trial in the order of their appearance. An independent-sample t test showed that L1 speakers [$N = 20$, ACC (accuracy) = 64.15% ± 6.78%] and L2 learners ($N = 20$, ACC = 61.55% ± 7.14%) yielded no significant group difference in working memory capacity, $t(38) = -1.18$, $p = 0.245$.

2.2. Materials

There were 120 experimental sentences in total, with 40 in each condition (NORM: semantically and syntactically normal sentences, SEM: sentence with semantic violation, and SEM + SYN: sentence with double syntactic-semantic violations; see Table 1 for examples). In addition, to counterbalance the number of correct and violated sentences, 40 filler sentences were included (see also Wang et al., 2015). Each experimental sentence contained a center-embedded relative clause, such as "小张拿着切水果的小刀过来了。." In the SEM condition "小张拿着切水果的钢琴过来了。," even though "钢琴" (piano) is a noun without syntactic violation, the sentence is semantically violated because "piano" could never be used to "cut the fruit." In addition, in the SEM + SYN condition, "举办" (hold) is a verb resulting in both syntactic and semantic violations because "hold" could neither "cut the fruit" nor "be taken." The filler sentences were similar to the experimental sentences in length but did not contain complex relative clauses, such as "小张拿走了孩子的玩具很高兴."
The rationale of designing such filler sentences is as following. In light

TABLE 1 Examples of experimental conditions and fillers.

Condition	Example sentence
NORM	小张 | 拿着 | 切 | 水果 | 的 |小刀 | 过来了。 Zhang | was taking | cut | fruit | de (*functional word*) |**knife** | came over. (Zhang was coming over with the **knife** which is used to cut fruits.)
SEM	小张 | 拿着 | 切 | 水果 | 的 |钢琴 | 过来了。 Zhang | was taking | cut | fruit | de (*functional word*) |**piano** | came over. (Zhang was coming over with the **piano** which is used to cut fruits.)
SEM + SYN	小张 | 拿着 | 切 | 水果 | 的 |举办 | 过来了。 Zhang | was taking | cut | fruit | de (*functional word*) |**hold** | came over. (Zhang was coming over with the **hold** which is used to cut fruits.)
Filler	小张 | 拿 | 走了 | 孩子 | 的 |玩具 | 很高兴。 Zhang | took | away | children | de (*functional word*) |**toy** | was very happy. (Zhang was very happy to take away the children's toy.)

NORM: semantically and syntactically normal. SEM: semantic violation. SEM + SYN: double violations. Examples are given in Chinese, with English literal glosses and translations below. The critical words are in bold.

TABLE 2 The number of strokes and word frequency of critical words in different conditions.

Condition	Number of the strokes	Frequency
NORM	13.800 ± 4.697	3.228 ± 0.646
SEM	14.200 ± 3.451	3.006 ± 0.539
SEM + SYN	15.150 ± 4.240	3.256 ± 0.571

of existing studies (e.g., Wang S. et al., 2013; Wang et al., 2015), the fillers should meet the following criteria. First, the purpose of filler was to balance the correct and incorrect sentences. Second, to avoid the participants' strategic responses, the fillers should be comparable with the experimental sentences regarding the superficial linguistic features (e.g., the number of nouns/verbs, sentence length). Thus, the filler sentence structure and the complex sentence structure in the present study share a similar syntactic frame: "N + V + V + N + de + N…" The correspondence of the overall structure between fillers and experimental sentences was not obligatory.

As the SEM and SEM + SYN conditions are primarily concerned with linguistic manipulations, three experts of Chinese linguistics verified the semantic and syntactic constraints of these sentences with perfect consistency, that is, all of them agreed that all the SEM sentences contained semantic violations, all the SEM + SYN sentences contained both syntactic and semantic violations, and that all the Norm and filler sentences were both semantically and syntactically acceptable. In addition, the NORM sentences as well as the filler sentences were randomly mixed with the SEM and the SEM + SYN sentences, and a group of Chinese native speakers ($N = 10$) who did not participate in the formal experiment were asked to determine whether these sentences were natural and thus acceptable to them or not by Yes/No responses through a questionnaire (without time limits). All NORM and filler sentences (100% correctly responded as "Yes") could be well distinguished from the SEM and the SEM + SYN sentences (100% correctly responded as "No") for each participant. Therefore, the results showed robust consistency across the participants and explanatory power regarding the validity of the materials.

The critical words in the experimental sentences were primarily selected from the glossary of HSK level-5 or below, with which L2 learners were asked to familiarize themselves before the experiment. The frequency and the number of strokes of the critical words were matched across three conditions (frequency information retrieved from Cai and Brysbaert, 2010; see Table 2). One-way ANOVA showed that there were no significant differences regarding the number of strokes and word frequency of the critical words among different conditions, $Fs(2, 117) < 2.18$, $ps > 0.05$. In particular, the critical words did not appear at the end of the sentences to avoid ERP contamination from the "wrap-up effect" at sentence-final positions (Hagoort, 2003). Except for the critical words, the sentential contexts were identical across conditions. Specifically, all experimental sentences used "小张" (Zhang) as the subject to avoid possible expectations caused by the sentence subject. Additionally, the frequency of the nouns following the functional word "的" (de) in the NORM condition was balanced in terms of the argument roles as "tool" (e.g., "knife") or "agent" (e.g., "child"). An additional single word reading task was conducted to provide a baseline for the ERP correlates underlying the critical words. These materials contained only nouns and verbs (45 words for each category), which were identical to the critical words adopted in the experimental sentences.

2.3. Procedure

Word reading and sentence comprehension tasks were conducted for both groups, while behavioral performance and EEG signals were recorded simultaneously. L2 speakers received the list of critical words 2 days before the experiments and were asked to consolidate the memory of these words. In addition, they were required to review these words again prior to the formal experiment until they reported total familiarity with the meaning and usage of the words.

In the formal experiment, participants were seated approximately 60–70 cm from the computer monitor in a fully shielded laboratory. Participants performed the word reading task before the sentence comprehension session. In the word reading task, each trial began with the presentation of a red fixation at the center of the screen for approximately 300 ms, followed by a blank screen for 200 ms, the critical word for 500 ms, and then a 1,000 ms blank screen. The participants were asked to familiarize themselves with the words again in this session. To ensure the participants' engagement, they were informed that their real-time mental activities would be monitored by the device. To note, first, the objective of asking participants (especially for the L2 learners) to familiarize themselves with the critical words before the formal test was to ensure that the processing differences of the critical words between the conditions during sentence reading could not be ascribed to the mere familiarity effects. Second, more critically, for each condition, a critical word could appear in all the tasks of word familiarization, word reading, and of sentence comprehension (see below). Therefore, the potential preview effect (i.e., effect of seeing the "critical words" multiple times before the formal study including both single-word reading and sentence comprehension tasks) could be canceled out in the comparison between different conditions if there were any for each stage.

The sentence comprehension task consisted of three blocks with 54, 54, and 52 sentences in each block. The order of the three blocks was randomized and there was a two-minute interval between every two consecutive blocks. In each block, three conditions of experimental sentences and filler sentences were presented in a pseudorandom order, such that the sentences of the same condition would not appear in more than three times consecutively. Each trial began with a red fixation cross at the center of the screen for approximately 300 ms, followed by sentences presented word by word, e.g., "小张 | 拿着 | 切 | 水果 | 的 | 小刀 | 过来了。" [Zhang | was taking | cut | fruit | de (functional word) | knife | came over.]. In particular, double-character words were presented for 500 ms, while single-character words lasted for 400 ms. Each word was followed by a 300 ms blank screen, while a Chinese period ("。") indicated the end of the sentence. After the sentence disappeared, the participants were required to judge whether the sentence was acceptable within 3,000 ms, which was followed by a blank screen lasting for 1,000 ms. The whole experiment took approximately 1.5 h for each participant.

2.4. Data recordings and analyses

Electroencephalogram (EEG) was recorded using a 32-channel (Ag-AgCl) NeuroScan system (NeuroScan Inc.) following the 10–20 system convention. For online recordings, the reference electrode was placed at the right mastoid (M2). Vertical electrooculogram (VEOG) was recorded from electrodes placed above and below the left eye, while horizontal EOG (HEOG) was recorded from electrodes at the outer cantus of each eye. The data were digitized at 1 kHz and amplified with a bandpass filter of 0.05–100 Hz. The impedance of each channel was maintained below 5 kΩ.

Off-line signal processing was carried out using Scan software (NeuroScan Inc.). The reference was converted to the averaged voltages of the bilateral mastoids (M1 and M2). EEG data were first adjusted by eliminating artifacts using the DC method and regression analysis and then segmented into 1,000 ms epochs, including a 200-ms prestimulus baseline and an 800-ms poststimulus. Eye blinks and other artifacts exceeding ±100 μV were rejected. ERPs were averaged across conditions for both word reading and sentence comprehension tasks and then filtered again with a low bandpass filter of 30 Hz (24 dB).

The ERP components induced by nouns and verbs in the word reading task (see Figure 1) were subtracted from the ERP components induced by the corresponding critical words in different types of experimental sentences (see Figure 2), so the word reading results were not included in the later statistical analyses (see Figure 1). As the L1 and L2 groups showed differing time windows of ERP components (i.e., N400 and P600) in the sentence comprehension task (see Figure 3), we first conducted statistical analysis within each group and then compared the components of interest between groups. For the within-group analyses, based on the visual inspection of brain topographies (see Figure 3) and previous studies (e.g., Yu and Zhang, 2008; Zhang et al., 2010; Wang et al., 2015), we selected three time windows of 100–200 ms, 250–430 ms, and 430–600 ms for the Chinese L1 group and 100–200 ms, 300–500 ms and 600–750 ms for L2 learners. They were used to denote ELAN, N400 (or larger negativities), and P600 (or late positivities), respectively. In light of existing ERP studies, we defined 7 regions of interest (ROIs) for statistical analyses: midline (Fz, FCz, Cz, CPz, and Pz), left anterior (F3, FC3, F7, and FT7), left central (C3 and T7), left posterior (CP3, P3, TP7, and P7), right anterior (F4, FC4, F8, and FT8), right central (C4 and T8), and right posterior (CP4, P4, TP9, and P8). As such, we first conducted a 3-way ANOVA on ELAN data at the midline sites, on which the L1 and L2 groups shared the same time window, with the condition (NORM, SEM, and SYN + SEM), electrode (Fz, FCz, Cz, CPz, and Pz), and group (L1 and L2) as factors. For the lateral regions, a 3-way ANOVA was conducted with condition, ROIs (left anterior, left central, left posterior, right anterior, right central, and right posterior), and group as factors. Then, identical analyses were conducted on N400 and P600 data, except that the group factor was removed due to the differing time windows across the two groups. Furthermore, we took a narrow-window approach (e.g., Sanmiguel et al., 2013; Ghani et al., 2020) by selecting the data including 10 ms before and 10 ms after the N400/P600 peaks since the wider time window mentioned above might overwhelm the subtle differences across conditions (as shown in Figure 4). Furthermore, representative electrodes were also selected based on the topography (Figure 4) to conduct the independent-sample t test between the experimental condition (SEM + SYN - SEM) and "baseline (0)." Importantly, narrow-window analysis provided a viable tool for group comparisons regarding N400 and P600 effects. Furthermore, *post hoc* power analysis using G*power (Faul et al., 2007) was

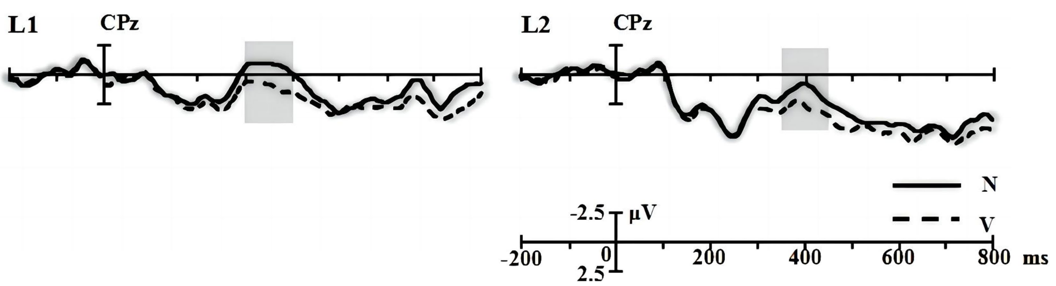


FIGURE 1
The averaged waveforms elicited by nouns (N) and verbs (V) in the word reading task. The grey rectangle marked the greatest difference between the two categories.

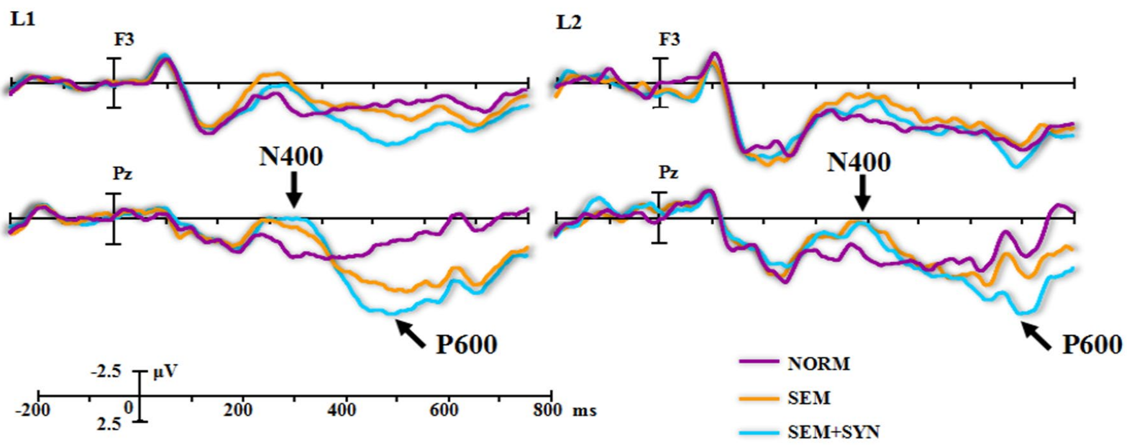


FIGURE 2
The original brain waveforms locked to the critical words embedded in the experimental sentences.

performed for *main* results so as to evaluate the overall statistic power of the present study.

3. Results

3.1. Behavioral results

The accuracy rate (ACC)[4] and reaction time (RT) data of L1 and L2 speakers across the three conditions are shown in Table 3 and Figure 5. ACC data showed a significant main effect of condition, $F(2, 76) = 17.91$, $p < 0.0005$, $\eta_p^2 = 0.320$, *power* = 0.97. Pairwise comparisons with Bonferroni corrections revealed that the ACC of the SEM + SYN condition was significantly higher than that of the SEM and NORM conditions ($ps < 0.005$), while the ACC of the SEM condition was significantly higher than that of the NORM condition ($p < 0.005$). The main effect of group was significant, $F(1, 38) = 26.77$, $p < 0.001$, $\eta_p^2 = 0.413$, *power* > 0.99, such that L1 speakers obtained significantly higher accuracy than L2 learners. The interaction between the two factors was not significant, $F(2, 76) = 2.30$, $p = 0.132$. In terms of RT data, the main effect of condition was significant [$F(2, 76) = 58.50$, $p < 0.0005$, $\eta_p^2 = 0.606$, *power* > 0.99]. Pairwise comparison showed that the RT of SEM + SYN was significantly shorter than that of SEM and NORM ($ps < 0.01$), while the responses to SEM were significantly faster than those to NORM ($p < 0.001$). The main effect of group was significant [$F(1, 38) = 8.59$, $p < 0.01$, $\eta_p^2 = 0.184$, *power* = 0.83]. The RT of L1 Chinese speakers was significantly shorter than that of L2 Chinese speakers. The interaction between the two factors above was not significant [$F(2, 76) = 1.58$, $p = 0.212$].

3.2. ERP results

As seen from the original waveforms of the three conditions across the two groups (Figure 2), the SEM + SYN condition elicited N400- and P600-like components, while ELAN was not detected at the frontal sites (F3 as the representative electrode) for both groups. In addition, single word reading results (Figure 1) showed that nouns induced more negative waves than verbs in the 300–500 ms time window for both groups. We then subtracted the waveforms of word categories from the original waveforms of the critical words embedded in sentential contexts (Figure 3) to better evaluate the ERP modulation of syntactic and semantic processing.

3.2.1 ELAN results

ELAN effects were examined at the frontal sites in the time window of 100–200 ms. At the midline sites, there was no significant effect for condition [$F(2, 76) = 1.40$, $p = 0.253$]. At the lateral sites, the main effect of condition was also not significant [$F(2, 76) = 2.79$, $p = 0.081$]. Thus, for both groups, the SEM + SYN condition did not induce an ELAN effect in the early time window.

3.2.2 N400 results

As shown in Figure 6, L1 data revealed a significant condition effect at both the midline and lateral sites [$Fs(2, 38) > 11.90$, $ps < 0.001$, $\eta_p^2s > 0.385$, *power* > 0.99] in the N400 time window (250–430 ms). Pairwise comparisons showed that SEM + SYN (midline: 0.407 ± 4.537 μV; lateral: 0.126 ± 2.406 μV) and SEM (midline: 1.049 ± 3.653 μV; lateral: 1.035 ± 2.140 μV) elicited more negative waves than NORM (midline: 2.953 ± 3.470 μV; lateral: 2.142 ± 2.101 μV) at both the midline and lateral sites ($ps < 0.05$), while no difference was detected between SEM + SYN and SEM. At the midline sites for L1 speakers, the main effect of electrode was also significant [$F(4, 76) = 4.57$, $p < 0.05$, $\eta_p^2 = 0.194$, *power* > 0.99] such that FCz (1.402 ± 2.168 μV) yielded more negative waves than Fz (2.168 ± 4.054 μV; $p < 0.05$). In general, the negativities in this time window showed a central-posterior distribution in the midline region. The interaction between electrode and condition was not significant [$F(8, 152) = 0.73$, $p = 0.524$]. For the lateral regions, the ROI effect of L1 speakers was not significant [$F(5, 95) = 1.00$, $p = 0.379$], while its interaction with condition was significant [$F(10, 190) = 2.78$, $p < 0.05$, $\eta_p^2 = 0.251$, *power* > 0.99]. Simple effect analyses showed comparable ERP patterns across all lateral ROIs as midline sites [$Fs(2, 38) \geq 6.30$, $ps < 0.005$].

4 ACC reflects the correct acceptance of NORM condition and the correct rejection of SEM and SEM+SYN conditions.

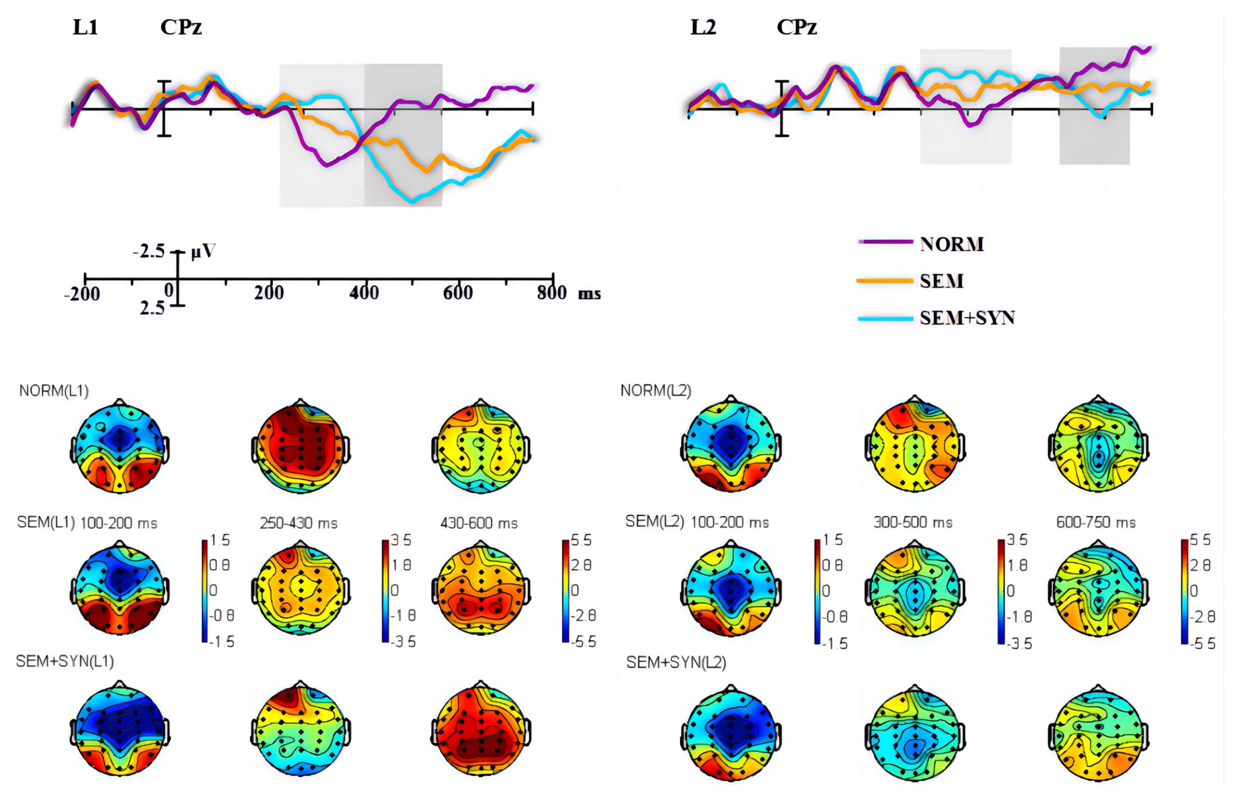


FIGURE 3
The grand-averaged difference waveforms and topographies. Light and dark gray rectangles marked the most obvious differences across conditions in the middle and late time-windows, respectively.

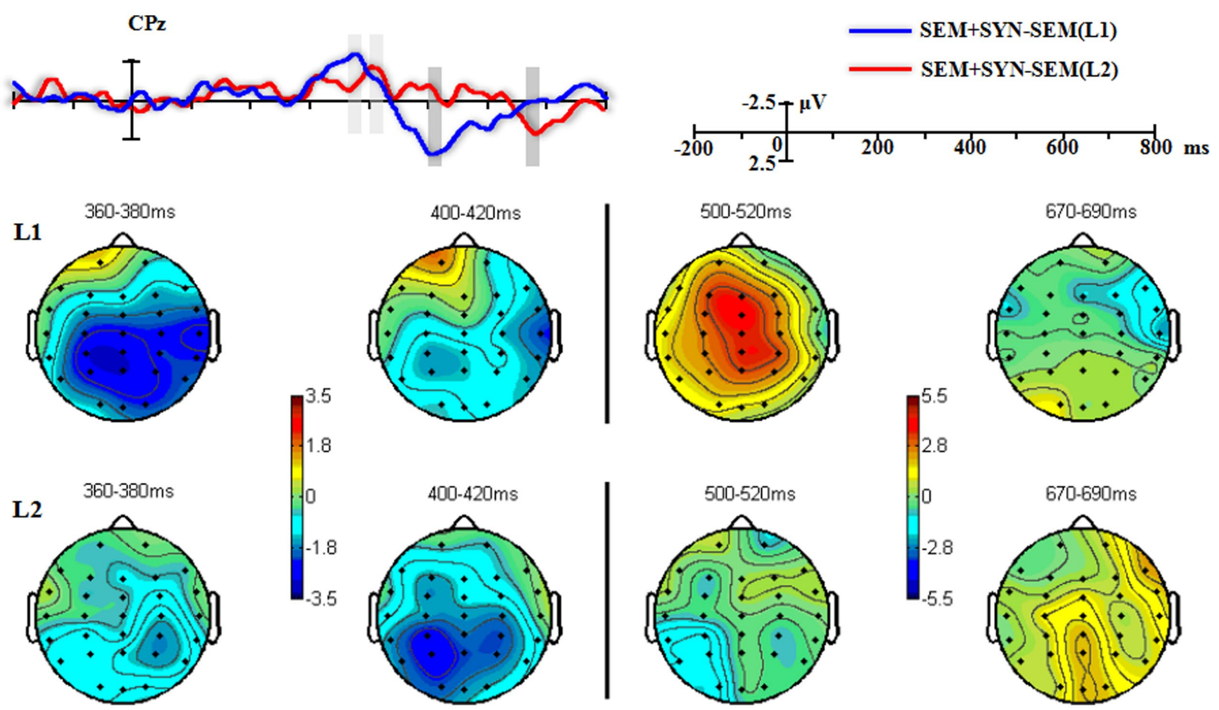


FIGURE 4
The grand-averaged difference waveforms and topographies of "SEM+SYN - SEM." Light and dark gray rectangles marked the most obvious differences across conditions in the middle and late time-windows, respectively.

In the time window of 300–500 ms, L2 data manifested similar patterns as L1 data. There was a significant condition effect for both midline [$F(2, 38) = 6.36$, $p < 0.05$, $\eta_p^2 = 0.251$, *power* > 0.99; see Figure 6] and lateral sites [$F(2, 38) = 7.34$, $p < 0.01$, $\eta_p^2 = 0.279$, *power* > 0.99] such that the difference between the SEM (midline: −0.737 ± 2.688 μV; lateral: −0.026 ± 1.987 μV) and SEM + SYN (midline: −1.624 ± 2.574 μV; lateral: −0.851 ± 1.854 μV) conditions was not significant, while the SEM and SEM + SYN conditions yielded more negative waves than the NORM condition (midline: 0.411 ± 3.129 μV; lateral: 0.961 ± 2.771 μV; *ps* < 0.05). The main effect of electrode was also significant at midline sites [$F(4, 76) = 4.57$, $p < 0.05$, $\eta_p^2 = 0.194$, *power* > 0.99]. FCz (−0.720 ± 2.639 μV) yielded more negative waves than Fz (0.388 ± 2.764 μV; $p < 0.05$). In general, the negative waves in this time window showed a centro-posterior distribution in the midline region. The interaction between distribution and condition was not significant [$F(8, 152) = 0.73$, $p = 0.524$]. No other significant main effect or interation was detected at lateral ROIs.

To eliminate semantic contaminations from the double violations, we analyzed the difference wave between SEM + SYN and SEM in light of a narrower window of 20 ms (i.e., 10 ms prior to the peak and 10 ms after the peak, see Figure 4). For L1 speakers, there were wide negativities in the posterior sites in the time window of 360–380 ms. Repeated-measures ANOVA with the experimental condition (SEM + SYN - SEM vs. baseline "0") and electrode (CPz, Pz, CP3, CP4, P3, P4) as factors revealed a significant effect of the experimental condition (SEM + SYN – SEM: −2.621 ± 3.760 μV), $F(1, 19) = 9.72$, $p < 0.01$, $\eta_p^2 = 0.338$, *power* > 0.99, while no other significant main effect or interaction was detected [$Fs(5, 95) \leq 0.78$, $ps > 0.05$]. Double violations elicited significantly greater negativities, which were widely distributed at bilateral centro-parietal regions. For L2 learners, SEM + SYN elicited more obvious negativities than SEM in 400–420 ms at centro-parietal sites (Figure 4). A similar repeated-measures ANOVA was conducted on representative sites (Cz, CPz, Pz, C3, C4, CP3, CP4, P3, P4). There was a significant effect of experimental condition [$F(1, 19) = 19.62$, $p < 0.0005$, $\eta_p^2 = 0.506$, *power* > 0.99] such that SEM + SYN was associated with significantly greater N400 than SEM (SEM + SYN – SEM: $M = -2.204$ μV, $SD = 2.937$ μV). No other significant main effect or interaction was identified.

We further analyzed the group difference in N400 difference waves (SEM + SYN - SEM) by conducting a 3-way repeated-measures ANOVA with electrodes (CPz, Pz, CP3, CP4, P3, P4), time window (360–380 ms, 400–420 ms), and group (L1 vs. L2) as factors (Figure 7). Only a significant interaction between time window and group was detected, $F(1, 38) = 11.78$, $p < 0.005$, $\eta_p^2 = 0.237$, *power* > 0.99, such that the N400 effect between two time windows was significant for both groups [$Fs(1, 38) \geq 4.40$, $ps < 0.05$]. In addition, N400 peak results averaged from the six centro-parietal sites (360–380 ms for L1 and 400–420 ms for L2) showed a significant group effect on peak latency [$t(38) = -18.27$, $p < 0.0005$; see Figure 8], while there was no significant effect on peak amplitudes across groups [$t(38) = -0.39$, $p = 0.698$; see Figure 8]. In particular, the latency of L1 speakers (372.592 ± 6.138 ms) was significantly earlier than that of L2 learners (408.975 ± 6.453 ms).

3.2.3 P600 results

At midline sites, the 2-way ANOVA yielded a significant main condition effect [$F(2, 38) = 18.05$, $p < 0.0005$, $\eta_p^2 = 0.487$, *power* > 0.99; see Figure 9]. The following comparisons showed that SEM + SYN (4.964 ± 4.462 μV) induced more positive waves than SEM (2.768 ± 4.200 μV), while SEM induced more positive waves than NORM (0.364 ± 3.067 μV; *ps* <0.05). The main effect of electrode was not significant [$F(4, 76) = 1.98$, $p = 0.170$]. The interaction between condition and electrode was significant [$F(8, 152) = 11.18$, $p < 0.001$, $\eta_p^2 = 0.371$, *power* > 0.99]. Simple effect analyses showed that the condition effect was significant at all electrodes [$Fs(2, 38) \geq 3.72$, $ps < 0.05$]. At lateral ROIs, the main effect of condition was significant [$F(2, 38) = 11.79$, $p < 0.0005$, $\eta_p^2 = 0.383$, *power* > 0.99] such that SEM + SYN (3.787 ± 2.896 μV) and SEM (2.800 ± 2.856 μV) elicited more positive waves than NORM (1.046 ± 2.049 μV; *ps* < 0.01), while no difference between the two violation conditions was detected ($p = 0.305$). The main effect of ROI was not significant [$F(5, 95) = 1.77$, $p = 0.191$], but its interaction with condition was significant [$F(10, 190) = 5.36$, $p < 0.005$, $\eta_p^2 = 0.220$, *power* > 0.99]. In

TABLE 3 Behavioral performance of the two groups.

Group	Condition	ACC	RT (ms)
L1 (*N* = 20)	NORM	0.93 ± 0.06	780.31 ± 246.73
	SEM	0.96 ± 0.03	627.84 ± 212.15
	SEM + SYN	0.98 ± 0.04	564.46 ± 195.70
L2 (*N* = 20)	NORM	0.77 ± 0.14	941.04 ± 206.59
	SEM	0.86 ± 0.11	786.36 ± 175.19
	SEM + SYN	0.89 ± 0.10	781.09 ± 180.70

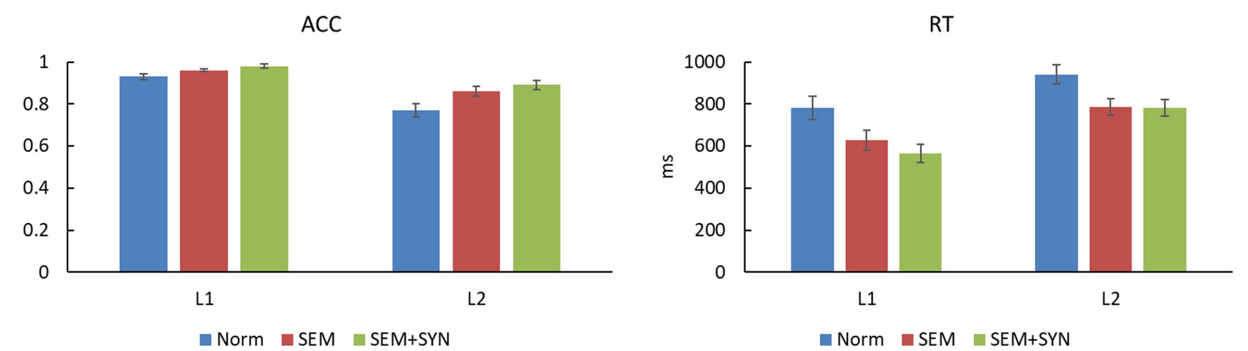


FIGURE 5
The ACCs and RTs of behavioral results (error bars indicate standard errors).

10.3389/fpsyg.2023.1143062

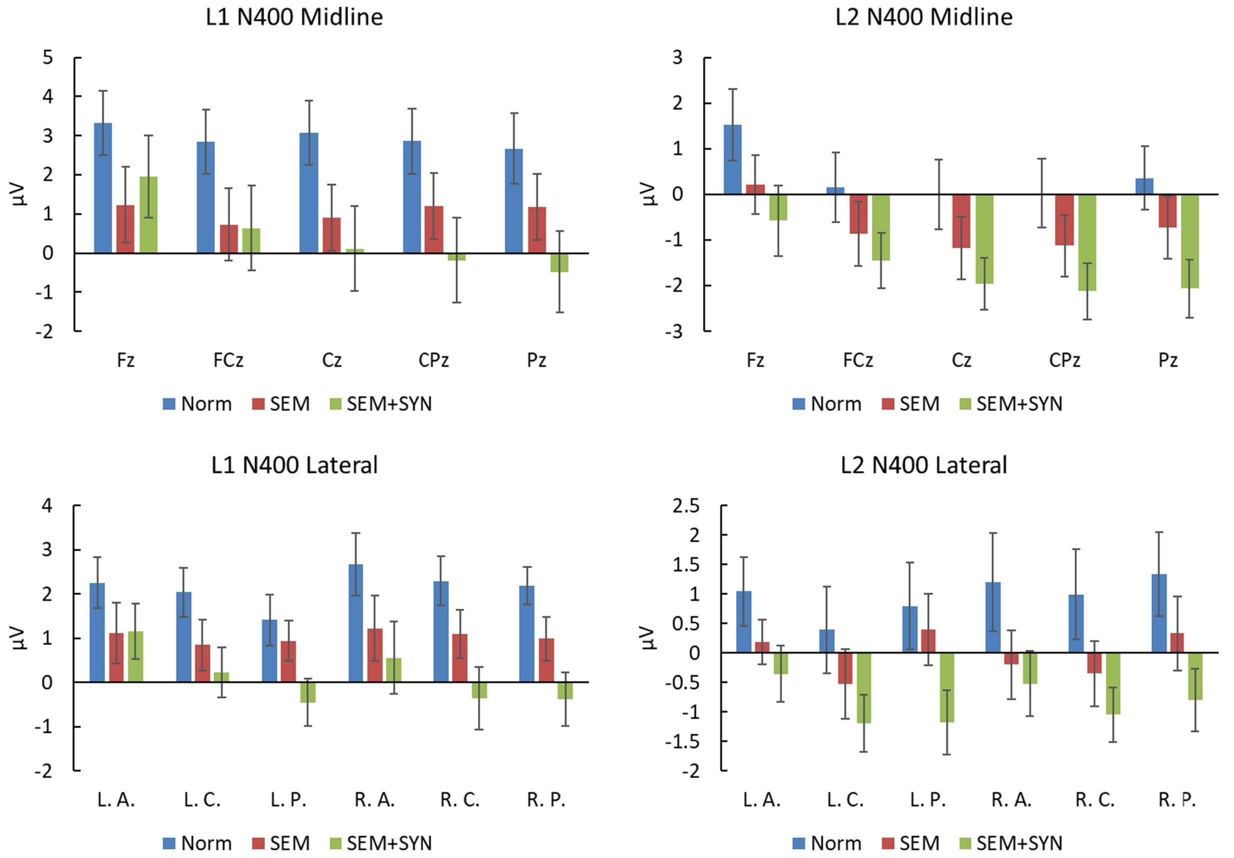


FIGURE 6
Amplitudes at midline electrodes and lateral ROIs in the N400 time window (250–430ms for L1 and 300–500ms for L2). L. A.: left anterior; L. C.: left central; L. P.: left posterior; R. A.: right anterior; R. C.: right central; R. P.: right posterior.

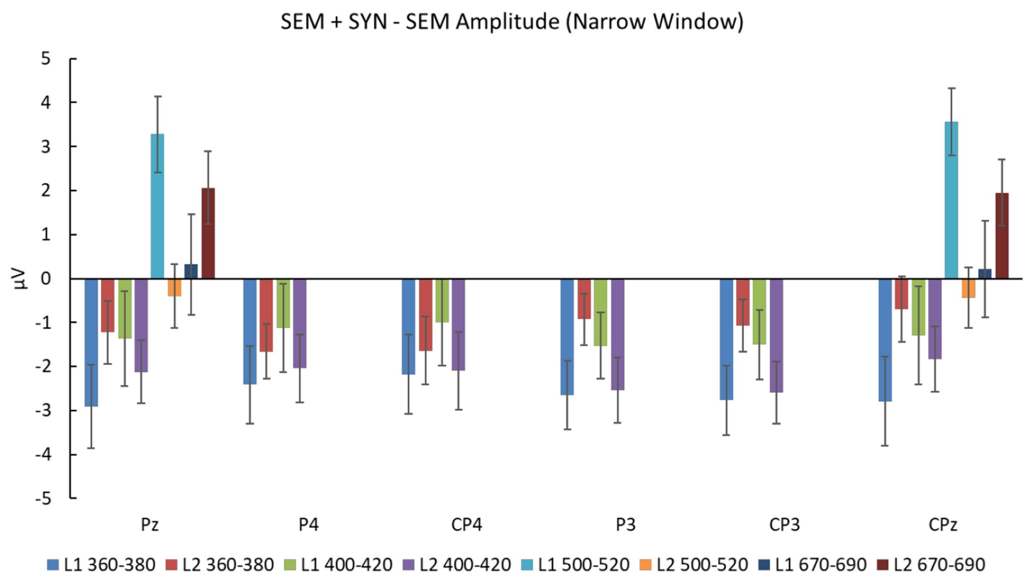


FIGURE 7
SEM+SYN − SEM amplitudes at selected electrodes for the chosen time window (360–380 and 400–420ms for N400 narrow window analyses at Pz, P4, CP4, P3, CP3, and CPz; 500–520 and 670–690ms for P600 narrow window analyses at Pz and CPz).

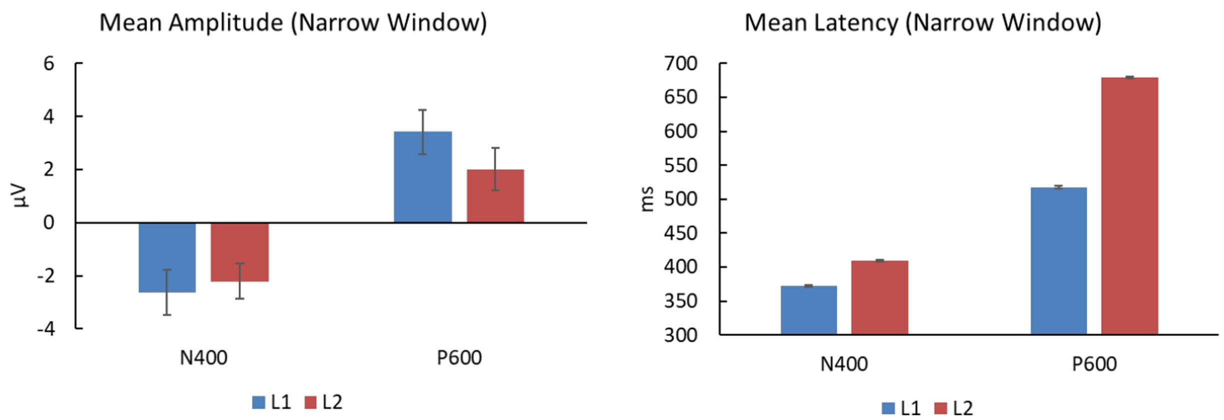


FIGURE 8
Mean amplitudes and latency averaged from selected electrodes for narrow-window analyses (360–380ms for L1 and 400–420ms for L2 in N400 time window; 500–520ms for L1 and 670–690ms for L2 in P600 time window).

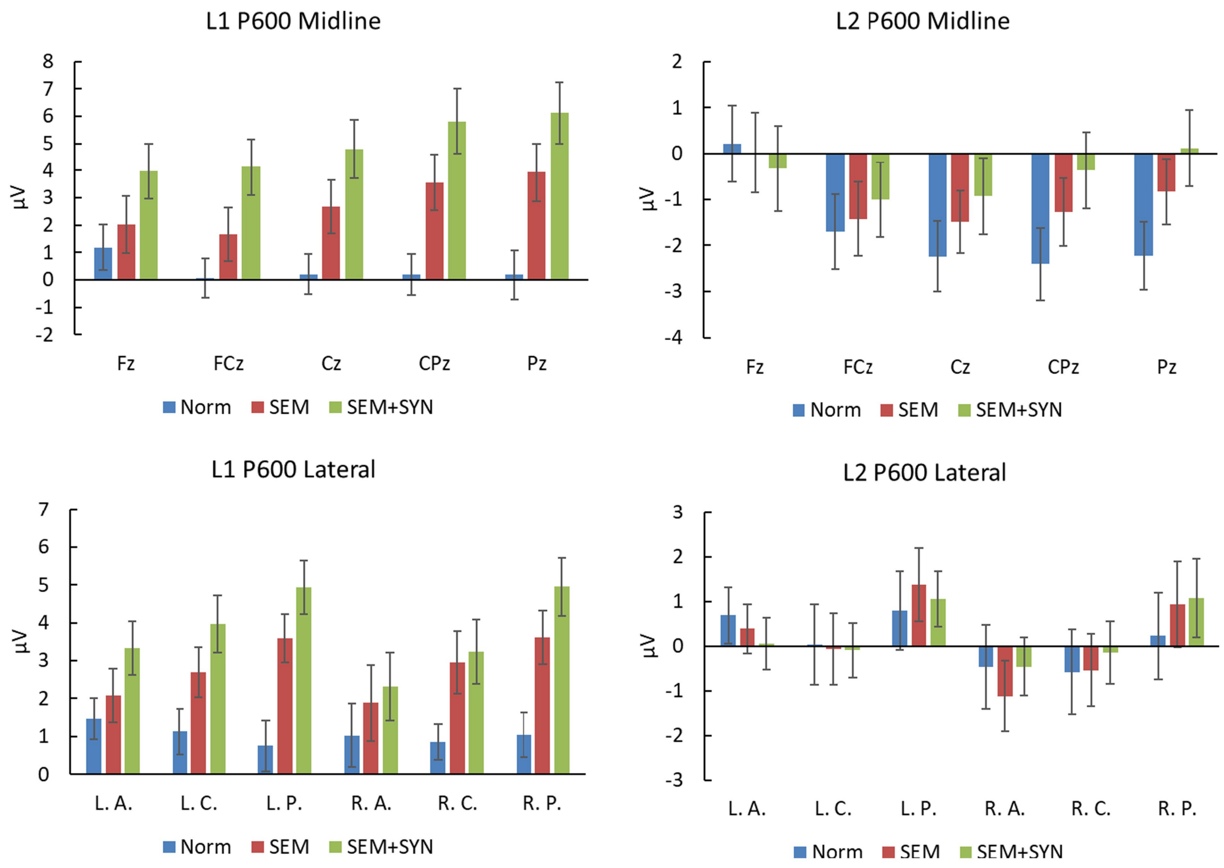


FIGURE 9
Amplitudes at midline electrodes and lateral ROIs in P600 time window (430–600ms for L1 and 600–750ms for L2). L. A.: left anterior; L. C.: left central; L. P.: left posterior; R. A.: right anterior; R. C.: right central; R. P.: right posterior.

particular, the condition difference was significant at all ROIs [$Fs(2, 38) \geq 4.36$, $ps < 0.05$], except at the right anterior region [$F(2, 38) = 1.36$, $p = 0.268$].

As seen in Figure 9, the P600 results in the L2 group manifested a distinct pattern from L1 speakers. At midline sites, the main effects of condition [$F(2, 38) = 1.55$, $p = 0.230$] and electrode [$F(4, 76) = 3.09$, $p = 0.077$] were not significant, while the interaction between them was significant [$F(8, 152) = 6.23$, $p < 0.005$, $\eta_p^2 = 0.247$, $power > 0.99$]. Simple effect analyses showed that condition effects were significant at Pz and CPz [$Fs(2, 38) \geq 4.175$, $ps < 0.05$] such that SEM + SYN and SEM elicited more positive waves than NORM ($ps > 0.05$), while no difference between the two violation conditions was detected

($p = 0.305$). At lateral ROIs, neither a significant main effect nor a significant interaction was identified.

Narrow-window analysis results based on difference waves in the L1 group (500–520 ms) showed wide positivities in bilateral centro-parietal regions (Figure 4). Repeated-measures ANOVA with the experimental condition (SEM + SYN - SEM vs. baseline "0") and electrode (CPz, Pz, CP3, CP4, P3, P4) as factors revealed a significant effect of the experimental condition, $F(1, 19) = 17.45$, $p < 0.01$, $\eta_p^2 = 0.479$, *power* > 0.99, such that SEM + SYN elicited significantly greater P600 than SEM (SEM + SYN – SEM: 3.417 ± 3.674 μV). There was no other significant main effect or interaction. For L2 learners, identical analysis was performed on CPz and Pz in the time window of 670–690 ms. There was a significant main effect of experimental condition, $F(1, 19) = 6.23$, $p < 0.05$, $\eta_p^2 = 0.247$, *power* > 0.99. SEM + SYN elicited greater P600 than SEM (SEM + SYN – SEM: $M = 2.008 \pm 3.597$ μV). The main effect of electrode and the interaction were not significant, $Fs(2, 38) < 0.22$, $ps > 0.05$.

As shown in Figure 4, the L1 and L2 groups produced obvious positivities associated with SEM + SYN – SEM at the time windows of 500–520 ms and 670–690 ms, respectively. We therefore examined the group difference at representative sites CPz and Pz (Figure 7). Repeated-measures ANOVA revealed a significant interaction only between time window and group [$F(1, 38) = 22.164$, $p < 0.0005$, $\eta_p^2 = 0.368$, *power* > 0.99]. Specifically, the group effect was significant only in the time window of 500–520 ms, $F(1, 38) = 12.45$, $p < 0.001$, while the time-window effect was significant for both groups, $Fs(1, 38) \geq 4.66$, $ps < 0.05$. Averaged results on P600 peaks across two representative sites showed comparable peak values for the two groups [$t(38) = 1.23$, $p = 0.228$], while the L1 group (517.425 ± 7.281 ms) showed significantly shorter peak latency than the L2 group (679.125 ± 7.527 ms), $t(38) = -69.06$, $p < 0.0005$ (Figure 8).

In addition, to examine the lateralization of ERP components, we subsequently averaged the lateral electrodes on left (F3, FC3, F7, FT7, C3, T7, CP3, P3, TP7, and P7) and right (F4, FC4, F8, FT8, C4, T8, CP4, P4, TP9, and P8) hemisphere and conducted a 2-way ANOVA on N400 and P600 data for L1 and L2 groups respectively, with condition (NORM, SEM, and SEM + SYN) and hemisphere (left and right) as factors. The results were shown in Supplementary material.

4. Discussion

Drawing on behavioral and ERP techniques, the current study investigated the relationship between syntactic and semantic processing when reading Chinese complex sentences with relative clauses center-embedded among L1 speakers and highly proficient L2 learners whose native language was Korean. Our findings showed that L1 speakers and L2 learners manifested a consistent behavioral and electrophysiological pattern of highly interactive syntactic and semantic processing during Chinese complex sentence reading.

The behavioral performance showed that SEM + SYN yielded higher ACC than SEM and NORM, while NORM was associated with the lowest accuracy. Likewise, SEM + SYN showed the fastest reactions, followed by SYN and NORM. Importantly, the behavioral patterns of the L1 and L2 groups were generally consistent. In particular, SEM + SYN involved both syntactic and semantic violations, which made error detection easier and faster than typical sentences. Although SEM involved semantic violations, its syntactic information remained correct, which led to relatively better recognition performance than double violations. In contrast, L1 and L2 speakers needed to make use of all the information available until all contents of the sentence were integrated when reading the semantically and syntactically normal sentences, which resulted in the longest reaction time and the lowest accuracy.

Furthermore, the ERP results provided more nuanced insights into the interaction between syntactic and semantic information in L1 and L2 Chinese complex sentence reading. The first important ERP finding from the current study was that L1 and L2 Chinese complex sentences with center-embedded relative clauses did not elicit the ELAN effect in the frontal sites from the double violations, while ELAN is an index of initial syntactic processing. This finding was consistent with the ERP patterns obtained from the double violations in simpler Chinese syntactic structures including BA, BEI, and SVO constructions (Yu and Zhang, 2008; Zhang et al., 2010; Wang S. et al., 2013; Wang et al., 2015; Zeng et al., 2020), where ELAN was also not identified. Those evidence could collectively suggest that syntactic processing is not the prerequisite for semantic processing in Chinese sentence comprehension regardless of structural complexities. In particular, ELAN has been recognized as an important index of temporal priority for syntactic processing and automaticity for local structure building (Friederici et al., 1993; Hahne and Friederici, 2002; Hagoort and Indefrey, 2014). However, ELAN is susceptible to various experimental manipulations. Steinhauer and Drury (2012) concluded that ELAN might be associated with a higher distribution probability of some stimuli (e.g., affix), the asymmetry of the precritical-word context with unchanged critical words, the "spillover effect," and the "offset effect." For instance, Fromont et al. (2016) created the French word category violation within a constant sentential context by using the homophone of definite articles "le/la/les" (equivalent to "the" in English) and accusative attachments "le/la/les" (equivalent to "him or her/them" in English), in which ELAN was not detected. Furthermore, Fromont et al. (2020) attempted to mitigate the interference from critical word alternation by changing presentence contexts and the type of critical words in the experimental sentences. Nevertheless, no ELAN effect was identified. Likewise, studies on Chinese sentences failed to find ELAN when keeping the sentential context unchanged (e.g., Yu and Zhang, 2008; Zhang et al., 2010, Experiment I, 2013; Wang S. et al., 2013; Wang et al., 2015; Yang et al., 2015). Furthermore, word category violation is deemed as syntactic violation from the linguistic perspective, while in Chinese, a word category violation is always accompanied with the semantic violation, and thus it can serve as the double-violation condition, which is valid for examining the interplay between syntactic and semantic processing in both alphabetic languages and Chinese. In light of the established rationales (e.g., Ye et al., 2006; Zhang et al., 2010, 2013), we could attribute the absence of ELAN to the null temporal and functional priority of syntax over semantics in Chinese, thus in support of the notion that in Chinese syntactic processing (esp., reflected by the word category combinations) is highly interactive with semantic processing as Chinese is assumed to depend on meanings heavily.

In addition, although ELAN was mostly identified from auditory experiments, the possibility cannot be ruled out that ELAN could be elicited from visual presentations (Gunter and Friederici, 1999; Friederici, 2017). For example, Dikker et al. (2009) noted that early syntactic processing might involve visual perception and analysis, which would facilitate the time course of syntactic processing. As the current study presented all sentences visually, the absence of ELAN in the current study might not relate to the input modality of language stimuli (but see Limitation for further discussion).

Even though ELAN was absent for both groups, L1 speakers and L2 learners manifested greater negativities in the N400 time window and greater positivities in the P600 time window in bilateral central-posterior sites when recognizing double violations than pure semantic violations. In particular, the priority of syntactic information processing could also be reflected by the block effect from syntactic violation on subsequent semantic processing (Steinhauer and Drury, 2012). Consistent with existing studies on Chinese simple structures (Yu and Zhang, 2008; Wang S. et al., 2013; Zhang et al., 2013; Zeng et al., 2020), the current study detected greater N400 from double violations than from the NORM condition, which is not supportive of a block effect. Meanwhile, this N400 was widely distributed at the central-posterior sites on the human scalp, which is in line with these related ERP studies. However, previous studies pointed out that the blocked semantic N400 might be affected by task demands and material properties. For instance, Hahne and Friederici (2002, Experiment II) found that N400 was elicited from double-violation sentences when the participants were engaged in a semantic judgment task. The authors therefore believed that semantic processing requires more cognitive control and is not as automatic as syntactic processing. Likewise, Zhang et al. (2010) admitted that the task of their Experiment I, which required the participants to answer questions related to the semantics of the experimental sentences, might also involve a bias toward semantic processing, thus resulting in larger N400 for double-violation sentences. However, they also identified significantly larger negativities associated with the double violations in experiment II, which employed an overall acceptability judgment task that was neutral with regard to syntactic and semantic processing.

In addition to task demands, N400 and the block effect could be impacted by the accessibility order of syntactic and semantic information in the experimental materials (Hahne and Friederici, 2002; van den Brink and Hagoort, 2004). In Hahne and Friederici's (2002) double-violation sentence "*Das Turschloß wurde im gegessen*.," syntactic violation was created by word category violation in "*im gegessen*" locally, while semantic violation was realized by the long-distance violation between "*Das Turschloß*" and "*gegessen*." It is thus debatable that the asymmetry of the violation distance may induce ELAN and further cause a block effect. In other words, the extraction of the subject N "*Das Turschloß*" from working memory and the establishment of the relationship between the subject and the predicate verb "*gegessen*" were obviously slower than the immediate collocation between adjacent words (i.e., "*im gegessen*"). As such, the first violation the participants encountered was the violation of word category. However, Zhang et al. (2010) employed a similar manipulation as Hahne and Friederici (2002) in Experiment II, where ELAN and block effects were still not found in the asymmetry of violation distance.

To resolve the confounding factors mentioned above on the N400 findings, the current study employed an overall acceptability judgment task (Zhang et al., 2010) to prevent participants' potential semantic bias. Importantly, our study focused on Chinese complex sentences where the disagreements between critical words and their collocation took place at the same distance for both the SEM and SEM + SYN conditions. As such, word syntactic information and semantic information could be accessed at the same time. Previous studies used Chinese simple sentences including "BA" and "BEI" structures as experimental materials, where syntactic processing may be weakened due to the relatively local syntactic violation. However, the syntactic complexity of our complex sentences was relatively higher, which could better reflect the role of word category information in long-distance dependency processing (see also Gibson, 1998). Furthermore, in our study, the critical words were placed in the middle of experimental sentences instead of at the end. This operation could effectively eliminate the wrap-up effect on ERP signals (Hahne and Friederici, 2002; Hagoort, 2003). Collectively, we can further verify that Chinese syntactic and semantic information are processed in parallel rather than in a serial manner (Kuperberg, 2007; Zhang et al., 2013; Fromont et al., 2020).

As expected, we observed that SEM + SYN induced significantly larger negativities than SEM in the N400 time window for both L1 and L2 speakers. As N400 conventionally indexes a process of semantic violation detection, our findings suggested an interactive pattern of syntactic and semantic violations in this time window. The consequences of a semantic violation on the N400 amplitude were boosted by an additional syntactic violation, while there was no boost of syntactic violation on P600 amplitude by additional semantic violation, thus manifesting an asymmetric pattern between semantic and syntactic processing (Hagoort, 2003). As such, we interpret the enhanced N400 in the SEM + SYN condition as additional difficulty in semantic integration from syntactic violation.

In line with previous findings (Wang S. et al., 2013; Wang et al., 2015), our study identified late positivities in the P600 time-window at bilateral centro-parietal sites for both SEM + SYN and SEM conditions. In addition, the enhanced P600 elicited by SEM + SYN compared with SEM tended to manifest an overlap between syntactic and semantic violations. Importantly, this pattern was consistent for both L1 speakers and L2 learners. P600 has been associated with the integration (Kaan et al., 2000; Friederici, 2011) and restoration (Hagoort et al., 1993; Friederici et al., 2002a; Kaan and Swaab, 2003) of various types of sentence information. In contrast, Zhang et al. (2013) found that double violations of verb transitivity and semantics did not induce the late positive component in the P600 time window compared with normal sentences. The authors thus interpreted the P600 as a sensitivity to the degree of sentence abnormality. Frisch et al. (2004) found that syntactic violation represented by word category violation could block the processing of argument structure, while the simple argument structure violation was set by verb transitivity violation. As such, verb transitivity violation may be lower than word category violation with regard to the degree of syntactic violation. Similarly, syntactic violation and double violations induced comparable P600 patterns among French L1/L2 speakers (Fromont et al., 2020) and Chinese L1 speakers (Lu, 2015; Li, 2016). However, as these studies used simple sentences where syntactic violations were always contaminated by semantic violations to some extent, the interplay between syntactic and semantic information in the P600 time window remains to be elucidated. Our study also detected a P600 associated with pure semantic violation in both groups. Kuperberg (2007) held that when semantic information was abnormal, a semantic memory-based stream could generate semantic illusion, which could decrease N400 amplitudes and merge the lexical entries into a combinatorial stream. This combination could enable a semantic (re) analysis and further cause a P600 component. In addition, P600 caused by semantic violation might reflect a combinatorial mechanism of semantics and syntax from a more general sense. Therefore, for both L1 and L2 speakers in our study, the observed P600 related to SEM + SYN and SEM might implicate an integral index of final repair and integration of semantics and syntax in complex sentence comprehension. One contribution of our study with regard to this issue is that we extended the P600 findings to Chinese L2 learners reading complex sentences. Even though their L1 and L2 exhibit linguistic differences, highly proficient L2 learners can still present similar patterns for syntactic processing as native speakers of the target language from the absence of ELAN to greater N400 and P600 associated with SEM + SYN than NORM. Taken together, our L2 findings are consistent with Friederici et al.'s (2002b) notion that highly proficient L2 learners can develop native-like processing strategies. For both groups, syntax does not present a temporal and functional priority over semantics, and there is an intensive interaction between syntactic and semantic information in the N400 time window such that double violations are associated with enhanced negativities due to accumulated semantic and syntactic information (Hagoort, 2003). In the later time window of P600, positivities could reflect participants' repair and integrity of the complex structures.

However, although L2 learners and L1 speakers had similar processing patterns shown by ERP, the difference waves' latency of "SEM + SYN - SEM" of L2 learners was longer than that of L1 speakers. This result was consistent with Yang (2012), who investigated the differences between L1 Chinese speakers and highly proficient German-speaking Chinese L2 speakers in processing the Mandarin "BEI" structure. According to the between-group analysis results as well as the visualization in Figure 4, L2 learners showed a significant delay of the N400 and P600 latencies. Thus, L2 speakers were still slower than L1 speakers in detecting, repairing, and integrating syntactic and semantic violations when reading Chinese complex sentences.

Specifically, the unified competition model (MacWhinney, 1997, 2005) holds that L2 learners employ the cognitive resources and processing strategies from their L1 to address L1-L2 shared structures. However, in regard to a marked contrast between L1 and L2, it remains unknown how L2 learners from different learning stages process the complex L1 structure. Our results revealed that highly proficient L2 learners developed parallel patterns to L1 speakers when processing center-embedded Chinese relative clauses, while the only difference was shown in the longer latency. The difference between L1 and L2 implies a difference in degree, not in kind, thus supporting a unified competition account. Although existing studies have established the role of L2 proficiency (Bowden et al., 2013; Yu and Dong, 2018; Jin et al., 2019), the similar processing pattern between L1 speakers and L2 learners indicates that given adequate language exposure, L2 learners can effectively suppress the influence of their mother tongue (i.e., L1) and show native-like sentence processing. In particular, the interaction between syntactic and semantic information processing in reading Chinese sentences plays a similar role for both L1 speakers and highly proficient L2 learners. Nevertheless, we found a difference in the ERP latencies, which might suggest the additional time that L2 learners need to process, repair, and integrate syntactic and semantic information. In light of the declarative/procedural (DP) model (Ullman, 2015), in both L1 and L2, the knowledge of syntax should initially be learned in declarative memory. In parallel, procedural memory should also gradually develop. After sufficient exposure to the language, procedural memory-based syntactic processing should take precedence over analogous declarative knowledge, resulting in increasing automatization of syntactic processing, which provides the opportunity for L2 learners to develop native-like automatic processing. However, Ullman (2015) also noted that even after years of exposure, adult L2 learners might not attain the degree of proceduralization of their syntax as L1 or early L2 learners because the ability of procedural memory gradually fades with increasing learners' age. In other words, there seems to be an unbridgeable gap in the automaticity of sentence processing between L1 speakers and late L2 learners. However, based on the current results, we propose that the nature of this gap might be quantitative rather than qualitative.

5. Conclusion

In light of ERP techniques, the current study examined how syntactic category information and semantics interact with each other regarding the time course of reading Chinese complex sentences among L1 speakers and highly proficient L2 learners. Our results revealed that double violations of semantics and syntax did not elicit an ELAN effect for either group. In addition, double violations evoked enhanced N400 and P600 compared with normal sentences, thus exhibiting a consistent biphasic waveform pattern. These findings indicate a highly interactive relationship between syntactic and semantic information during Chinese complex sentence reading and suggest that syntax does not manifest a temporal and functional priority, which could relate to the typological specialties of the Chinese language system. Importantly, Chinese L2 learners with a morphologically rich language background could effectively suppress the influence from their L1 and show a similar ERP pattern to native Chinese speakers. Our findings further suggest that the syntax-first pattern in L2 might be limited to specific languages such that there might exist an interaction between L2 proficiency and language typology. Languages with differing morphological diversity might exhibit different electrophysiological patterns regarding the interplay between syntactic and semantic processes.

6. Limitation

The current study has several limitations warranting discussion.

First, participants' gender differences were not well controlled in each group. Even though both L1 and L2 groups shared a similar gender ratio (L1: male:female = 6/17; L2: male:female = 7/16) in the present study (see also Wang S. et al., 2013: 6 to 15; Wang et al., 2015: 4 to 17 for a similar ratio), making the results comparable between the groups and the studies, whether and to what extend the gender factor would modulate the relationship between semantic and syntactic processing during Chinese complex sentence reading still await to be explored.

Second, the working memory capacities were comparable across the two groups (L1 vs. L2), and for the materials, as the critical word position in the sentence was identical across all the conditions (Norm, SEM, and SEM+,SYN), WM variations (if any) should not be confounded with our results. Nevertheless, we did not further examine whether and how the individual working memory differences might modulate the sentence processing, which is a valuable research question for further investigations.

Third, we only aimed to investigate whether relatively-high-proficiency Chinese L2 learners with a distinct language background from Chinese could process the syntactic and semantic information as Chinese L1 speakers, Thus, the potential modulation effect of L2 proficiency was not in the focus of the present study. Nevertheless, future studies may include more languages and various language proficiency levels to address these issues in a more systematic fashion.

Last, although ELAN was mostly identified from auditory experiments, evidence that it could appear in reading studies (i.e., in the visual modality) was also reported (e.g., Neville et al., 1991; Dikker et al., 2009; Wang S. et al., 2013). As such, we attributed the absence of ELAN to the absence of initial syntactic processing, the rationale of which is also in line with existing studies in Chinese (Yu and Zhang, 2008; Zhang et al., 2010; Wang S. et al., 2013; Wang et al., 2015; Zeng et al., 2020), rather than the mere modality effect. However, conducting the present experiment in the auditory domain is expected in the near future so as to evaluate whether the ELAN effect would be amplified in the auditory modality.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Beijing Normal University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LC and LF: came up with the original idea of this study. LC: conducted the ERP experiment, and analyzed the data. LC, MY, and FG: completed the first draft of this manuscript, which was further revised by ZF, PW, and LF. All authors participated in the discussion of the results and prepared the revised version of the manuscript for submission.

Funding

This work was funded by National Social Science Fund of China (22CYY017).

Acknowledgments

The authors sincerely feel grateful for the reviewers' insightful comments. Special thanks are extended to Di Gao, Yuanyue Zhang, Jia Guo, Shuhua Li, and Mingzhu Pan for collecting the data. The authors also thank Prof. Baoguo Chen, Prof. Taomei Guo, and Prof. Yan Song from Beijing Normal University for their helpful comments and constructive inputs. We sincerely feel grateful for AJE for English polishing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1143062/full#supplementary-material

References

Adger, D. (2013). *A syntax of substance*. Cambridge, MA: MIT Press.

Bemis, D. K., and Pylkkänen, L. (2011). Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J. Neurosci.* 31, 2801–2814. doi: 10.1523/JNEUROSCI.5003-10.2011

Bowden, H. W., Steinhauer, K., Sanz, C., and Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia* 51, 2492–2511. doi: 10.1016/j.neuropsychologia.2013.09.004

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

Chen, B., and Ning, A. (2008). The comparison of processing difficulty between Chinese subject-relative and object-relative clauses [in Chinese]. *Chin. J. Appl. Linguist.* 14, 029–034. doi: 10.3969/j.issn.1006-6020.2008.01.005

Chen, B., Ning, A., Bi, H., and Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychol.* 129, 61–65. doi: 10.1016/j.actpsy.2008.04.005

Chen, L., Wu, J., Fu, Y., Kang, H., and Feng, L. (2019). Neural substrates of word category information as the basis of syntactic processing. *Hum. Brain Mapp.* 40, 451–464. doi: 10.1002/hbm.24386

Chesi, C., and Moro, A. (2015). "The subtle dependency between competence and performance" in *50 years later: Reflections on Chomsky's aspects*. eds. Á. J. Gallego and D. Ott (Cambridge, MA: MIT Working Papers in Linguistics), 33–46.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu, HI: University of Hawaii Press

Dikker, S., Rabagliati, H., and Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition* 110, 293–321. doi: 10.1016/j.cognition.2008.09.008

Epstein, S. D., Kitahara, H., and Daniel Seely, T. (2014). Labeling by minimal search: implications for successive-syclic A-movement and the conception of the postulate phase. *Linguist. Inq.* 45, 463–481. doi: 10.1162/ling_a_00163

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Feng, L., and Wang, Y. (2013). A study on processing difficulty of Chinese relative clauses and the affecting factors [in Chinese]. *Lang. Teach. Linguist. Stud.* 5, 1–9.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. Cambridge, MA: The MIT Press

Friederici, A. D., Hahne, A., and Saddy, D. (2002a). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *J. Psycholinguist. Res.* 31, 45–63. doi: 10.1023/A:1014376204525

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi: 10.1016/0926-6410(93)90026-2

Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002b). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 99, 529–534. doi: 10.1073/pnas.012611199

Friederici, A. D., and Weissenborn, J. (2007). Mapping sentence form onto meaning: the syntax-semantic interface. *Brain Res.* 1146, 50–58. doi: 10.1016/j.brainres.2006.08.038

Frisch, S., Hahne, A., and Friederici, A. D. (2004). Word category and verb-argument structure information in the dynamics of parsing. *Cognition* 91, 191–219. doi: 10.1016/j.cognition.2003.09.009

Fromont, L. A., Royle, P., Perlitch, I., and Steinhauer, K. (2016). Re-evaluating the dynamics of phrase-structure processing using event related potentials: the case of syntactic categories in French. *Int. J. Psychophysiol.* 108:86. doi: 10.1016/j.ijpsycho.2016.07.270

Fromont, L. A., Royle, P., and Steinhauer, K. (2020). Growing random forests reveals that exposure and proficiency best account for individual variability in L2 (and L1) brain potentials for syntax and semantics. *Brain Lang.* 204:104770. doi: 10.1016/j.bandl.2020.104770

Fujita, K. (2014). "Recursive merge and human language evolution" in *Recursion: Complexity in cognition. Studies in theoretical psycholinguistics*. eds. T. Roeper and M. Speas (Switzerland: Springer International Publishing), 243–264. doi: 10.1007/978-3-319-05086-7_11

Gao, F., Wang, R., Armada-da-Silva, P., Wang, M. Y., Lu, H., Leong, C., et al. (2022). How the brain encodes morphological constraints during Chinese word reading: an EEG-fNIRS study. *Cortex* 154, 184–196. doi: 10.1016/j.cortex.2022.05.016

Ghani, U., Signal, N., Niazi, I. K., and Taylor, D. (2020). A novel approach to validate the efficacy of single task ERP paradigms to measure cognitive workload. *Int. J. Psychophysiol.* 158, 9–15. doi: 10.1016/j.ijpsycho.2020.09.007

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1

Goucha, T., and Friederici, A. D. (2015). The language skeleton after dissecting meaning: a functional segregation within Broca's Area. *NeuroImage* 114, 294–302. doi: 10.1016/j.neuroimage.2015.04.011

Goucha, T., Zaccarella, E., and Friederici, A. D. (2017). A revival of homo loquens as a builder of labeled structures: neurocognitive considerations. *Neurosci. Biobehav. Rev.* 81, 213–224. doi: 10.1016/j.neubiorev.2017.01.036

Gunter, T. C., and Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology* 36, 126–137. doi: 10.1017/S004857729997155X

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 883–899. doi: 10.1162/089892903322370807

Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483. doi: 10.1080/01690969308407585

Hagoort, P., and Indefrey, P. (2014). The neurobiology of language beyond single words. *Annu. Rev. Neurosci.* 37, 347–362. doi: 10.1146/annurev-neuro-071013-013847

Hagoort, P., Wassenaar, M., and Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cogn. Brain Res.* 16, 38–50. doi: 10.1016/S0926-6410(02)00208-2

Hahne, A., and Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cogn. Brain Res.* 13, 339–356. doi: 10.1016/S0926-6410(01)00127-6

Hahne, A., and Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cogn. Brain Res.* 11, 199–212. doi: 10.1016/S0926-6410(00)00071-9

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). Neuroscience: the faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

He, W., Chen, B., and Cui, P. (2012). The processing asymmetry effect of subject/object relative clause and its neural mechanism [in Chinese]. *Adv. Psychol. Sci.* 20, 981–994. doi: 10.3724/sp.j.1042.2012.00981

Hornstein, N. (2009). *A theory of syntax: Minimal operations and universal grammar*. Cambridge, UK: Cambridge University Press.

Hornstein, N., and Nunes, J. (2008). Adjunction, labeling, and bare phrase structure. *Biolinguistics* 2, 057–086. doi: 10.5964/bioling.8621

Hornstein, N., and Pietroski, P. (2009). Basic operations: minimal syntax-semantics. *Catalan J. Linguist.* 8, 113–139. doi: 10.5565/rev/catjl.148

Hoshi, K. (2018). Merge and labeling as descent with modification of categorization: a neo-Lennebergian approach. *Biolinguistics* 12, 039–054. doi: 10.5964/bioling.9135

Hoshi, K. (2019). More on the relations among categorization, merge and labeling, and their nature. *Biolinguistics* 13, 1–21. doi: 10.5964/bioling.9147

Hsiao, F., and Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition* 90, 3–27. doi: 10.1016/S0010-0277(03)00124-0

Isel, F., Hahne, A., Maess, B., and Friederici, A. D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biol. Psychol.* 74, 337–346. doi: 10.1016/j.biopsycho.2006.09.003

Jin, L. (1997). Syntactic, semantic and contextual characteristics of the BA sentence [in Chinese]. *Stud. Chin. Lang.* 6, 415–423.

Jin, H., Gao, F., and Chen, Z. (2019). Applications of event related potentials (ERP) in studies of second language syntactic acquisition [in Chinese]. *Chin. Teach. World* 33, 522–547. doi: 10.13724/j.cnki.ctiw.20190927.008

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084

Kaan, E., and Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *J. Cogn. Neurosci.* 15, 98–110. doi: 10.1162/089892903321107855

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain Lang.* 109, 68–74. doi: 10.1016/j.bandl.2008.06.002

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11, 539–550. doi: 10.3758/BF03196991

Lenneberg, E. (1967). *Biological foundations of language*. New York, NY: Wiley, *2*, 59–67

Li, X. (2016). *Syntactic and semantic processing of Chinese long passive Bei sentences: Evidence from event-related potentials [in Chinese]* Master's thesis Changsha: Hunan University.

Liu, T. (2011). *Neurolinguistics research on syntactic movement in Chinese [in Chinese]*. Ph.D.'s thesis, Nanjing: Nanjing Normal University.

Liu, T., Zhou, T., and Yang, Y. (2011). Universality of processing advantage for subject relative clause: an ERP study of Chinese relative clauses [in Chinese]. *Linguist. Sci.* 10, 1–20.

Lu, Q. (2015). *Syntactic and semantic processing of Chinese middle sentences: Evidence from event-related potentials [in Chinese]*. Master's thesis Changsha: Hunan University.

Ma, Z., Guo, T., Peng, D., and Yan, Y. (2007). ERP evidence for neural dissociation of processing Chinese nouns and verbs. *Prog. Nat. Sci.* 17, 77–80.

MacWhinney, B. (1997). "Second language acquisition and the competition model," in *Tutorials in bilingualism: Psycholinguistic perspectives*, ed. GrootA. M. B. de and J. Kroll Mahwah, NJ: Erlbaum, 113–142.

MacWhinney, B. (2005). "A unified model of language acquisition," in *Handbook of bilingualism: Psycholinguistic approaches*, ed. J. Kroll and GrootA. M. B. de Oxford: Oxford University Press, 49–67.

Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proc. Natl. Acad. Sci. U. S. A.* 106, 8362–8367. doi: 10.1073/pnas.0810928106

Makuuchi, M., Grodzinsky, Y., Amunts, K., Santi, A., and Friederici, A. D. (2013). Processing noncanonical sentences in Broca's region: reflections of movement distance and type. *Cereb. Cortex* 23, 694–702. doi: 10.1093/cercor/bhs058

Maran, M., Numssen, O., Hartwigsen, G., and Zaccarella, E. (2022). Online neurostimulation of Broca's area does not interfere with syntactic predictions: a combined TMS-EEG approach to basic linguistic combination. *Front. Psychol.* 13:968836. doi: 10.3389/fpsyg.2022.968836

Miyagawa, S., Berwick, R. C., and Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Front. Psychol.* 4:71. doi: 10.3389/fpsyg.2013.00071

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., and Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *J. Cogn. Neurosci.* 3, 151–165. doi: 10.1162/jocn.1991.3.2.151

Niharika, M. K., and Prema Rao, K. S. (2020). Processing syntax: perspectives on language specificity. *Int. J. Neurosci.* 130, 841–851. doi: 10.1080/00207454.2019.1707818

Petkov, C. I., and ten Cate, C. (2020). Structured sequence learning: animal abilities, cognitive operations, and language evolution. *Top. Cogn. Sci.* 12, 828–842. doi: 10.1111/tops.12444

Sanmiguel, I., Todd, J., and Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N 1 component of the auditory ERP. *Psychophysiology* 50, 334–343. doi: 10.1111/psyp.12024

Shen, J. (2016). *Nouns and verbs [in Chinese]*. Beijing: The Commercial Press.

Steinhauer, K., and Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120, 135–162. doi: 10.1016/j.bandl.2011.07.001

Steinhauer, K., White, E. J., and Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second. Lang. Res.* 25, 13–41. doi: 10.1177/0267658308098995

Ullman, M. T. (2015). "The declarative/procedural model: a neurobiologically motivated theory of first and second language" in *Theories in second language acquisition: An introduction*. eds. B. VanPatten and J. Williams. *2nd* ed (London, UK: Routledge), 135–158.

Unsworth, N., Heitz, R. P., Schrock, J. C., and Engle, R. W. (2005). An automated version of the operation span task. *Behav. Res. Methods* 37, 498–505. doi: 10.3758/BF03192720

van den Brink, D., and Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPS. *J. Cogn. Neurosci.* 16, 1068–1084. doi: 10.1162/0898929041502670

Wang, Y. (2011). *A study of the acquisition of relative clause in Chinese by English native speakers [in Chinese]*. Master's thesis Beijing: Beijing Normal University.

Wang, H., and Bing, W. (2013). Inconsistencies between the frequency of Chinese relative clauses and the difficulty of processing: evidence from ERP [in Chinese]. *Foreign Lang. Res.* 3, 13–22. doi: 10.3969/j.issn.1005-7242.2013.03.002

Wang, S., Mo, D., Xiang, M., Xu, R., and Chen, H. C. (2013). The time course of semantic and syntactic processing in reading Chinese: evidence from ERPs. *Lang. Cogn. Process.* 28, 577–596. doi: 10.1080/01690965.2012.660169

Wang, F., Ouyang, G., Zhou, C., and Wang, S. (2015). Re-examination of Chinese semantic processing and syntactic processing: evidence from conventional ERPs and reconstructed ERPs by residue iteration decomposition (RIDE). *PLoS One* 10:e0117324. doi: 10.1371/journal.pone.0117324

Wang, X., Zhong, Y., Fan, W., and Lei, X. (2013). The timing of interaction between syntax and semantic in the Chinese sentence processing: evidence from ERP research [in Chinese]. *J. Psychol. Sci.* 36, 827–831. doi: 10.16719/j.cnki.1671-6981.2013.04.027

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., et al. (2020). Non-adjacent dependency learning in humans and other animals. *Top. Cogn. Sci.* 12, 843–858. doi: 10.1111/tops.12381

Xu, K., Duann, J. R., Hung, D. L., and Wu, D. H. (2019). Preference for object relative clauses in Chinese sentence comprehension: evidence from online self-paced reading time. *Front. Psychol.* 10:2210. doi: 10.3389/fpsyg.2019.02210

Yamada, Y., and Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Res.* 1130, 167–180. doi: 10.1016/j.brainres.2006.10.052

Yan, X. (2018). *The second language processing and production of Chinese relative clauses [in Chinese]*. Master's thesis Guangzhou: South China University of Technology.

Yang, Y. (2012). *Processing of Chinese BEI sentences by Chinese and German speakers—An ERP study [in Chinese]* Master's thesis Shanghai: Shanghai International Studies University.

Yang, S., Cai, Y., Xie, W., and Jiang, M. (2021). Semantic and syntactic processing during comprehension: ERP evidence from Chinese QING structure. *Front. Hum. Neurosci.* 15:701923. doi: 10.3389/fnhum.2021.701923

Yang, Y., Wu, F., and Zhou, X. (2015). Semantic processing persists despite anomalous syntactic category: ERP evidence from Chinese passive sentences. *PLoS One* 10:e013193. doi: 10.1371/journal.pone.0131936

Ye, Z., Luo, Y. J., Friederici, A. D., and Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: evidence from event-related potentials. *Brain Res.* 1071, 186–196. doi: 10.1016/j.brainres.2005.11.085

Yu, Z., and Dong, Y. (2018). The neurocognitive mechanism of gender agreement processing in L2 pronoun resolution [in Chinese]. *Mod. Foreign Lang.* 41, 467–480.

Yu, J., and Zhang, Y. (2008). When Chinese semantics meets failed syntax. *Neuroreport* 19, 745–749. doi: 10.1097/WNR.0b013e3282fda21d

Zaccarella, E., Schell, M., and Friederici, A. D. (2017). Reviewing the functional basis of the syntactic merge mechanism for language: a coordinate-based activation likelihood estimation meta-analysis. *Neurosci. Biobehav. Rev.* 80, 646–656. doi: 10.1016/j.neubiorev.2017.06.011

Zeng, T., Li, Y., and Wu, M. (2020). Syntactic and semantic processing of passive BEI sentences in Mandarin Chinese: evidence from event-related potentials. *Neuroreport* 31, 979–984. doi: 10.1097/WNR.0000000000001507

Zhang, Y., Li, P., Piao, Q., Liu, Y., Huang, Y., and Shu, H. (2013). Syntax does not necessarily precede semantics in sentence processing: ERP evidence from Chinese. *Brain Lang.* 126, 8–19. doi: 10.1016/j.bandl.2013.04.001

Zhang, Y., Piao, Q., Yu, J., and Yang, Y. (2011). The functional status of syntactic category processing in sentence comprehension [in Chinese]. *Adv. Psychol. Sci.* 19, 1741–1748. doi: 10.3724/SP.J.1042.2011.01741

Zhang, Q., and Yang, Y. (2010). Object preference in the processing of relative clause in Chinese: ERP evidence [in Chinese]. *Linguist. Sci.* 9, 337–353.

Zhang, Y., Yu, J., and Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 765–781. doi: 10.1037/a0019254

Zhao, Z. (2011). *A study on the acquisition of Chinese relative clauses [in Chinese]*. Master's thesis Beijing: Beijing Normal University.

Zhou, C., Li, F., and Chen, R. (2020). An eye-tracking study on the processing of Chinese relative clause [in Chinese]. *Foreign Lang. Educ.* 41, 40–46. doi: 10.16362/j.cnki.cn61-1023/h.2020.01.009

Zhou, T., Zheng, W., Shu, H., and Yang, Y. (2010). The superiority of processing Chinese object-extracted relative clause: evidence from aphasic studies [in Chinese]. *Linguist. Sci.* 9, 225–243.

Zhu, D. (1982). *Lecture notes on grammar [in Chinese]*. Beijing: The Commercial Press.

Zhu, Y., Xu, M., Lu, J., Hu, J., Kwok, V. P. Y., Zhou, Y., et al. (2022). Distinct spatiotemporal patterns of syntactic and semantic processing in human inferior frontal gyrus. *Nat. Hum. Behav.* 6, 1104–1111. doi: 10.1038/s41562-022-01334-6

TYPE Original Research
PUBLISHED 11 May 2023
DOI 10.3389/fpsyg.2023.1132039

OPEN ACCESS

EDITED BY
Yang Zhang,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Zhilong Xie,
Jiangxi Normal University, China
Shifa Chen,
Ocean University of China, China

*CORRESPONDENCE
Yufang Yang
✉ yangyf@psych.ac.cn

RECEIVED 26 December 2022
ACCEPTED 14 April 2023
PUBLISHED 11 May 2023

CITATION
Ding J, Liang P, Guo X and Yang Y (2023) The influence of conceptual concreteness on the reading acquisition and integration of novel words into semantic memory via thematic relations.
Front. Psychol. 14:1132039.
doi: 10.3389/fpsyg.2023.1132039

The influence of conceptual concreteness on the reading acquisition and integration of novel words into semantic memory via thematic relations

Jinfeng Ding[1,2], Panpan Liang[1,2], Xinyu Guo[1,2] and Yufang Yang[1,2]*

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China, [2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Plenty of studies have been conducted to reveal neurocognitive underpinnings of conceptual representation. Compared with that of concrete concepts, the neurocognitive correlates of abstract concepts remain elusive. The current study aimed to investigate the influence of conceptual concreteness on the reading acquisition and integration of novel words into semantic memory. We constructed two-sentence contexts in which two-character pseudowords were embedded as novel words. Participants read the contexts to infer the meaning of novel words which were either concrete or abstract, and then performed a lexical decision task and a cued-recall memory task. In lexical decision task, primed by the learned novel words, their corresponding concepts, thematically related or unrelated words as well as unlearned pseudowords were judged whether they were words or not. In memory task, participants were presented with the novel words and asked to write down their meaning. The contextual reading and memory test can demonstrate the modulation of conceptual concreteness on novel word learning and the lexical decision task can reveal whether concrete and abstract novel words are integrated into semantic memory similarly or not. During contextual reading, abstract novel words presented for the first time elicited a larger N400 than concrete ones. In memory task, the meaning of concrete novel words was recollected better than abstract novel words. These results indicate that abstract novel words are more difficult to acquire during contextual reading, and to retain afterwards. For lexical decision task behavioral and ERPs were graded, with the longest reaction time, the lowest accuracy and the largest N400s for the unrelated words, then the thematically related words and finally the corresponding concepts of the novel words, regardless of conceptual concreteness. The results suggest that both concrete and abstract novel words can be integrated into semantic memory via thematic relations. These findings are discussed in terms of differential representational framework which posits that concrete words connect with each other via semantic similarities, and abstract ones via thematic relations.

KEYWORDS

concreteness, thematic relation, novel word learning, contextual reading, EEG

1. Introduction

Concepts are a component of knowledge system and play an important role in human cognition. It is well-established that concrete concepts are mainly represented by sensory-motor information (for reviews see Capitani et al., 2003; Binder and Desai, 2011; Lambon Ralph et al., 2017). But the representation of abstract concepts – concepts that have no perceivable referents – remains elusive (but see Jessen et al., 2000; Binder et al., 2005; Wang et al., 2010; Pulvermüller, 2013; Hoffman et al., 2015; Binder, 2016; Borghi, 2022). One of the debated topics in this area is whether the manner of connections with other concepts is different for concrete and abstract concepts. The current study addressed the issue by examining how concrete and abstract novel words are learnt in contextual reading and integrated into semantic memory. To this end, we simulated vocabulary learning in second language (L2) study by associating novel words with known concepts in contexts, a crucial way to novel word learning (Batterink and Neville, 2011; Elgort and Warren, 2014). The findings will shed new light on the influence of conceptual concreteness on L2 reading acquisition.

1.1. Concreteness effect in word acquisition

Previous studies have shown that concrete words are learned better over abstract words in first language (L1, e.g., Palmer et al., 2013) and L2 (e.g., De Groot and Keijzer, 2000; Elgort and Warren, 2014) acquisition. For instances, Palmer et al. (2013) explored the influence of conceptual concreteness on the acquisition and subsequent processing of novel words. They asked native English speakers to learn real but rare English words which were paired with their definitions. It was found that participants acquired concrete words better than abstract words as indicated by the higher accuracy in the word-definition pairing test for the former ones. Meanwhile, participants responded faster to concrete words compared to abstract words in the later semantic categorization and lexical decision tasks, indicating a processing advantage of concrete (vs. abstract) words.

In L2 vocabulary learning, using a paired-associate learning in which English pseudowords were paired with Dutch translations, De Groot and Keijzer (2000) found that concrete novel words were easier to learn and less susceptible to forgetting for native speakers of Dutch, as indicated by higher recall accuracies for concrete words compared to abstract words. In contextual reading, Mestres-Missé et al. (2014) examined the influence of conceptual concreteness on novel word learning. They asked participants to read sentences ending with novel words and found that the reading times in the second (vs. first) sentences increased for abstract novel words but decreased for concrete novel words. These results indicated that the meaning of concrete words was discovered and learned faster than abstract words.

The concreteness advantage effect in word learning could be accounted for by dual coding theory in which an imagery-based system and a verbal-based system are associated with concepts in semantic memory (Paivio, 1986, 1991). According to the dual coding theory, while both the imagery-based and verbal-based systems are associated with concrete words, only the verbal-based system is involved for abstract words. With the support of the two systems, concrete words were learned better than abstract words. The neuroanatomical structure for dual-coding knowledge has been depicted by the neural framework of Bi (2021), in line with the direct neural evidence provided by previous study (Mestres-Missé et al., 2009). In the study, participants were asked to read sentences and to derive the meaning of novel words associated to concrete and abstract concepts. The important finding was that the ventral anterior fusiform gyrus, a region associated with high-level visual processing, showed a selective activation for novel concrete words, indicating the involvement of the imagery-based/sensory-derived system in concrete word learning.

Besides the dual coding theory, the concreteness effect could also be explained by context availability hypothesis (Schwanenflugel and Shoben, 1983; Schwanenflugel and Stowe, 1989). This theory argues that the difference between concrete and abstract words could be attributed to the discrepancy in the availability of contextual information. That is, abstract words tend to appear in a wider range of contexts than concrete words, which results in a looser contextual constraint for the abstract words (Hill et al., 2014). Based on this theory, the context wherein concrete and abstract concepts appear will modulate the concreteness effect. For instance, the concreteness effect in lexical decision task was absent after concrete and abstract words were processed in sentences (Schwanenflugel et al., 1988; van Hell and de Groot, 2008), showing a facilitation effect of sentence context on the processing of abstract words. The stronger contextual facilitation to abstract words processing than concrete words was also observed in other studies (Hoffman et al., 2015; Bechtold et al., 2021, 2022). However, in contrast to this theory, Taylor et al. (2019) found that concreteness effects were reliable for words presented both in isolation and in contexts, and similar for words in low-and high-constraint contexts.

Therefore, how the conceptual concreteness influences word acquisition in sentence reading needs to be further explored. If the facilitation effect stems from the context availability, there should be no difference between the acquisition of concrete and abstract novel words in sentence reading. In contrast, according to the dual coding theory the superiority of concrete concepts in learning through sentence reading should be observed.

1.2. Semantic integration of novel words

Previous studies have shown that, in contextual reading, novel words can be rapidly learned and integrated into the semantic network via different semantic relations (Borovsky et al., 2012; Chen et al., 2014; Ding et al., 2017a; Zhang et al., 2017, 2019). Among these studies, two important semantic relations, taxonomic and thematic relations (Mirman et al., 2017), are mainly explored. Taxonomic relation is established on the bases of semantic similarities between concepts (Murphy, 2010), and thematic relation organizes concepts by their complementary roles in the same event or episode (Estes et al., 2011).

In contextual reading, the experiments on novel word learning usually include a learning phase and a following testing phase. In the learning phase, novel words are embedded in contexts and participants are asked to read the contexts for deriving the meaning of novel words incidentally or intentionally. In the testing phase, the novel words serving as primes are paired with different types of semantically (taxonomically or thematically) related or unrelated target words.

Participants are asked to judge whether the target word is a word or not. By comparing the behavioral and neural responses to the different types of target words, researchers could examine whether the novel words have been integrated into semantic memory through certain semantic relations.

Using the contextual learning paradigm, it is found that when contexts provide semantic features of concepts, the taxonomically but not thematically related words could be primed by the novel words, as indicated by the smaller N400s for the taxonomically related words than the unrelated words (Ding et al., 2017a). The results suggest that semantic integration of novel words into semantic memory is through taxonomic relations but not thematic relations (Ding et al., 2017a). However, when contexts describing events or scenarios were provided in the learning phase, the novel words could prime both taxonomically and thematically related words, indicating semantic integration of novel words into semantic memory via both taxonomic and thematic relations (Zhang et al., 2017, 2019). Taken together, the results suggest that the information provided by learning contexts influences the way of novel words integration into semantic memory.

Most of the aforementioned contextual reading studies are concerned with concrete concepts. How novel words associated with abstract concepts are integrated into semantic memory needs to be addressed. An influential theory, different representational frameworks (Crutch and Warrington, 2005, 2007; Crutch et al., 2006), assumes that concrete concepts are organized by semantic similarity and abstract ones by semantic association. In other words, concrete concepts connect with other concepts via taxonomic relations, and abstract ones via thematic relations. The theory stemmed from the results of a series of neuropsychological studies with patients. In the first study (Crutch and Warrington, 2005) a patient with semantic refractory assess deficit was asked to perform a task of matching spoken word to a written one. The target written word was paired with other three taxonomically (similarity-based) or thematically (association-based) related words. The results revealed a taxonomic but not thematic interference for concrete words, and a reverse pattern for abstract ones. The different representations for concrete and abstract concepts are also confirmed by studies using different tasks or techniques in normal individuals (Crutch et al., 2009; Duñabeitia et al., 2009). However, the findings were not replicated by some subsequent studies (e.g., Hamilton and Coslett, 2008; Zhang et al., 2013; Geng and Schnur, 2015). For instance, Geng and Schnur (2015) also used a spoken word to written word matching task and reported higher error rates and longer reaction times for both taxonomically and thematically related words compared with unrelated words irrespective of the concreteness of the target words. In addition, the interference was larger for the taxonomic than thematic words, indicating the prominent role of taxonomic relation in the organizations of both concrete and abstract concepts (Xu et al., 2018).

Given the abovementioned debate, how abstract novel words are integrated into semantic memory after contextual reading needs to be further explored. Ding et al. (2017b) examined whether concrete and abstract novel words learnt through contextual reading could be integrated into semantic memory via taxonomic relations by using event-related potential (ERP) technique with paradigm similar to previous studies (e.g., Chen et al., 2014; Ding et al., 2017a; Zhang et al., 2017, 2019). Participants were asked to derive the meaning of novel words through reading episodic contexts and performed a lexical decision task with novel words as semantic primes. The results showed that both concrete and abstract novel words could prime their taxonomically related words, indicating that both concrete and abstract novel words can be integrated into semantic memory via taxonomic relations, which was compatible with Geng and Schnur (2015).

However, the taxonomic and thematic relations are not mutually exclusive (Hill et al., 2014), that is, the taxonomically related words could also be thematically related with each other. Moreover, thematic relations are more dominant to abstract words (Zhang et al., 2013; Troyer and McRae, 2022). For example, Zhang et al. (2013) found both categorical (taxonomic) and associative (thematic) blocking interference for abstract words, but only categorical blocking interference for concrete words, indicating more relevance of abstract concepts to thematic relations. It is, hence, needed to further examine the (a)symmetry of semantic integration of concrete and abstract novel words via taxonomic and thematic relations.

Based on the aforementioned studies, the current study aimed to explore how the concreteness of concepts influences word acquisition during contextual reading, and whether the concrete and abstract novel words could be integrated into semantic memory via semantic relations in a similar way.

2. Research methodology

2.1. Research questions and theoretical hypotheses

The study aimed to address two research questions. Whether the concrete novel words could be learned better than the abstract ones in contextual reading? After contextual reading, whether the concrete and abstract novel words could be integrated into semantic memory through thematic relations in the same way? The two questions were examined through three experimental tasks. First, in the context reading phase, EEG data were recorded when participants read contexts to explore the difference in brain responses to concrete and abstract novel words in thematic contexts. Second, in the testing phase, lexical decision task was used to test how the concrete and abstract novel words were integrated into semantic memory when primed by the novel words. Third, a cued-recall memory test was conducted to explore the effect of concreteness on memory consolidation. The cued-recall memory test being conducted after the lexical decision task is to exclude the influence of explicit semantic retrieval of novel words on the implicit semantic priming to their corresponding concepts. In sum, the results of contextual reading and cued-recall memory tasks would show the learning effects for novel words, and the lexical decision task further elucidate integration processes of novel words into semantic memory.

As to the effect of the conceptual concreteness on word acquisition during contextual reading, in terms of the dual coding theory (Paivio, 1986, 1991), the concrete novel words would be learned better than the abstract novel words with the support of both verbal-and imagery-based systems, which would be reflected by smaller N400s and/or higher cued-recall accuracy for the concrete (vs. abstract) novel words. However, on the basis of the context availability hypothesis (Schwanenflugel and Shoben, 1983; Schwanenflugel and Stowe, 1989),

TABLE 1 Examples of the stimuli in the learning phase and lexical decision task.

Learning discourses in the learning phase		
Concrete condition	Abstract condition	
小蝌蚪长大之后会变成一只芋沌，此刻池塘里的荷叶上蹲着一只芋沌在捕食昆虫。 A little tadpole grows up into a **yudun**. Right now on a lotus leaf there is a **yudun** catching insects.	获得诺贝尔奖是科学家的最高栗芸，大家都在为赢得这份最高的栗芸而努力。 To a scientist wining the Nobel prize is the highest **liyun**. Everyone is trying hard to win this highest **liyun**.	
Targets in the lexical decision task		
	Concrete condition	Abstract condition
Corresponding concept (CC)	青蛙(frog)	荣誉(honor)
Thematically related word (TR)	稻田(paddy)	掌声(applause)
Unrelated word (UR)	裤子(pants)	说法(statement)
Pseudoword	晾岌(liang ji)	贡颠(gong dian)
Pseudoword	甚筋(shen jin)	募旺(mu wang)
Pseudoword	泉愧(quan kui)	屑泊(xie bo)

The original stimuli are Chinese followed by English translations. The novel words are in boldface in the discourses.

the learning effects would be of no difference for the concrete and abstract novel words since the learning contexts provided equal context availability for them.

For the integration of concrete and abstract novel words into semantic memory via thematic relations, it was predicted that the novel words would connect with the thematically related words as indicated by the behavioral and/or neural semantic priming effects, namely, smaller N400s, shorter reaction time, and/or higher accuracy for the thematically related targets than for the unrelated targets, based on previous studies (e.g., Zhang et al., 2017, 2019). If the different representational framework (Crutch and Warrington, 2010) applies to the learning process, the semantic priming effects would be different in the two conditions. In contrast, if concrete and abstract concepts connect with other concepts through similar semantic relations, similar semantic priming effects in the concrete and abstract conditions would be observed.

2.2. Participants

Twenty-four university students (mean age 21.4 years, range 18–27, 12 males), right-handed native speakers of Chinese, took part in the experiment. They did not report dyslexia or neural impairment with normal or corrected-to-normal vision. This research was approved by the Ethics Committee of Institute of Psychology, Chinese Academy of Sciences. Before the EEG experiment, all subjects were given a written informed consent in accordance with the principles of the Declaration of Helsinki.

2.3. Materials

The materials in the learning phase were the same with those used in our previous study (Ding et al., 2017b). Sixty-six two-character pseudowords which can be pronounced but lack meaning served as novel words. They were embedded in two-sentence contexts to represent known concepts, half of which are concrete and the other half abstract. The novel words in the first sentences were always at the sentence-final position (see Table 1 for an example of stimuli). According to the pretest results (Table 2), most of participants could infer the meaning of the novel words in the two conditions with no difference between them [$t_{(64)} = 0.20$, $p = 0.840$]. In addition, most of participants expected the corresponding concepts of the novel words at the terminal positions of the first sentences with equally high cloze probabilities [$t_{(64)} = -0.10$, $p = 0.924$]. The results showed that the sentence constraints were not significantly different for concrete and abstract conditions, therefore excluding the potential influence of sentence constraint on the concreteness effect (Taylor et al., 2019; Bechtold et al., 2022). In addition, considering the potential influence of sentence complexity on word recognition, we used sentences with simple structures and both the numbers of characters [Concrete: *Mean* (*SD*) = 33.67 (2.34); Abstract: *Mean* (*SD*) = 33.64 (2.66), $t_{(64)} = 0.05$, $p = 0.961$] and words [Concrete: *Mean* (*SD*) = 12.00 (1.09); Abstract: *Mean* (*SD*) = 11.94 (1.20), $t_{(64)} = 0.22$, $p = 0.830$] were matched between the concrete and abstract conditions.

The lexical decision task was performed in the semantic priming paradigm, with the novel words serving as primes. Three types of real words: the corresponding concepts of novel words (CC), thematically related words (TR), and unrelated words (UR) and three pseudowords served as target words. Sixteen participants (mean age 22.7 years, range 19-27, 8 males) rated the semantic relatedness of the TR and UR targets to the CC targets on a 7-point Likert scale (7 indicates the most closely related and 1 indicates unrelated). Since the novel words were pseudowords, their concreteness was assigned based on that of their corresponding concepts. Sixteen participants (mean age 22.5 years, range 19-27, 8 males) rated the concreteness of all the target words on a 7-point Likert scale (7 indicates the most concrete, 1 indicates the most abstract). In addition, 21 participants (mean age 22.5 years, range 18-29, 11 males) rated the valence (7 indicates the most positive, 1 indicates the most negative) and arousal (7 indicates the highest arousal, 1 indicates the lowest arousal) of all the target words on 7-point Likert scales. Meanwhile, word frequency (Cai and Brysbaert, 2010) and number of strokes of all the target words were calculated. The descriptive results of all the aforementioned pretests are presented in Table 2. As can be seen in Table 3, the statistical analyses results indicated that the TR targets were more semantically related to the CC targets than the UR targets irrespective of the conceptual concreteness. In addition, the concrete words are rated to be more concrete than the abstract words irrespective of the target condition. Besides, all the target words were not significantly different in valence, arousal, word frequency, or number of strokes.

Meanwhile, for the TR targets, we calculated their cloze probabilities in the first sentences and their inferring probabilities in the whole discourses. Due to the non-normal distribution of the ratios, we conducted Mann-Whitney tests and found that the cloze probabilities [$n = 66$, $z = -1.00$, $p = 0.317$] or the inferring probabilities

TABLE 2 Means (*SD*s) of the stimuli properties.

	Corresponding concept		Thematically related word		Unrelated word	
	Abstract	Concrete	Abstract	Concrete	Abstract	Concrete
Relatedness	–	–	5.97 (0.47)	5.95 (0.63)	2.03 (0.38)	1.86 (0.50)
Concreteness	2.73(0.56)	6.15 (0.55)	2.74 (0.64)	5.95 (0.61)	2.53 (0.42)	6.03 (0.55)
Valence	4.55 (0.95)	4.57 (0.81)	4.62 (0.83)	4.45 (1.00)	4.24 (0.54)	4.55 (0.71)
Arousal	3.61 (0.87)	3.43(0.80)	3.50 (0.96)	3.43 (1.00)	3.37 (0.96)	3.45 (1.18)
Word frequency	2.52 (0.89)	2.57 (0.97)	2.62 (0.84)	2.26 (0.75)	2.34 (0.86)	2.51 (0.73)
Number of strokes	16.72 (5.10)	15.61 (4.77)	16.89 (5.19)	17.06 (4.87)	16.09 (4.08)	17.00 (3.62)
Cloze probability	81.06% (20.86%)	80.80% (21.60%)	0	0.25% (1.45%)		
Inferring probability	96.72% (5.49%)	96.97% (4.57%)	0	1.52% (8.70%)		

Cloze probability represents the proportion of the participants who gave the target words to complete the first sentences which were presented without the last words. Inferring probability represents the proportion of the participants who gave the target words as the meaning of novel words. SD: standard deviation.

TABLE 3 *F*-values of the ANOVAs on the stimuli properties.

	Relatedness	Concreteness	Valence	Arousal	Frequency	Strokes
Target condition	1662.07***	1.46	0.85	0.23	1.03	0.50
Conceptual concreteness	1.52	1502.16***	0.17	0.16	0.02	< 0.001
Target condition by Conceptual concreteness	0.51	1.32	1.72	0.31	2.24	0.83

The dfs of target condition and the interaction between target condition and conceptual concreteness are (1, 64) for relatedness, and (2, 128) for other properties. The df of conceptual concreteness is (1, 64) for all the properties. *** Significant at.001 level.

[n = 66, z = -1.00, p = 0.317] of the TR targets were not significantly different between the concrete and abstract conditions, indicating that participants learned the correct meaning of novel words.

2.4. Procedure

Participants who were not recruited in any pretests took part in EEG experiment. The 66 discourses were divided into six blocks with each of three blocks including 10 discourses (five concrete and five abstract) and each of the other three blocks including 12 discourses (six concrete and six abstract). In each block, participants read the contexts in a pseudo-random order with no more than three discourses of the same condition being presented consecutively. Following a 1,000-ms fixation cross in the center of the screen, the contexts were presented one word or two-word phrase at a time (500-ms duration, 300-ms inter-stimulus interval (ISI)). The novel words were always presented in isolation for 1,000 ms. The two-sentence discourse was presented again on the screen after the last phrase and participants could press the space button if they had inferred the meaning of the novel words. After a 2000-ms resting screen, the next trial began.

After reading all 10 or 12 discourses in a block, participants performed a lexical decision task. Following a 1,000-ms fixation cross in the center of the screen, a prime word and a target word were presented in serial for 300 ms with a 200-ms ISI. Participants pressed the “F” or “J” on the keyboard to indicate whether the target word was a real word or not as quickly and accurately as possible. The button press was counterbalanced across participants. The word pairs were presented in a pseudo-random order: the same trial type was not presented consecutively more than three times and word pairs containing the same novel word were not presented in succession.

There was a short break between blocks. Approximately 20 min after all the blocks, participants performed a cued-recall memory task in which they were asked to write down the meaning of novel words learned in the experiment. Before the EEG experiment, participants had been told that there would be a memory test for the novel words.

2.5. Electrophysiological recording and Preprocessing

EEG was recorded (0.05–100 Hz, sampling rate 500 Hz) during contextual reading and lexical decision task with 64 Ag/AgCl electrodes, with the online reference linked to the right mastoid electrode. Vertical blinks and horizontal eye-movements were monitored. All electrodes impendence was kept below 5 kΩ. With NeuroScan software 4.5, automatic correction of the ocular artifacts (Semlitsch et al., 1986) and a band-pass filter at 0.1–30 Hz were conducted. Then the EEG data were segmented into 1,200-ms epochs from 200 ms before to 1,000 ms after the onset of critical words. After baseline correction with the mean amplitudes of prestimulus interval, an artifact correction was performed with criteria of ±80 μV at all electrodes except for electrooculograms, and the ERPs were re-referenced offline to the algebraic average of two mastoids.

2.6. ERP data analysis

Based on previous studies (Ding et al., 2017b) and visual inspections, 300–400 ms (for learning phase)/300–500 ms (for testing phase) and 500–1,000 ms (for both learning and testing phases) time windows were selected for the N400 and late components, respectively. For the learning phase, conceptual concreteness (abstract, concrete), presentation time (first, second), laterality (left, middle, right), and anteriority (anterior, central, posterior) served as within-subject factors in repeated measures ANOVAs. For the testing phase, conceptual concreteness (abstract, concrete), target condition (CC, TR, UR), laterality and anteriority were taken as within-subject factors in repeated measures ANOVAs. The combination of laterality and anteriority resulted in nine regions of interest with each containing six electrodes: left anterior (F3, F5, F7, FC3, FC5, and FT7), left central (C3, C5, T7, CP3, CP5, and TP7), left posterior (P3, P5, P7, PO5, PO7, and O1), middle anterior (F1, FZ, F2, FC1, FCZ, and FC2), middle central (C1, CZ, C2, CP1, CPZ, and CP2), middle posterior (P1, PZ, P2, PO3, POZ, and PO4), right anterior (F4, F6, F8, FC4, FC6, and FT8), right central (C4, C6, T8, CP4, CP6, and TP8), and right posterior (P4, P6, P8, PO6, PO8, and O2). In addition, for the pair-wise comparisons involving factors with three levels, corrected *p* values were reported with Bonferroni correction.

3. Results

3.1. Behavioral data

Due to the non-normal distribution of the data, generalized linear mixed-effects model analyses were conducted in R (R Core Team, 2013) using glmer function for the reaction time (RT) with a logarithmic link function (Lo and Andrews, 2015) and accuracy (ACC). Meanwhile, hypr package (Rabe et al., 2020) was used to construct custom contrast coding mainly for the pair-wise comparisons for the target condition. In the models, conceptual concreteness, target condition and their interaction were included as fixed effect factors. The models were initially constructed with maximal random effect structures. When the models failed to converge, we simplified the random structure according to Bates et al. (2015). The final models for the RT and ACC were the same:

$$\text{RT} / \text{ACC} \sim \text{concreteness}^{*}\text{condition} + (1|\text{subject}).$$

For the reaction time, error trials (4.31%) and outlier data points (beyond 2.5 *SD*s as well as shorter than 300 ms or longer than 1,500 ms, 4.88%) were excluded from the analysis. Figure 1A presents the reaction time of the remained trials in each condition. Table 4 presents the results of custom contrast coding for the reaction time. The results showed that participants responded fastest to the CC targets, then to the TR targets and slowest to the UR targets. Neither the main effect of conceptual concreteness nor the interaction between conceptual concreteness and target condition were significant.

Table 5 presents the results of custom contrast coding for the accuracy. As also can be seen from Figure 1B, the accuracy decreased in a graded manner from the CC targets, to the TR targets, finally to the UR targets. Similar to the results for the reaction time, the main effect of conceptual concreteness or the interaction between conceptual concreteness and target condition were not significant.

In the cued-recall memory test, participants recollected the meaning of concrete novel words better than that of abstract novel words [Concrete: *Mean* (*SD*) = 12.29 (10.93); Abstract: *Mean* (*SD*) = 9.38 (9.88), $t_{(23)} = 4.34$, $p < 0.001$].

3.2. ERP data

The grand average waveforms elicited by the novel words in the learning phase at CZ electrode are presented in Figure 2. For the learning phase, in the time window of 300–400 ms, we found a significant interaction between concreteness and time [$F_{(1,23)} = 11.23$, $p = 0.003$, $\eta_p^2 = 0.328$]. The following simple-effects tests showed that the abstract novel words elicited smaller N400s for the first time than the second time [$F_{(1,23)} = 6.60$, $p = 0.017$, $\eta_p^2 = 0.223$]. This effect was absent for the concrete novel words [$F_{(1,23)} = 0.01$, $p = 0.910$, $\eta_p^2 = 0.001$]. Meanwhile, the abstract novel words elicited larger N400s than the concrete novel words for the first time [$F_{(1,23)} = 4.47$, $p = 0.045$, $\eta_p^2 = 0.163$], but not for the second time [$F_{(1,23)} = 4.25$, $p = 0.051$, $\eta_p^2 = 0.156$]. In the time window of 500–1,000 ms, there was a significant main effect of time [$F_{(1,23)} = 43.38$, $p < 0.001$, $\eta_p^2 = 0.653$] and a significant interaction between time and hemisphere [$F_{(2,46)} = 5.72$, $p = 0.006$, $\eta_p^2 = 0.199$]. The following ANOVAs showed that the novels words elicited larger late positivities for the first time than the second

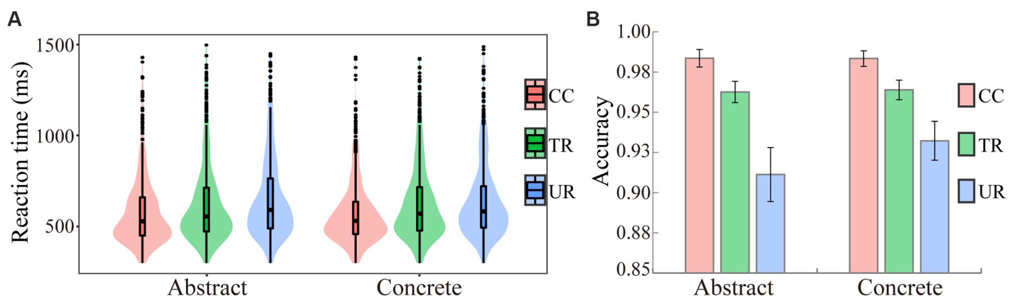


FIGURE 1
The reaction times of the correct responses **(A)** and the accuracies **(B)** for target words in each condition. CC: corresponding concepts; TR: thematically related words; UR: unrelated words.

TABLE 4 GLMM results for the reaction time.

	Estimate	Std. Eror	*t* value	Pr (>|t|)
(Intercept)	−1.78	0.120	−14.78	< 2e-16 ***
Conceptual concreteness	0.012	0.011	1.16	0.245
CC vs. UR	−0.182	0.013	−13.95	< 2e-16 ***
TR vs. UR	−0.056	0.012	−4.39	1.16e-05 ***
CC vs. TR	−0.126	0.013	−9.69	< 2e-16 ***
Abstract (CC-UR) vs. Concrete (CC-UR)	−0.002	0.013	−0.22	0.830
Abstract (TR-UR) vs. Concrete (TR-UR)	−0.022	0.013	−1.68	0.823
Abstract (CC-TR) vs. Concrete (CC-TR)	0.019	0.013	1.45	0.147

CC: corresponding concepts; TR: thematically related words; UR: unrelated words.

TABLE 5 GLMM results for the accuracy.

	Estimate	Std. Eror	z value	Pr (>|z|)
(Intercept)	3.48	0.17	19.97	< 2e-16 ***
Conceptual concreteness	−0.11	0.17	−0.65	0.516
CC vs. UR	1.63	0.22	7.49	7.13e-14 ***
TR vs. UR	0.81	0.16	4.95	7.28e-07 ***
CC vs. TR	0.81	0.24	3.43	0.000598 ***
Abstract (CC-UR) vs. Concrete (CC-UR)	0.15	0.22	0.68	0.498
Abstract (TR-UR) vs. Concrete (TR-UR)	0.13	0.16	0.78	0.43
Abstract (CC-TR) vs. Concrete (CC-TR)	0.02	0.24	0.08	0.937866

CC: corresponding concepts; TR: thematically related words; UR: unrelated words.

time in the left [$F_{(1,23)} = 83.33$, $p < 0.001$, $\eta_p^2 = 0.784$], middle [$F_{(1,23)} = 29.30$, $p < 0.001$, $\eta_p^2 = 0.560$], and right [$F_{(1,23)} = 16.60$, $p < 0.001$, $\eta_p^2 = 0.419$] regions.

Figure 3 presents the grand average waveforms elicited by the target words at CZ electrode. For the testing phase, in the N400 time window, we found a significant main effect of target condition [$F_{(2,46)} = 16.63$, $p < 0.001$, $\eta_p^2 = 0.420$]. Pair-wised comparisons revealed that the CC targets elicited the smallest N400s, the TR targets the moderate, and the UR targets the largest [CC vs. UR: $t_{(23)} = 4.52$, $p < 0.001$; TR vs. UR: $t_{(23)} = 3.08$, $p = 0.016$; CC vs. TR: $t_{(23)} = 3.70$, $p = 0.004$]. We also found a significant main effect of concreteness [$F_{(1,23)} = 20.65$, $p < 0.001$, $\eta_p^2 = 0.473$], with the concrete condition eliciting larger N400s than the abstract condition. The interactions between concreteness and hemisphere [$F(2,46) = 3.52$, $p = 0.038$, $\eta_p^2 = 0.133$], as well as between concreteness and anteriority [$F_{(2,46)} = 16.14$, $p < 0.001$, $\eta_p^2 = 0.412$] were also significant. The following ANOVAs showed that the concrete targets elicited larger N400s than the abstract targets over the whole brain regions [Left: $F_{(1,23)} = 10.73$, $p = 0.003$, $\eta_p^2 = 0.318$; Middle: $F_{(1,23)} = 19.41$, $p < 0.001$, $\eta_p^2 = 0.458$; Right: $F_{(1,23)} = 18.91$, $p < 0.001$, $\eta_p^2 = 0.451$; Anterior: $F_{(1,23)} = 32.08$, $p < 0.001$, $\eta_p^2 = 0.582$; Central: $F_{(1,23)} = 18.07$, $p < 0.001$, $\eta_p^2 = 0.440$] except the posterior region [$F_{(1,23)} = 2.66$, $p = 0.117$, $\eta_p^2 = 0.104$]. In the time window of 500–1,000 ms, only a significant main effect of concreteness was found [$F_{(1,23)} = 8.68$, $p = 0.007$, $\eta_p^2 = 0.274$], with the abstract targets eliciting larger late positivities than the concrete targets.

4. Discussion

The current study investigated the influence of conceptual concreteness on reading acquisition and integration of novel words into semantic memory. During contextual reading, the abstract novel words elicited a larger N400 than the concrete novel words for the first presentation. In the memory test, concrete novel words were better recollected than abstract novel words. These results indicated that the meaning of abstract novel words was derived more difficultly and more susceptible to forgetting than that of concrete ones. During the testing phase, both behavioral and ERP results revealed semantic integration of novel words into semantic network via thematic relations, as indicated by the graded RTs and accuracy as well as N400s for the CC, TR and UR targets, regardless of conceptual concreteness.

4.1. Modulation of concreteness on reading acquisition of novel words

In contextual reading, the abstract novel words elicited a larger N400 than the concrete novel words for the first presentation. N400 is a well-established component, indicating the easiness of semantic processing such as semantic retrieval or semantic integration (Lau et al., 2008; Kutas and Federmeier, 2011). Semantic violation would elicit a larger N400 during language comprehension (Kutas and Hillyard, 1980). When participants encountered the novel words in the first sentences, they would experience a semantic violation because the CC targets were strongly expected as indicated by their high cloze probabilities. Therefore, participants should retrieve semantic information (i.e., the CC targets) in semantic memory for the novel words. Due to the more consistency for words with more sensory-based representations and more variations for words with more language-based representations (Wang and Bi, 2021), the larger N400s for the abstract than concrete novel words indicated more difficulty in

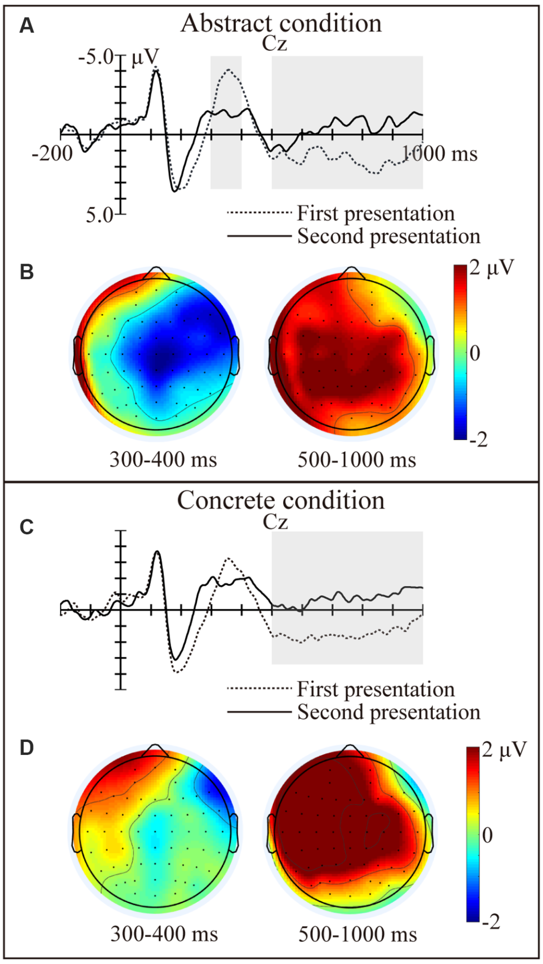


FIGURE 2
ERP results in the learning phase. Grand average waveforms elicited by the novel words in the learning phase in the abstract condition **(A)** and concrete condition **(C)** at Cz electrode. Topographies showing the N400 and LPC effects (first presentation vs. second presentation) in the abstract condition **(B)** and concrete condition **(D)**. The solid lines represent the novel words presented for the first time. The dotted lines represent the novel words presented for the second time.

deriving the meaning of abstract novel words. This result was in line with the prediction of the dual coding theory that the concrete concepts would be better retrieved than abstract concepts with the support of both the verbal and imagery systems (Paivio, 1986, 1991).

However, this result seems to contradict the pretest results that showed no difference between the predictabilities of the corresponding concepts of concrete and abstract novel words (80.80 and 81.06% for concrete and abstract conditions, respectively). This contradiction between the ERPs and behavioral measures might be attributed to the different sets of cognitive operations which were caught by the two techniques. The ERPs reflect the online semantic processing of the stimuli, while the behavioral measures are associated with a series of cognitive processes, such as semantic access, decision making and motor operations within a relatively long-time duration (Holcomb, 1993; Chen et al., 2014). Therefore, without time pressure, the difficulty of deriving meaning of abstract novel words may not be reflected by the cloze probability test.

In addition, the novel words presented for the first time elicited larger late positivities than the second time in both conditions. This later effect, in terms of LPC or P600, in sentence comprehension is associated with semantic re-analysis of the perceptual input which is in conflict with the contextual prediction (Kuperberg, 2007; Brouwer et al., 2012; Kuperberg et al., 2020). As discussed above, the novel words in the first sentences were semantically anomalous compared with the corresponding concepts of novel words which were predicted based on the preceding contexts. After retrieving the meaning of novel words, participants needed to re-analyze the novel words by associating the corresponding concepts to them to make sense of the sentences. When the novel words were encountered at the second sentences, this semantic re-analysis was easier relative to that in the first sentences, resulting in smaller late positivities. In particular, the later effect was not influenced by conceptual concreteness. We postulated that the difficulty in semantic retrieval might have been resolved in the N400 time window. After the corresponding concepts were retrieved, the semantic re-analysis was not that demanding with no differences between the concrete and abstract contexts due to the strong constraints in both conditions. This later effect, to some extent, can be considered as a supportive evidence of context availability hypothesis in contextual reading at electrophysiological level. Combined with the N400 results, the ERPs in contextual reading indicate a cognitive neurodynamic processing of concrete and abstract words during sentence comprehension.

The influence of conceptual concreteness manifested in the lexical decision task. The concrete target words elicited larger N400s and smaller late positivities than the abstract ones, regardless of the semantic relatedness with the prime words. Previous studies have reported this larger long-lasting negativities for concrete words than for abstract words in a variety of processing tasks (Kounios and Holcomb, 1994; Holcomb et al., 1999; West and Holcomb, 2000; Kanske and Kotz, 2007; Huang et al., 2010; Adorni and Proverbio, 2012; Barber et al., 2013; Palmer et al., 2013; Bechtold et al., 2018, 2022). The long-lasting negativity effect includes an initial N400 component and a late N700 component. Holcomb et al. (1999) extended the dual coding theory and the context availability theory and proposed that concrete concepts are supported by superior linguistic information and an additional imagery-based system relative to abstract concepts. The N400 effect in response to conceptual concreteness is proposed to reflect stronger involvement of semantic activation or integration for concrete (vs. abstract) words which is in line with context availability theory; consistent with the dual coding theory, the N700 effect is associated with enhanced retrieval of visual imagery information (Kounios and Holcomb, 1994; Holcomb et al., 1999; Bechtold et al., 2018, 2022). It is worth noting that, in line with previous studies, the current study obtained a fronto-central N400 effect in response to conceptual concreteness in single-word processing (Barber et al., 2013; Palmer et al., 2013). While words were accompanied by contexts (Holcomb et al., 1999; West and Holcomb, 2000; Bechtold et al., 2022), the N400 effect distributed broadly or centro-parietally, which might be attributed to the high demands of integrating the words into contexts (Bechtold et al., 2022).

In addition, different from the ERP results, in the lexical decision task the behavioral results did not show the concreteness effect, that is, no difference in RTs or accuracies between the concrete and abstract conditions. These results are in line with the context availability hypothesis which predicts that the concreteness effect

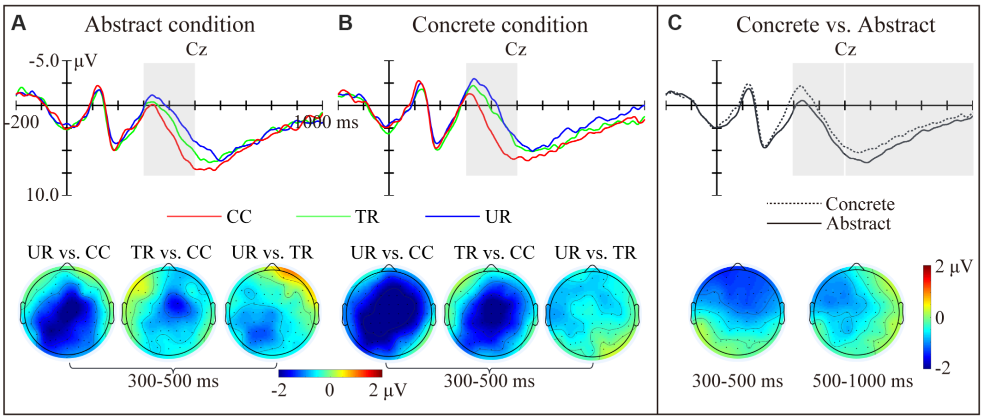


FIGURE 3
ERP results in the lexical decision task. Grand average waveforms elicited by the target words and topographies showing the N400 effects in the abstract condition **(A)** and concrete condition **(B)** at Cz electrode. Panel **(C)** presents the grand average waveforms elicited by the concrete and abstract targets at Cz electrode, as well as the topographies showing the N400 and LPC effects in response to the conceptual concreteness. The red lines represent the corresponding concepts (CC). The green lines represent the thematically related words (TR). The blue lines represent the unrelated words (UR). The solid line represents the concrete condition. The dotted line represents the abstract condition.

would disappear in the case words were embedded in contexts, especially in highly constraining contexts (Schwanenflugel et al., 1988; van Hell and de Groot, 2008). It was shown that highly constraining sentences in the current study facilitated the semantic access of both the concrete and abstract novel words, and in turn facilitated the implicit semantic processing of the target words. However, in the explicit semantic processing, namely cued-recall memory task, the performance for the concrete words was better than the abstract words. This may be ascribed to the strategies that participants used in the task. At the beginning of the experiment, they were told that there would be a memory test for the novel words. In this case they had to try to retain the corresponding concepts of novel words. As discussed above, in the support of (extended) verbal and imagery information for the concrete concepts, concrete novel words would be recollected better than abstract ones. It is due to the memory test, participants would be more sensitive to the conceptual concreteness, resulting in the N400 and N700 effects for the target words in the current study. Taken together, all these results suggest a modulation of conceptual concreteness on reading acquisition of novel word.

4.2. Similar semantic integration of abstract and concrete novel words

In the lexical decision task, the CC targets elicited the smallest N400s than other targets, and participants responded fastest to the CC targets with the highest accuracy. These results, along with the observed effects in the learning phase, indicated that the meaning of novel words was acquired in contextual reading, in line with previous studies exploring novel word learning in contextual reading (Mestres-Missé et al., 2007; Borovsky et al., 2012, 2013; Chen et al., 2014; Ding et al., 2017a,b; Zhang et al., 2017, 2019).

In addition, compared to the UR targets the TR target elicited smaller N400s with shorter RT and higher accuracy, indicating the establishment of thematic relations between novel and existing words in semantic memory. Using concrete words as stimuli, previous studies found that novel words could be integrated into semantic network through thematic relations (Zhang et al., 2017, 2019). In addition, this kind of semantic integration was modulated by the richness of learning context content. In contexts describing a single episode, novel words could only connect with words through thematic relation gained in the learning episodes; while in multiple episodic contexts, could connect with more words through extended thematic relations, even not implicated in the learning contexts (Zhang et al., 2019). In the current study, two sentences in a context depicted two episodes for the novel word, and the thematically related word in the testing phase was not related to the learning episodes. Therefore, the results of the present study did replicate the previous study in the concrete condition (i.e., Zhang et al., 2019).

Most importantly, it was found that the integration of novel words was not modulated by conceptual concreteness. In other words, both concrete and abstract novel words were integrated into semantic memory in a similar way through thematic relations. The findings are not predicted by different representational frameworks theory, which posits that abstract concepts connect with other concepts mainly via thematic relations while concrete concepts via taxonomic relations. According to Crutch and Warrington (2010), the theory was a graded but not binary model. Specifically, both concrete and abstract concepts connect with other concepts through similarity-based (taxonomic) and associative (thematic) information; the dependence on taxonomic or thematic relations varies along the continuum of concreteness, with more concrete concepts being more dependent on taxonomic representation and more abstract concepts more on thematic organization. In the current study, the corresponding concepts of concrete and abstract novel words were significantly different in conceptual concreteness on a 7-point Likert scale (6.15 and 2.73 for concrete and abstract conditions, respectively). The abstract novel words would connect with known words in semantic memory relying on much more thematic relations than the concrete novel words, which would be reflected by different semantic priming effects for the thematically related target words in abstract and concrete conditions. However, the observed smaller N400s for the thematically related (vs. unrelated) words did not interact with conceptual concreteness.

Previous studies have found that novel words learned in different types of contexts are integrated into semantic memory via different sematic relations (Ding et al., 2017a; Zhang et al., 2017, 2019). For instance, novel words are integrated into semantic memory either in contexts describing semantic features of concepts via taxonomic relations (Ding et al., 2017a) or in episodic contexts depicting the events or episodes via thematic relations (Zhang et al., 2017). One might argue that the similar integration through thematic relations for concrete and abstract novel words could be due to the learning episodic contexts, which emphasized the thematic associations among the concepts in the events. Nevertheless, a subsequent study has shown that novel words learned from episodic contexts could be also integrated into semantic memory via taxonomic relations (Zhang et al., 2019). Especially, our previous study using the same learning contexts with the current study found similar semantic integration of concrete and abstract novel words via taxonomic relations (Ding et al., 2017b). The results of our previous and current studies could not provide evidence for an asymmetric semantic integration of concrete and abstract concepts into semantic memory via taxonomic and thematic relations. Although the concrete novel words were learned better than the abstract ones, as predicted by the (extended) dual coding theory, the concrete and abstract novel words could establish both taxonomic and thematic relations with semantic memory via episodic contexts. Combined with the learning effects, it can be concluded that the conceptual concreteness only influences the learning processes of novel words but not the way the novel words are integrated into semantic memory.

However, the similar integration pattern of concrete and abstract novel words could not completely rule out the different dependence on taxonomic and thematic relations for concrete and abstract concepts due to the following reason. In the lexical decision task, only one word which was typically thematically related to the novel word was selected as stimuli. Due to the balance between semantic relatedness of the TR target words to both concrete and abstract novel words (CC target words in the pretest), the observed semantic priming effects were not significantly different in the two conditions. If using the blocked-translation paradigm in which a series of thematically related words served as competitors (e.g., Zhang et al., 2013), the semantic integration of novel words as indicated by semantic interference might be different between the concrete and abstract conditions.

4.3. Implications on L2 reading acquisition

The contextual learning paradigm used in the current and previous studies (e.g., Mestres-Missé et al., 2007, 2008; 2010; Borovsky et al., 2010; Chen et al., 2014; Ding et al., 2017a,b; Zhang et al., 2017, 2019; Liu et al., 2019) associated novel words with known concepts in semantic memory. This paradigm can be treated as a simulation of L2 vocabulary leaning, in which the concepts of novel words already exist in learner's semantic memory (Mestres-Missé et al., 2007, 2014; Ferreira et al., 2015). Based on the findings of the current and our previous study (Ding et al., 2017b), the novel words could be learned rapidly and integrated into semantic memory via both taxonomic and thematic relations irrespective of conceptual concreteness. In addition, Zhang et al. (2019) found that in multiple episodic contexts novel words could be integrated into semantic memory via extended thematic relations and taxonomic relations. Taken together, all these results indicate that embedding words in episodic contexts, especially in multiple episodic contexts, will facilitate the learning efficiency for L2 vocabulary learning.

However, the generalization of the conclusion should be cautious. First, although previous studies revealed rapid acquisition of novel words in contextual learning paradigm in different languages, such as English (e.g., Borovsky et al., 2010, 2012, 2013), Spanish (e.g., Mestres-Missé et al., 2007, 2014), Chinese (e.g., Ding et al., 2017a,b; Zhang et al., 2017, 2019) and so on, only in Chinese the modulations of contextual information on the integration of novel words into semantic memory were examined here. Whether the facilitation effect of multiple episodic contexts on L2 vocabulary learning could be found in various languages needs to be further addressed. Second, individual differences such as age might modulate the observed effects. The weight of taxonomic and thematic relations in conceptual system organization changes during life span development. For instance, during children's growing up, there is a developmental shift from thematic to taxonomic thinking; and thematic thinking gains more weight for aging people (for a review see Mirman et al., 2017). The participants recruited in the current study were adults which may rely on both taxonomic and thematic relations. Third, there has been evidence for the modulation of sentential constraint and L2 proficiency level on L2 word learning (Pulido, 2003; Borovsky et al., 2010; Elgort et al., 2015; Ma et al., 2015, 2016; Chen et al., 2017). Only in highly constraining sentences could the meaning of L2 novel words be derived (Ma et al., 2015), and higher-proficiency L2 learners outperformed than lower-proficiency L2 learners (Ma et al., 2015; Chen et al., 2017). Most of sentences used in these studies are highly constraining and presented in L1 for participants, who could be treated as high proficiency L2 language learners. Therefore, using weakly constraining sentences and recruiting L2 learners with different language proficiency levels and ages to investigate reading acquisition and semantic integration of novel words would increase the generalizability of the current conclusion in future studies. In addition, future studies are also needed due to the following limitation of the current study. In order to compare the current study with our previous study (i.e., Ding et al., 2017b), we used the same learning procedure in which the same novel word was repeated six times to each participant. Although the presenting order was counterbalanced across conditions, the repetition effects could not be completely excluded. Constructing more items and using a Latin square design will increase the robustness of the current conclusion.

5. Conclusion

The current study explored the modulation of conceptual concreteness on reading acquisition and semantic integration of novel words using contextual learning paradigm. During contextual reading, the meaning of abstract novel words was inferred with more difficulty than concrete novel words. In a cued-recall memory test assessing the learning effects of novel words, concrete novel words were preserved better than abstract ones. All these results, consistent with the (context-extended) dual coding theory, indicate the support of verbal and imagery information for concrete words processing. Meanwhile, both concrete and abstract novel words could construct thematic relations with known concepts in semantic memory, which is in conflict with the different representational frameworks theory under this circumstance (i.e., in lexical decision task). The results suggest a modulation of conceptual concreteness on reading acquisition but not on the type of semantic relations through which novel words are integrated into semantic memory.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JD: formal analysis, data curation, writing–original draft, writing–review and editing, and funding acquisition. PL: methodology, software, investigation, and visualization. XG: investigation and writing–review and editing. YY: conceptualization, writing–review and editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31900762 and 61433018).

Acknowledgments

We deeply appreciated Chen Feng for his help in data analyses and the excellent work of the technical support staff at the Institutional Center for Shared Technologies and Facilities of Institute of Psychology, Chinese Academy of Sciences.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adorni, R., and Proverbio, A. M. (2012). The neural manifestation of the word concreteness effect: an electrical neuroimaging study. *Neuropsychologia* 50, 880–891. doi: 10.1016/j.neuropsychologia.2012.01.028

Barber, H. A., Otten, L. J., Kousta, S. T., and Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain Lang.* 125, 47–53. doi: 10.1016/j.bandl.2013.01.005

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. ArXiv150604967 Stat. Available at: http://arxiv.org/abs/1506.04967

Batterink, L., and Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *J. Cogn. Neurosci.* 23, 3181–3196. doi: 10.1162/jocn_a_00013

Bechtold, L., Bellebaum, C., and Ghio, M. (2022). When a sunny day gives you butterflies: an electrophysiological investigation of concreteness and context effects in semantic word processing. *J. Cogn. Neurosci.* 35, 241–258. doi: 10.1162/jocn_a_01942

Bechtold, L., Bellebaum, C., Hoffman, P., and Ghio, M. (2021). Corroborating behavioral evidence for the interplay of representational richness and semantic control in semantic word processing. *Sci. Rep.* 11:6184. doi: 10.1038/s41598-021-85711-7

Bechtold, L., Ghio, M., and Bellebaum, C. (2018). The effect of training-induced visual imageability on electrophysiological correlates of novel word processing. *Biomedicine* 6. :75. doi: 10.3390/biomedicines6030075

Bi, Y. (2021). Dual coding of knowledge in the human brain. *Trends Cogn. Sci.* 25, 883–895. doi: 10.1016/j.tics.2021.07.006

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychon. Bull. Rev.* 23, 1096–1108. doi: 10.3758/s13423-015-0909-1

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536. doi: 10.1016/j.tics.2011.10.001

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., and Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *J. Cogn. Neurosci.* 17, 905–917. doi: 10.1162/0898929054021102

Borghi, A. M. (2022). Concepts for which we need others more: the case of abstract concepts. *Curr. Dir. Psychol. Sci.* 31, 238–246. doi: 10.1177/09637214221079625

Borovsky, A., Elman, J. L., and Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Lang. Learn. Dev.* 8, 278–302. doi: 10.1080/15475441.2011.614893

Borovsky, A., Kutas, M., and Elman, J. L. (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition* 116, 289–296. doi: 10.1016/j.cognition.2010.05.004

Borovsky, A., Kutas, M., and Elman, J. L. (2013). Getting it right: word learning across the hemispheres. *Neuropsychologia* 51, 825–837. doi: 10.1016/j.neuropsychologia.2013.01.027

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055

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

Capitani, E., Laiacona, M., Mahon, B., and Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn. Neuropsychol.* 20, 213–261. doi: 10.1080/02643290244000266

Chen, B., Ma, T., Liang, L., and Liu, H. (2017). Rapid L2 word learning through high constraint sentence context: an event-related potential study. *Front. Psychol.* 8:e2285. doi: 10.3389/fpsyg.2017.02285

Chen, S., Wang, L., and Yang, Y. (2014). Acquiring concepts and features of novel words by two types of learning: direct mapping and inference. *Neuropsychologia* 56, 204–218. doi: 10.1016/j.neuropsychologia.2014.01.012

Crutch, S. J., Connell, S., and Warrington, E. K. (2009). The different representational frameworks under-pinning abstract and concrete knowledge: evidence from odd-one-out judgements. *Q. J. Exp. Psychol.* 62, 1377–1390. doi: 10.1080/17470210802483834

Crutch, S. J., Ridha, B. H., and Warrington, E. K. (2006). The different frameworks underlying abstract and concrete knowledge: evidence from a bilingual patient with a semantic refractory access dysphasia. *Neurocase* 12, 151–163. doi: 10.1080/13554790600598832

Crutch, S. J., and Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain* 128, 615–627. doi: 10.1093/brain/awh349

Crutch, S. J., and Warrington, E. K. (2007). Semantic priming in deep-phonological dyslexia: contrasting effects of association and similarity upon abstract and concrete word reading. *Cogn. Neuropsychol.* 24, 583–602. doi: 10.1080/02643290701577351

Crutch, S. J., and Warrington, E. K. (2010). The differential dependence of abstract and concrete words upon associative and similarity-based information: complementary semantic interference and facilitation effects. *Cogn. Neuropsychol.* 27, 46–71. doi: 10.1080/02643294.2010.491359

De Groot, A. M., and Keijzer, R. (2000). What is hard to learn is easy to forget: the roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Lang. Learn.* 50, 1–56. doi: 10.1111/0023-8333.00110

Ding, J., Chen, S., Wang, L., and Yang, Y. (2017a). Thematic and taxonomic relations of novel words learned from action and perceptual features. *J. Neurolinguistics* 41, 70–84. doi: 10.1016/j.jneuroling.2016.10.002

Ding, J., Liu, W., and Yang, Y. (2017b). The influence of concreteness of concepts on the integration of novel words into the semantic network. *Front. Psychol.* 8:e02111. doi: 10.3389/fpsyg.2017.02111

Duñabeitia, J. A., Aviles, A., Afonso, O., Scheepers, C., and Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: evidence from the visual-world paradigm. *Cognition* 110, 284–292. doi: 10.1016/j.cognition.2008.11.012

Elgort, I., Perfetti, C. A., Rickles, B., and Stafura, J. Z. (2015). Contextual learning of L2 word meanings: second language proficiency modulates behavioural and ERP indicators of learning. *Lang. Cogn. Neurosci.* 30, 506–528. doi: 10.1080/232737982014.942673

Elgort, I., and Warren, P. (2014). L2 vocabulary learning from reading: explicit and tacit lexical knowledge and the role of learner and item variables. *Lang. Learn.* 64, 365–414. doi: 10.1111/lang.12052

Estes, Z., Golonka, S., and Jones, L. L. (2011). Thematic thinking: the apprehension and consequences of thematic relations. *Psychol. Learn. Motiv. Adv. Res. Theor.* 54, 249–294. doi: 10.1016/B978-0-12-385527-5.00008-5

Ferreira, R. A., Göbel, S. M., Hymers, M., and Ellis, A. W. (2015). The neural correlates of semantic richness: evidence from an fMRI study of word learning. *Brain Lang.* 143, 69–80. doi: 10.1016/j.bandl.2015.02.005

Geng, J., and Schnur, T. T. (2015). The representation of concrete and abstract concepts: categorical versus associative relationships. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 22–41. doi: 10.1037/a0037430

Hamilton, A. C., and Coslett, H. B. (2008). Refractory access disorders and the organization of concrete and abstract semantics: do they differ? *Neurocase* 14, 131–140. doi: 10.1080/13554790802032218

Hill, F., Korhonen, A., and Bentz, C. (2014). A quantitative empirical analysis of the abstract/concrete distinction. *Cogn. Sci.* 38, 162–177. doi: 10.1111/cogs.12076

Hoffman, P., Binney, R. J., and Ralph, M. A. L. (2015). Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex* 63, 250–266. doi: 10.1016/j.cortex.2014.09.001

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology* 30, 47–61. doi: 10.1111/j.1469-8986.1993.tb03204.x

Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742. doi: 10.1037/0278-7393.25.3.721

Huang, H. W., Lee, C. L., and Federmeier, K. D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *NeuroImage* 49, 1116–1123. doi: 10.1016/j.neuroimage.2009.07.031

Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., et al. (2000). The concreteness effect: evidence for dual coding and context availability. *Brain Lang.* 74, 103–112. doi: 10.1006/brln.2000.2340

Kanske, P., and Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Res.* 1148, 138–148. doi: 10.1016/j.brainres.2007.02.044

Kounios, J., and Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 804–823. doi: 10.1037//0278-7393.20.4.804

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063

Kuperberg, G. R., Brothers, T., and Wlotko, E. W. (2020). A tale of two Positivities and the N400: distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *J. Cogn. Neurosci.* 32, 12–35. doi: 10.1162/jocn_a_01465

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Lambon Ralph, M. A., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Liu, W., Ding, J., Li, L., and Yang, Y. (2019). Metaphorical meaning learning in contexts: an event-related potential study. *J. Neurolinguistics* 49, 57–70. doi: 10.1016/j.jneuroling.2018.08.004

Lo, S., and Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* 6:e01171. doi: 10.3389/fpsyg.2015.01171

Ma, T., Chen, R., Dunlap, S., and Chen, B. (2016). The effect of number and presentation order of high-constraint sentences on second language word learning. *Front. Psychol.* 7:e01396. doi: 10.3389/fpsyg.2016.01396

Ma, T., Chen, B., Lu, C., and Dunlap, S. (2015). Proficiency and sentence constraint effects on second language word learning. *Acta Psychol.* 159, 116–122. doi: 10.1016/j.actpsy.2015.05.014

Mestres-Missé, A., Càmara, E., Rodriguez-Fornells, A., Rotte, M., and Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *J. Cogn. Neurosci.* 20, 2153–2166. doi: 10.1162/jocn.2008.20150

Mestres-Missé, A., Münte, T. F., and Rodriguez-Fornells, A. (2009). Functional neuroanatomy of contextual acquisition of concrete and abstract words. *J. Cogn. Neurosci.* 21, 2154–2171. doi: 10.1162/jocn.2008.21171

Mestres-Missé, A., Münte, T. F., and Rodriguez-Fornells, A. (2014). Mapping concrete and abstract meanings to new words using verbal contexts. *Second. Lang. Res.* 30, 191–223. doi: 10.1177/0267658313512668

Mestres-Missé, A., Rodriguez-Fornells, A., and Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cortex* 17, 1858–1866. doi: 10.1093/cercor/bhl094

Mestres-Missé, A., Rodriguez-Fornells, A., and Münte, T. F. (2010). Neural differences in the mapping of verb and noun concepts onto novel words. *NeuroImage* 49, 2826–2835. doi: 10.1016/j.neuroimage.2009.10.018

Mirman, D., Landrigan, J.-F., and Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychol. Bull.* 143, 499–520. doi: 10.1037/bul0000092

Murphy, G. L. (2010). "What are categories and concepts?" in *The making of human concepts*. eds. D. Mareschal, P. C. Quinn and S. E. G. Lea (Oxford New York: Oxford University Press).

Paivio, A. (1986). *Mental representations: a dual coding approach*. New York: Oxford University Press.

Paivio, A. (1991). Dual coding theory: retrospect and current status. *Can. J. Psychol.* 45, 255–287. doi: 10.1037/h0084295

Palmer, S. D., Macgregor, L. J., and Havelka, J. (2013). Concreteness effects in single-meaning, multi-meaning and newly acquired words. *Brain Res.* 1538, 135–150. doi: 10.1016/j.brainres.2013.09.015

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vcabulary acquisition through reading. *Lang. Learn.* 53, 233–284. doi: 10.1111/1467-9922.00217

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* 17, 458–470. doi: 10.1016/j.tics.2013.06.004

R Core Team (2013). A language and environment for statistical computing (Version 4.2.1). Retrieved from https://www.r-project.org

Rabe, M. M., Vasishth, S., Hohenstein, S., Kliegl, R., and Schad, D. (2020). Hypr: an R package for hypothesis-driven contrast coding. *J. Open Sour. Softw.* 5:2134. doi: 10.21105/joss.02134

Schwanenflugel, P. J., Harnishfeger, K. K., and Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *J. Mem. Lang.* 27, 499–520. doi: 10.1016/0749-596X(88)90022-8

Schwanenflugel, P. J., and Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 82–102. doi: 10.1037/0278-7393.9.1.82

Schwanenflugel, P. J., and Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Read. Res. Q.* 24, 114–126. doi: 10.2307/748013

Semlitsch, H. V., Anderer, P., Schuster, P., and Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23, 695–703. doi: 10.1111/j.1469-8986.1986.tb00696.x

Taylor, R. S., Francis, W. S., Borunda-Vazquez, L., and Carbajal, J. (2019). Mechanisms of word concreteness effects in explicit memory: does context availability play a role? *Mem. Cogn.* 47, 169–181. doi: 10.3758/s13421-018-0857-x

Troyer, M., and McRae, K. (2022). Thematic and other semantic relations central to abstract (and concrete) concepts. *Psychol. Res.* 86, 2399–2416. doi: 10.1007/s00426-021-01484-8

van Hell, J. G., and de Groot, A. M. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychol.* 128, 431–451. doi: 10.1016/j.actpsy.2008. 03.010

Wang, X., and Bi, Y. (2021). Idiosyncratic tower of babel: individual differences in word-meaning representation increase as word abstractness increases. *Psychol. Sci.* 32, 1617–1635. doi: 10.1177/09567976211003

Wang, J., Conder, J. A., Blitzer, D. N., and Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 31, 1459–1468. doi: 10.1002/hbm.20950

West, W. C., and Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* 12, 1024–1037. doi: 10.1162/08989290051137558

Xu, Y., Wang, X., Wang, X., Men, W., Gao, J. H., and Bi, Y. (2018). Doctor, teacher, and stethoscope: neural representation of different types of semantic relations. *J. Neurosci.* 38, 3303–3317. doi: 10.1523/JNEUROSCI.2562-17.2018

Zhang, M., Chen, S., Wang, L., Yang, X., and Yang, Y. (2017). Episodic specificity in acquiring thematic knowledge of novel words from descriptive episodes. *Front. Psychol.* 8:e00488. doi: 10.3389/fpsyg.2017.00488

Zhang, M., Ding, J., Li, X., and Yang, Y. (2019). The impact of variety of episodic contexts on the integration of novel words into semantic network. *Lang. Cogn. Neurosci.* 34, 214–238. doi: 10.1080/23273798.2018.1522446

Zhang, X., Han, Z., and Bi, Y. (2013). Are abstract and concrete concepts organized differently? Evidence from the blocked translation paradigm. *Appl. Psycholinguist.* 34, 1059–1092. doi: 10.1017/S0142716412000124

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 25 May 2023
DOI 10.3389/fpsyg.2023.1144530

OPEN ACCESS

EDITED BY
Zaizhu Han,
Beijing Normal University, China

REVIEWED BY
Xiaolu Wang,
Zhejiang University City College, China
Lavern Byfield,
Southern Illinois University Carbondale,
United States

*CORRESPONDENCE
Yang Zhao
✉ zhaoyang@pku.edu.cn

RECEIVED 14 January 2023
ACCEPTED 24 April 2023
PUBLISHED 25 May 2023

CITATION
Hu W and Zhao Y (2023) Code-switching costs from Chinese-English relative clauses processing.
Front. Psychol. 14:1144530.
doi: 10.3389/fpsyg.2023.1144530

Code-switching costs from Chinese-English relative clauses processing

Wanying Hu[1] and Yang Zhao[2]*

[1]College of International Education, Minzu University of China, Beijing, China, [2]School of Chinese as a Second Language, Peking University, Beijing, China

Introduction: The source of costs is a primary concern in code-switching, yet a consensus has not yet been reached. This study investigates whether code-switching during syntactic processing in Chinese-English dual languages results in a cost.

Methods: We use Chinese and English relative clauses in either object (Experiment 1) or subject (Experiment 2, which has a more complex structure) positions to test the costs in syntactic processing. Forty-seven Chinese-English bilinguals and 17 English-Chinese bilinguals participated in acceptability judgment tests and self-paced reading experiments.

Results: The statistical findings indicate that syntactic processing is a source of the costs incurred in code-switching, as evidenced by the code-switching costs observed in the head movement during relative clause comprehension.

Discussion: The outcomes are consistent with the implications of the 4-Morpheme Model and the Matrix Language Framework. Additionally, the experiment shows that the processing of relative clauses depends on the underlying structures, which is consistent with the Dependency Locality Theory.

KEYWORDS

code-switching cost, Chinese-English bilingual, relative clause, syntactic processing, bilingualism

1. Introduction

Code-switching, shortened as CS, refers to the alternating use of two languages in a single utterance, a sentence, or other language components, which is one of the prominent features of bilingual language life (Toribio and Bullock, 2009). Within the field of psycholinguistic research on code-switching, the switching cost is a widely discussed phenomenon that refers to the cognitive consumption required by bilingual individuals when transitioning from one language to another, resulting in longer response times, higher error rates, and higher consumption of cognitive resources. This cost is quantifiable, easy to measure, and comparable, which reflects the language control ability of bilinguals (Zhang et al., 2020). Most research has found costs in code-switching understanding and producing with the discussions on influencing factors and sources of the costs.

There is currently no consensus on whether the cost of code-switching comes solely from task switching or also from language processing. There are discussions primarily on the lexical level and little research on the sentence level. This study provides evidence for the mechanism of code-switching costs at the syntactic level by examining whether code-switching will occur during the code-switching in the head movement of relative clauses comprehension in two experiments based on the different word orders of Chinese and English relative clauses. Experiment 1 examines the cost in object position relative clauses. While Experiment 2 involves the modification of materials to incorporate subject position relative clauses within an indirect speech context, thereby eliminating the potential influence of sentence-ending effects and the ambiguity associated with the definition of the matrix language.

1.1. The sources of code-switching costs

The cost of code-switching remains a topic of ongoing debate with questions surrounding its origin and whether language processing serves as the underlying reason or not. Currently there are two main explanations. The first proposes that only non-verbal processing is accountable for code-switching costs (Thomas and Allport, 2000; Zhang and Wang, 2012; Bosma and Pablos, 2020). While some studies contend that the cost of code-switching is also influenced by language processing, particularly bilingual mental lexicon processing (Grainger and Beauvillain, 1987; Casaponsa et al., 2020).

The first view is based on the idea that the code-switching process is controlled by a general cognitive control mechanism. Some studies have found that code-switching cost equals task-switching cost, which refers the increased response times and error rates that occur when individuals switch from one cognitive task to another (Monsell, 2003; Huang and Lin, 2010). Since code-switching and non-code-switching belong to two types of tasks with different difficulties using different strategies (Dijkstra and van Heuven, 2002), code-switching could be an inherent process of task-switching and entirely unrelated to language processing. Task-switching costs have been found in many code-switching studies too (Thomas and Allport, 2000; Cui and Zhang, 2010; Yim and Bialystok, 2012; Zhang and Wang, 2012; Liu et al., 2015; Timmermeister et al., 2020).

Liu et al. (2015) used alpha-number classification and picture-naming to investigate the relationship between code-switching and task-switching. Simple and mixed conditions were presented as separate blocks. In the alpha-number classification task, the alpha-number combination's position was stationary in the simple condition. The task was to determine whether a letter was a vowel or a consonant. Under the mixed condition, the alpha-number pair's position was not stationary. The task was to determine whether the letter was a vowel or consonant or to judge the parity of the number based on where the alpha-number combination appeared. The mixed conditions tested task-switching costs. The picture-naming analyzed the code-switching costs with similar sets. The simple condition required the subjects to use their mother tongue. In the mixed condition, they used either their mother tongue or second language depending on the position. In regard to the findings, there was no statistically significant interaction observed between task types and task conditions (simple/mixed). This suggests that the performance patterns of code-switching and task-switching costs are similar. According this study, it can be inferred that task-switching is responsible for all the costs associated with code-switching. However, Liu et al. (2015) concentrated on the output process of Cantonese-Putonghua bidialectal switching, which took place within the same language system and thus did not reflect the switching of two distinct languages. Switching between these two dialects may differ from switching between two distinct languages. Furthermore, their research, as well as several other experiments (Zhang and Wang, 2012), employed materials from the lexical level, which did not include switching in syntactic processing. No evidence is provided for whether there are costs from syntactic processing.

To the second view, certain studies have indicated that the cost of code-switching may arise from processing in the mental lexicon. As per this view, two distinct languages are stored in separate mental lexicons. When bilingual individuals process the orthographic or phonological information in one language, they activate only the corresponding mental lexicon. When code-switching occurs, it requires activation of another mental lexicon, leading to switching costs due to increased cognitive resource consumption. Grainger and Beauvillain (1987) has proved that there is a bilingual competition in sub-word processing by observing orthographic processing, and showed that the code-switching cost did occur at this stage.

The researchers categorized the words into two groups based on the independent variable of whether the targets were consistent with the orthographies of both languages or just one language, as demonstrated in Example (1):

(1) A. Consistent with both English-French orthographies: brain (En), Canot (boat, Fr).
B. Consistent with English or French orthographies: white (En), proie (spoils of war, Fr).

In group A, the word "brain" is an English word, and its letter combination is also acceptable in French, and "canot" too. In group B, the "wh" in "white" is a letter combination that is allowed in English but not in French, so the word does not fit in both English-French orthographies, as does "proie." In the word-nonword judgment task, there was a significant switching cost in group B but not in group A. Grainger and Beauvillain (1987) suggested that this phenomenon is related to the language-specific orthography in the mental lexicon, wherein different orthographies from different languages are stored separately and not activated concurrently. The change of orthographies leads to code-switching cost, indicating that the orthographic processing stage is the source of code-switching cost: in cases where a word possesses language-specific orthographic features, such as the English word "white" with the letter combination "wh," the decoding process is limited to activating only the orthography that is consistent with the word (other English words), while the orthography that is inconsistent with the word (i.e., French words) is not activated. When switching to the French word, bilinguals need to activate French orthography, resulting in costs. If a word's orthography is consistent with that of both languages, it activates dual language orthographies. Subsequently switching to French does not necessitate reactivation of the orthography, thereby reducing consumption.

The main focus of recent research has been on whether the costs of code-switching originate from the mental lexicon or simply from task-switching as a general cognitive process. Syntactic processing is rarely discussed. Myers-Scotton (1993, 2002) proposed that the Matrix Language Framework (shortened as MLF) and later the 4-M model, dividing code-switching into two phases: mental lexicon and formulator. The MLF puts forth the Uniform Structure Principle and the Differential Access Hypothesis. The former claims that when only one lexeme in a mixed language originates from the embedded language, the order should be consistent with that of the matrix language. The differential access Hypothesis states that no natural code-switching occurs in the formulator. This assertion implies that the source of the cost is the formulator. Li et al. (2018) use eye movement technology and discover switching costs in language processing. Dijkstra and van Heuven (2002) also argues that the code-switching cost may have multiple sources. Language processing and task switching are both potential sources of code-switching cost. The verification of syntactic processing costs, however, is still inadequate at this time.

1.2. Research on relative clauses processing

The relative clause is one of the most studied structural phenomena in language research. As the Chinese relative clause has a non-dominant word order (SVO language with a Relative-N order), the study of the relative clause holds typological significance. Research conducted on English and several other Indo-European languages has demonstrated that subject clauses are acquired prior to object clauses in both children and adults, and possess advantages in processing (Liu et al., 2011; Gutierrez-Mangado and Ezeizabarrena, 2012). Many theories, including Dependency Locality Theory, Relativized Minimality, Accessibility Theory, and Perspective Conversion Theory, have been proposed to explain this bias. Relativized Minimality and Dependency Locality Theory have received the most attention (Zhang, 2015; Hu et al., 2016). The former focuses on the distance between components in linear processing. The latter, which is based on the underlying structure, believes that movement across a greater number of components necessitates a higher level of cognitive consumption (Rizzi, 1990; Hu et al., 2016).

There is no agreed conclusion in the studies on Chinese relative clauses: some studies have found the advantage of subject clauses (Liu et al., 2011; Wei and Chen, 2020). Some studies prove that object clauses have advantages in acquisition and processing (Feng and Wang, 2013; Wang and Bing, 2013). These studies not only examine the processing of relative clauses in native Chinese speakers but also examine the development and processing of relative clauses in second and third language learners, as well as individuals with language impairments (He and Yu, 2001). Wei and Chen (2020) investigated the acquisition of relative clauses by high-level English learners in China through a Chinese-English translation test. Their study found out that the scores of Chinese English learners in translating subject relative clauses are significantly higher than that of object relative clauses. The main effect of clause types is significant, indicating that the two kinds of relative clauses are different in acquisition and actual use. Their research also examined the influence of position (clauses occupy subject or object position) on the acquisition and found out that subject-position relative clauses are easier to be learned.

As the Chinese relative clause has a specific word order, it is an excellent syntactic structure to test the word order hypothesis in the Matrix Language Framework (Myers-Scotton, 1993), which proposes that the word order of the embedded language should be consistent with that of the matrix language. At the same time, with different complexity of subject and object clauses, our research will use it to test the influence of grammatical difficulties on code-switching, which is also a test for Dependency Locality Theory and learners' acquisition of relative clauses.

1.3. Code-switching on relative clauses

According to the principle of morpheme order in MLF, the word order of the embedded language should be consistent with that of the matrix language. Therefore, when code-switching is in relative clauses, the word order of the embedded language should be the same as the matrix language. Chinese and English relative clauses build up the word orders with opposite movement directions. By investigating code-switching occurring in the movement, this study aims to observe whether a cost is incurred during syntactic processing or not. As Chinese relative clause (clause-antecedent) is different from the English word order (antecedent-clause), this study chooses the opposite word orders of the relative clause to investigate Chinese-English code-switching costs in syntactic processing. The word order of Chinese relative clauses is unique in typology. Among the 879 languages in Dryer (2013), there are only five languages with VO and Rel-N, including Chinese, as shown in example (2a). Meanwhile VO and N-Rel co-occurs in 416 languages, including English in example (2b).

(2) a. *wo renshi dai maozi de shushu.*
I know wear a hat aux uncle
b. I know this uncle who wears a hat.

Examples (2a) and (2b) mean the same. The word order of the Chinese relative clause in the sentence a is "verb + noun + *de* + antecedent," where *de* is the marker of the Chinese relative clause, which is in front of the antecedent. In the sentence b, the word order is "antecedent +who+ verb + noun," where "who" is the marker of the English relative clause, which is placed after the antecedent.

According to the principle of morpheme order, when Chinese and English code-switching occurs in the relative clause with only one embedded word, the sentence order should be consistent with the matrix language. Therefore, if the matrix language is English with an embedded Chinese relative clause, the word order of the Chinese clause should be the same as in English, with the clause coming after the antecedent. As is shown in example (3).

(3) a. I knew this uncle who wore *maozi* (hat).
b. *I knew this wore *maozi* (hat) *de* (aux) uncle.

Example (3a) and (3b) are two sentences that includes English-Chinese code-switching, where English is the matrix language, and Chinese is the embedded language. The correct word order should be the one in English (3a). Whereas sentence (3b) has a Chinese noun in the relative clause with the Chinese word order, and is not in conformity with the matrix language which may produce cost in processing. The difference between the two sentences (3a and b) is only in the direction of the antecedent movement. The direction of sentence (3a) is the same as that of the English relative clause, which has moved from the base generation position (in front of wear) to the left before WHO, while the direction of the sentence (3b) is the same as that of the Chinese relative clause (Myers-Scotton and Jake, 2009). At the same time, the movement distances in both CS sentences (3a and 3b) are the

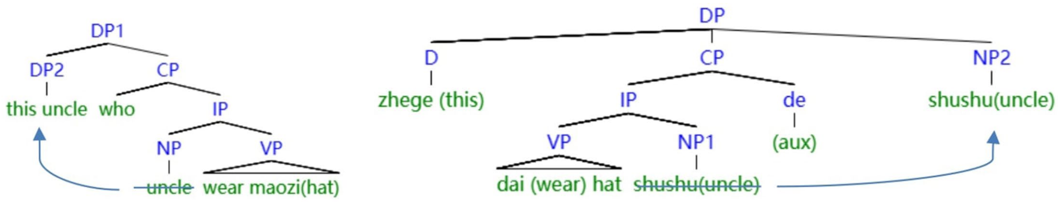


FIGURE 1
Tree structures of CS relative clauses.

same, and there is no need to cross other components (see Figure 1). Therefore, when the direction of movement affects the code-switching costs, it suggests that the cost may come from the sentence processing stage.

1.4. The present study

Previous research has examined the lexical-level costs of code-switching. Some theories and empirical research also provide evidence that syntactic processing may be a source of costs associated with code-switching. However, very few, if any, studies have directly looked at the costs at the syntactic level. The current study thus aims to examine whether costs are from switching in syntactic processing.

The current study conducts two experiments with self-paced reading and acceptability judgment tasks to see if the correspondence of the matrix language to movement directions affects code-switching costs. The results indicate that costs may or may not be elicited during syntactic processing.

Several factors, such as domestic languages, matrix languages, and tasks, have an impact on the costs (Chang et al., 2017). Therefore, Chinese-English and English-Chinese bilinguals, as well as code-switching from two directions (English to Chinese or the other way round), are all included in this study. Additionally, it employs offline and online tasks. We predict that costs will be present in a number of situations, demonstrating syntactic processing as a source of code-switching costs.

2. Experiment 1: code-switching on the relative clauses in the object position

2.1. Design

Experiment 1 is a 2×2 two-factor repeated measurement design. The first independent variable is word order, a within-subject design with two levels: Chinese and English order. The second independent variable is the matrix languages, a within-subject design with two levels: Chinese and English. The dependent variable is the acceptability in Acceptability judgment task (shortened as AJT) and RT in self-paced reading.

2.2. Participants

Twenty-six Chinese-English (Chinese native English learners, shortened as Ch-En) bilinguals participated in self-paced reading and acceptability judgment tasks. They are mainly undergraduates in Beijing with an age range of 18–30. The English proficiencies of participants are upper intermediate (self-rated score 6.26/10). The Wilcoxon signed rank test showed a significant difference between the Chinese-English self-assessment, $p < 0.001$.

After the experiment, each participant received 30 *yuan*, and an additional 15 *yuan* as a reward when the accuracy was over 95%. The study was approved by the School of Chinese as a Second Language, Peking University.

2.3. Materials

Thirty-two Chinese sentences containing object-position relative clauses, as well as their English translations, are the original materials in the study, seen in Supplementary material. Some expressions are modified in the non-interest region of English sentences. The acceptability of sentences and the word frequencies of corresponding Chinese and English components in relative clauses are controlled.

This study only switches one noun in each relative clause and does not constitute any island to satisfy the Morpheme Order Principle. Additionally, to make subjects activate both languages before the relative clause, the determiner in the antecedent DP is switched. And to prevent the participants from guessing the purpose, parts of the following sentence are also switched. Finally, the word orders of both Chinese as matrix language, English as embedded one (Ch-En) and En-Ch relative clauses are switched to the embedded-language word orders or kept as the original ones to make up all four conditions in the experiment. Each group of sentences has one comprehension question, and the language of comprehension questions is consistent with the matrix language of the experimental sentences. Table 1 is the examples, and all Chinese words are Chinese characters (but Pinyin here for easy access).

Experimental materials form four groups, with 32 sentences in each group and eight in each type. Each participant could only see one sentence in a group. At the same time, participants also read 64 filler sentences with comprehension questions. 96 sentences are presented pseudo-randomly to ensure that three experimental sentences do not appear in a row.

2.4. Procedure

Before the experiment, the participants completed the Chinese and English proficiency tests and get familiar with the word list of the experimental sentences.

OpenSesame V3.3.5 is used to present the program and record the data for the self-paced reading. The instruction is presented first, and

TABLE 1 Examples of materials in Ex1.

Types	Sentences
En-Ch En order	I know 这个 **gentleman who always wears** 眼镜. He is a 朋友of my father's.
	➤ 这个-zhege-this; 眼镜-yanjing-glasses; 朋友-pengyou-friend.
En-Ch Ch order	*I know 这个 **always wears** 眼镜/的 **gentleman**. He is a 朋友 of my father's.
	➤ 的-de-AUX.
Ch-En Ch order	我/认识/**this/** 总是/戴 **/glasses/** 的/叔叔, /他/是/我/爸爸的/friend.
	➤ 我-wo-I; 认识-renshi-know; 总是-zongshi-always; 戴-dai-wear; 叔叔-shushu-gentleman; 他-ta-he; 是-shi-is; 爸爸-baba-father.
Ch-En En order	*我/认识/**this/** 叔叔 **/who/** 总是/戴/ **glasses**, /他/是/我/爸爸的/friend.
Comprehension Q	My father has a friend.—T

The underlined and bold parts are the areas of interest. The italic parts are Chinese words, and Pinyin (pronunciation) and translation are under the sentences. The slash lines show the Chinese word segmentation. The sentences with * are theoretically incorrect sentences.

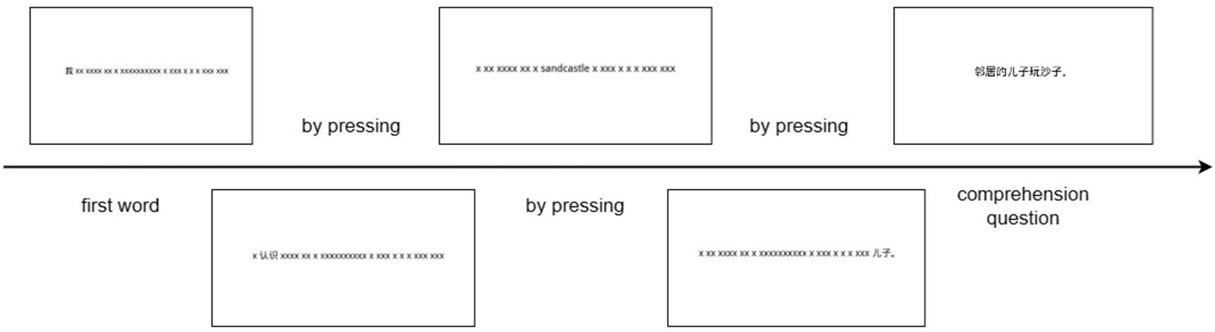


FIGURE 2
The illustration of the experiments.

then the practice session. The experimental sentence is presented in the non-cumulative form of a moving window. The first word is presented while other words are presented as "x" (each English word is replaced by an x. Each Chinese character uses two x instead), and then the word is presented by pressing the space bar. At the same time, the previous one disappears. After the whole sentence is presented, the reading comprehension question will appear on the next screen after a press. The participants press "Z" or "/" to represent "yes" or "no" to answer the questions. The illustration is in Figure 2.

After the online experiment, the participants complete the acceptability judgment of the experimental sentences and fillers (seven-point scale). Each participant judges 96 new sentences.

2.5. Data analysis

The correct rate of reading comprehension tasks is 95.99% ($SD = 2.10\%$). For the reaction time, we first delete incorrect reactions and remove outliers.

SPSS 20.0 and R 3.6.3 are used for data statistics: SPSS 20.0 is used for descriptive statistical analysis, paired *t*-test, and repeated measurement test. The POLR function in the MASS package in R is used to conduct logistics analysis on the acceptability judgment data. And the LME4 package (Bates et al., 2015) does linear mixed-effect model analysis on the experimental data. In the mixed effects model analysis by R, logarithms reaction time is used.

2.6. Results

2.6.1. Acceptability judgment task

The descriptive statistical results are shown in Table 2. Repeated measurement variance analysis by the subject shows that the main effect of the matrix language is not significant, and the main effect of consistency is near significant [$F_1(1, 25) = 4.123$, $p = 0.053$, $\eta^2 = 0.142$, $\beta = 0.497$]. The interaction is significant, $F_1(1, 25) = 4.306$, $p = 0.048$, $\eta^2 = 0.147$, $\beta = 0.514$. The results of simple effect analysis show that when Chinese is the matrix language, there are significant differences between the consistent and inconsistent groups [$p = 0.005$, *CI* [0.078, 0.384], $\eta^2 = 0.279$, $\beta = 0.848$]. When English is the matrix language, there is no significant difference in main effects and interaction effects. Repeated measurement variance analysis by item also finds a near significant main effect of consistency. The results show that consistency influences native speakers' acceptability of code-switching sentences.

2.6.2. Self-paced reading

Descriptive statistics results from each region of interest are shown in Table 3. The results of linear mixed model analysis are in Supplementary material.

In the antecedent D (*zhege*/this) region, the linear mixed model analysis shows that the main effect of matrix language is significant, the main effect of consistency and the interaction are not significant.

In order to examine the individual effects of the matrix and native languages, a simple effect analysis is performed, even in cases where the interaction is not found to be significant. The results of the simple effect analysis show no significant difference between the consistent and inconsistent groups when Chinese and English are the matrix language, respectively.

In the verb (*dai*/wears) region, the linear mixed model analysis shows that the main effect of subject language is significant, the main effect of consistency is not significant, and the interaction between the two is not significant. The results of the simple effect analysis show no significant difference between the consistent and inconsistent groups when Chinese and English are the matrix language, respectively.

In the switched "inner N" (glasses/*yanjing*) region, the linear mixed model analysis shows that the main effect of matrix language is significant, the main effect of consistency is not significant, and the interaction between the two is significant. The results of simple effect analysis show that when Chinese is the matrix language, there is no significant difference between the consistent and inconsistent groups. When English is the matrix language, there is a significant difference between the consistent and inconsistent groups [$t(662) = 2.837$, $p = 0.005$].

In the "outer N" (*shushu*/gentleman) region, the results of linear mixed model analysis show that the main effect of matrix language is significant, the main effect of consistency is near significant, and the interaction between the two is significant. The results of simple effect analysis show that when Chinese is the matrix language, there is a nearly significant difference between the consistent and inconsistent groups [$t(636) = 1.710$, $p = 0.088$]. When English is the matrix language, there is no significant difference between the consistent and inconsistent groups.

2.7. Discussion

The findings demonstrate that the code-switching cost is significantly influenced by whether the relative clause word order is consistent with the matrix language. AJT results show that when the relative clause word order is consistent with the matrix language, bilinguals are more likely to accept the switching sentences. In other case, when the matrix language is Chinese, Ch-En bilinguals are more sensitive to it. Self-paced reading provides more details on time course: the main effect of consistency and the interaction between consistency and matrix languages appear in the outer N period. The interaction also appears in the switched inner N area. Last but not least, Ch-En bilinguals performed differently when the matrix languages are different.

The cost of inconsistency, found in both the acceptable judgment task and the self-paced reading task, indicates that the word order of the relative clause affects CS cost. From this, we can infer that the cost comes from the word-order processing of relative clauses at the sentence-processing stage. Chinese and English declarative sentences have the same SVO word order, but the orders of the relative clause are not consistent. The movement directions of the Chinese and English relative clauses are not the same. During the movement, Chinese goes to the right while English to the left, which is the main difference between Chinese and English relative clauses (Xiong, 2005). In Experiment 1, code-switching costs occur in the Noun-movement where the opposite moving need to be processed. The results also show that the Noun-movement is one of the code-switching cost sources.

In Experiment 1, some results do not follow the expectations: in the "inner N" region, when English is the matrix language, and in the "outer N" and "V" region, when Chinese is the matrix language, the inconsistent condition processes faster. These phenomena may be due to the unclear definition of the matrix languages or the sentence ending effect.

First, there are many possibilities for the matrix languages of relative clauses. The definition of the matrix language of a clause is controversial, which may be the language of the main clause or the language used by the core component of the clause. At present, the research on code-switching has not provided a certain confirmation method to define the matrix language of subordinate clauses (Myers-Scotton, 1993). In this study, the content words in the core are consistent with the matrix language, and thus we define the language of the main clause as the matrix language of the relative clause. For example, in "*Wo renshi* (I know) this *zongshi* (always) *dai* (wear) glasses *de* (aux) *shushu* (gentleman), *ta shi wobaba de* (he is my father's) friend"(Chinese in italics, and translation in brackets), the language of the main clause "*wo renshi* (I know)" and the content word "*shushu* (gentleman)" are Chinese, so the matrix language is Chinese, and the word order of the clause should follow the Chinese.

However, there are other possibilities for the matrix language of relative clauses. One is the language of the antecedent's determiner. The core component of the relative clause is DP, that is, "this" in the sentences. To enable the participants to activate both languages before entering the relative clause, "this (*zhege*)" is switched in this study, which may lead to the change of the matrix language, resulting in costs contrary to the prediction. This possibility could explain some of the non-conformity, but not all. Experiment 2 clarifies the matrix language by changing the switching position in the main sentence.

TABLE 2 descriptive statistics results of Ex 1.

ML	Agree	Disagree
Chinese	3.05(1.46)	2.82(1.41)
English	2.80(1.43)	2.75(1.42)

TABLE 3 Descriptive results of reaction time in Ex 1 (ms, and SD in brackets).

ROI	Ch-En agree	Ch-En disagree	En-Ch agree	En-Ch disagree
D	388.87(161.20)	376.79(148.80)	400.94(122.41)	403.65(118.85)
V	414.42(153.22)	389.11(127.43)	521.77(198.14)	513.96(227.01)
Inner-n	573.71(325.71)	574.84(362.58)	610.40(319.16)	541.67(240.00)
Outer-n	453.58(191.89)	425.72(155.72)	494.52(252.85)	513.98(243.82)

Another possibility for the matrix language is the mother tongue of the bilinguals. Previous studies have proven that proficiencies affect cognitive control abilities and code-switching costs (Meuter and Allport, 1999; Costa and Santesteban, 2004; Zhang and Cui, 2008; Linck et al., 2012). Therefore, Chinese-English bilinguals have higher rate of activation of their mother tongue with the dominant word order from Chinese. The inconsistency condition (Chinese order) processes faster when English is the matrix language. This possibility is also limited as it cannot explain the processing advantage of the inconsistency condition when Chinese is the dominant language. To solve this problem, Experiment 2 includes both English and Chinese bilinguals.

The phenomenon may also relate to the sentence-ending effect. In this experiment, the relative clause is at the end of the first clause. Although the participants know the sentence does not end after the relative clauses, they are still likely to have a long pause when reading punctuation marks, showing the sentence-ending effect (Ni et al., 1998; Chang et al., 2009). Moreover, by the end of the sentence, bilinguals, acquiring enough information, may start to integrate information causing the reactions to delay. The sentence ending effect can explain parts of inconsistent advantages. However, it cannot explain the inconsistent advantage in the V region or the performance of the "outer N" when Chinese is the matrix language and "inner N" when English is the matrix. In addition, previous studies have found that punctuation processing tends to trigger skip reading and shortened fixation time (Wang et al., 2018). This facilitating effect is inconsistent with the sentence ending effect found in this study. In view of this possibility, the positions of relative clauses change in Experiment 2, to eliminate the possible influence of punctuation or sentence-ending effect.

3. Experiment 2: code-switching on the relative clauses in the subject position

To further examine the costs from syntactic processing, we design experiment 2 for Chinese-English and English-Chinese (shortened as En-Ch) bilinguals to eliminate the effects of matrix languages and sentence structures with updated materials.

3.1. Design

This experiment is a $2 \times 2 \times 2$ three-factor repeated measurement design. The first independent variable is word order, a within-subject design with two levels: Chinese and English. The second independent variable is the matrix language, a within-subject design with two levels: Chinese and English too. The third variable is the language background, a between-subject design with two levels: Chinese-English and English-Chinese bilinguals. The dependent variable is acceptability in AJT and RT in self-paced reading.

3.2. Participants

Twenty-one Chinese-English and 16 English-Chinese bilinguals (shortened as En-Ch) participated in self-paced reading and acceptability judgment tasks. Ch-En bilinguals are students in Beijing and they participated in the experiment in the lab. En-Ch bilinguals were recruited online with an age range of 18–35. They conducted the experiments remotely under experimenter's supervision by Tencent Meeting.

These bilinguals are upper-intermediate learners of Chinese or English. The differences within the in-group regarding bilingual proficiencies are significant. After the experiment, each participant receives 30 *yuan*, and an additional 15 *yuan* as a reward when the accuracy is over 95%.

3.3. Materials

We revised the sentences in Experiment 1. First, the main clauses are extended to "I heard from XX (name) that." Second, the relative clauses move to the subject position of the indirect speech, e.g., "I heard from my father that the uncle always wears glasses …." Finally, the predicate is added directly without punctuations, such as "plans to come to my house and have dinner next week ". The whole sentence is "I heard from my father that *the gentleman who always wears glasses* plans to come to my house and have dinner next week" (The italics are relative clauses and their antecedents, including the four areas of interest in this experiment). All sentences without switching are in Supplementary material.

Thirty-two Chinese sentences containing object-position relative clauses are the original materials in the study, as well as their English translations. Then we make switching materials with the same steps in Experiment 1. Table 4 is the examples for the materials in Experiment 2.

3.4. Procedure and data analysis

The procedure and data analysis are the same as in Experiment 1. The average correct rate of reading comprehension tasks is 95.54% ($SD = 2.10\%$).

3.5. Results

3.5.1. Acceptability judgment task

The descriptive statistical results are in Table 5. For Ch-En bilinguals, the logistic regression model for ordered variables analysis shows that the main effect of the matrix language is not significant; The main effect of consistency is significant ($\beta = 1.503$, $t = 7.562$, $p < 0.001$, OR value 4.494). The interaction is significant ($\beta = -1.021$, $t = 7.562$, $p < 0.001$, OR value 4.494). The results of simple effect analysis show that when Chinese is the matrix language, there are significant differences between the consistent and inconsistent groups ($Z = 7.562$, $p < 0.001$). When English is the matrix language, there are also significant differences between the consistent and inconsistent groups ($Z = 2.441$, $p = 0.015$). Ch-En bilinguals are sensitive to the relative clauses' word orders.

For En-Ch bilinguals, the logistic regression model for ordered variables analysis finds no significant effect of the matrix language, consistency, or interaction. En-Ch bilinguals exhibit a lack of responsiveness to the relative clause word order's conformity to the matrix language.

TABLE 4 Examples of materials in Ex 2.

Types	Sentences
En-Ch En order	I heard from 爸爸 that **the gentleman who always wears** 眼镜 plans to come to my house 吃饭 next week.
	➤ 爸爸-baba-father; 眼镜-yanjing-glasses; 吃饭-chifan-have dinner.
En-Ch Ch order	*I heard from 爸爸 that **this always wears** 眼镜/的 **gentleman** plans to come to my house 吃饭 next week.
	➤ 的-de-AUX.
Ch-En Ch order	我/听/my father/说/那个/总是/戴/ **glasses** /的/叔叔/下周/要/来我家, /和/我们/一起/have dinner。
	➤ 我-wo-I; 听-ting-hear from; 说-shuo-say; 那个-nage-that; 总是-zongshi-always; 戴-dai-wear; 叔叔-shushu-gentleman; 下周-xiazhou-next week; 要-yao-will; 来我家-lai wojia-come to my home; 和-he-with; 我们-women-we; 一起-yiqi-together.
Ch-En En order	*我/听/my father/说/那个/叔叔 **/who/** 总是/戴 **/glasses/**下周/要/来我家 ，/和/我们/一起/have dinner。
Comprehension Q	A gentleman is coming to my house next week. —T

The underlined and bold part is the area of interest. The italic parts are Chinese words, and Pinyin (pronunciation) and translation are under the sentences. The slash lines show the Chinese word segmentation. The sentences with * are theoretically incorrect sentences.

TABLE 5 descriptive statistics results of Ex 2 (ms, *SD* in blankets).

	ML	Agree	Disagree
Ch-En bilingual	Chinese	3.80(1.24)	2.64(1.28)
	English	3.13(1.47)	2.76(1.32)
En-Ch	Chinese	3.87(1.36)	3.75(1.45)
Bilingual	English	3.85(1.36)	3.86(1.40)

3.5.2. Self-paced reading

Descriptive statistics results of each region of interest are shown in Table 6. The results of linear mixed model analysis are in Supplementary material.

In the antecedent D (*zhege*/this) region, the linear mixed model analysis shows that the interaction between matrix language and consistency is near significant. But all main effects and other interaction effects are not significant. The results indicate the validity of the experiment and analysis.

In the verb V (*dai*/wears) region, the linear mixed model analysis shows that the main effects of matrix language and the interaction between matrix languages and background are significant. The matrix language and background influence bilinguals' code-switching processing.

In the noun "inner N" (glasses/*yanjing*) region, the linear mixed model analysis shows no significant effect. For Ch-En and En-Ch bilinguals respectively, but there are significant differences when English is the matrix language.

In the noun "outer N" (*shushu*/gentleman) region, the results of linear mixed model analysis show that the main effect of consistency is nearly significant, and the interaction between the matrix languages and consistency is significant. For Ch-En bilinguals, the main effect of matrix language, consistency, and their interaction are significant. When Chinese is the matrix language, the effect of consistency is significant. But for En-Ch bilinguals, only the effect of matrix language is found.

For the whole relative clauses region, significance is found in the main effect of the matrix languages and the interaction between the matrix languages and background. For En-Ch bilinguals, the interaction between consistency and matrix language is significant, and the consistency effect is significant when English is the matrix language.

Bootstrap paired sample *t*-tests are conducted with a sample size of 10,000 to analyze the data of Chinese-English bilinguals under the condition of consistency and inconsistency. The results show that in the "V" and "outer N" regions, the consistency effect in Chinese is nearly significant by subject analysis. In the "inner N" region, the consistency effect is significant when English is the matrix language.

3.6. Discussion

Experiment 2 improves the materials and resolves remaining problems in Experiment 1. Overall, the findings are in line with the theoretical hypothesis.

Some conjectures from Ex1 are verified. First, Experiment 2 provides a clear definition of matrix language. The main clause's objects are the first portions of the sentences to switch, which are independent from the relative clause and have no bearing on the relative clause's choice of matrix language. Experiment 2 increases the number of components additionally in the section prior to the RC from 3 to 5, which makes the matrix language exactly the one used in the main clause. The findings demonstrate that when English is the matrix language, condition with no inconsistency has an advantage. Additionally, the matrix language, not the linguistic background, has a large impact on the outer N and V areas. With the results of Experiment 1, we conclude that the matrix language of a relative clause may be determined by the language used by the core component and antecedent DP (the/this/that), not the mother tongues of bilinguals.

Second, Experiment 2 excludes the sentence ending effect caused by punctuation. In Experiment 2, the relative clause serves as the indirect speech's subject and is followed by a predicate component without punctuation. The results of Experiment 2 did not find the ending effect of pauses in the relative clause, indicating that punctuation may bring the end effect. This result goes against the earlier findings (Wang et al., 2018). However, due to the lack of other relevant evidence on punctuation processing, the role of punctuation in language processing needs to be further investigated.

The modified experiment illuminates the role of consistency. The consistency effect in AJT is significant. In self-paced reading, the

TABLE 6 Descriptive results of reaction time in Ex 2.

	ROI	Ch-En agree	Ch-En disagree	En-Ch agree	En-Ch disagree
Ch-En bilingual	D	396.66(135.29)	391.63(156.23)	398.72(156.26)	419.88(158.72)
	V	394.19(157.82)	414.60(164.54)	475.45(192.52)	502.61(214.96)
	Inner-n	608.56(358.86)	609.75(360.16)	577.96(290.40)	639.75(327.99)
	Outer-n	443.86(210.01)	406.36(204.10)	530.09(250.31)	561.84(274.43)
En-Ch bilingual	D	371.65(140.78)	374.72(151.89)	343.05(138.05)	335.81(124.44)
	V	359.04(149.16)	349.29(136.70)	353.90(134.39)	343.45(127.56)
	Inner-n	435.63(211.38)	441.03(223.73)	414.22(226.04)	484.22(250.69)
	Outer-n	391.47(183.40)	399.93 (188.04)	369.58(154.51)	385.19(156.42)

Chinese-English and English-Chinese bilinguals exhibit significant consistency effects in some areas. It demonstrates how the matrix language and word order have an impact on code-switching costs. In this study, the word order of relative clauses—basically the direction of movement—is manipulated. Nouns should be relativized to the right when Chinese is the matrix language and to the left when English is the matrix language. Movement is an operation in the syntax formulator. We can conclude that code-switching in the syntax formulator results in costs as we discover the cost of movement processing.

Experiment 2 finds the code-switching costs have different patterns when matrix languages, language proficiencies and tasks are diversified. When English is the matrix language, Ch-En and En-Ch process more quickly in the inner N (glasses) region. But when Chinese is the matrix language, there are no effects. When processing relative clauses including code-switching, bilinguals need to process code-switching and relationalization together, which requires high consumption. The inner N is the last word in relative clauses when English is the matrix language. Bilinguals can quickly process relative clauses after processing this word. They are intelligent enough to notice the contradiction as a result. However, when Chinese is the matrix language, inner N is in the middle. Bilinguals need to get more input to conduct movement and processing RC. Therefore, they are insensitive to the inconsistency. The proficiency of matrix languages is also important. Bilinguals have different control capacities over two languages during complex processing (Schilling et al., 2014). The fact that bilinguals who are En-Ch have consistency effect when English is the matrix language is the clue.

The offline AJT shows significant effects, especially when Chinese is the matrix language. In the online self-paced reading task, the consistency effect sizes are small. The results are different from Ex1. These phenomena may be due to task and structure difficulties. In experiment 2, we moved the relative clause from the object position to the subject. Previous studies have shown that for children and second language speakers, the subject relative clause in the subject position is the easiest (Ma, 2012; Feng and Wang, 2013). Structural difficulty reduction helps bilinguals to process inconsistencies in online tasks, while they could find inconsistencies easily in offline tasks. The structural difficulty affects the generation of code-switching costs (Tarłowski et al., 2013; Zeller, 2020). If bilinguals have enough cognition resources to analyze the syntactic structure in AJT, they can fully invoke explicit knowledge to make judgments and are more sensitive to consistency differences. Especially when the matrix language is the mother tongue of bilinguals, they make decisions quicker and more accurate.

There is a finding that our prediction cannot explain: When Chinese is the matrix language, Ch-En's RT to the consistent condition is slower in "outer N (uncle)." This phenomenon may be related to the unique relative clauses to Chinese. In most languages, with SVO as the basic word order and noun-relative clause order, the core components are in the front of the RC. However, Chinese is an SVO language with a postpositional antecedent, with core components on different positions between the main sentence and relative clauses. Chinese native speakers can accept two word-orders. When Chinese and English are activated, Ch-En bilinguals activate Chinese and English basic word order with a prepositive core component, English and Chinese RC order. In real-time language processing system, the order of the prepositive core component may have an advantage. It may be why they can better understand the relative clause with the prepositive core component. However, this supposition lacks evidence. The reason why Ch-En bilinguals have fast processing in the condition of inconsistent "outer N (uncle)" position needs to be investigated.

4. General discussion

Experiments 1 and 2 investigate the Chinese-English code-switching costs in syntax processing by manipulating the relative clause word orders. The results show that costs are from the processing of movement, which indicates the syntax process is one of the sources of code-switching costs. Also, the experiments also show evidence of RC's processing mechanism.

4.1. The source of code-switching costs in syntax processing

This study uses two experiments to examine the code-switching costs in syntax processing. The results show that when relative clause word order is inconsistent with the matrix language, code-switching produces more costs with slower processing. As the order processing is at the stage of the syntactic operation, this stage is the source of code-switching costs. This result conforms to the prediction of the Morpheme-Order Principle in the Matrix Language Framework, Uniform Structure Principle, and Differential Access Hypothesis (Myers-Scotton and Jake, 2009).

According to the Uniform Structure Principle, code-switching should keep in consistency with the matrix language generally. The

Different Access Hypothesis holds that different morphemes have different switching mechanisms: The bridging late system morphemes, external late system morphemes, and other syntax structures are prominent in the syntactic generator. Morphemes salient in different portions have different difficulty degrees in code-switching: the morphemes generated in the mental lexicon can switch automatically, while the morphemes and other operators in the syntactic generation cannot switch, and understanding these switching structures is difficult with extra cognition resources. Therefore, code-switching in the process of movement is unnatural and labeled, requiring additional cognitive resources and resulting in consumption. That is to say, the code-switching cost will occur on the processing of switched syntax structures. The results of these experiments confirm the hypothesis and show that the syntax process is one of the sources of code-switching costs, which is in line with the predictions of the Matrix Language Framework Model (Myers-Scotton, 1993).

The component realized in syntactic generators is less code-switched. According to definition of external late system morphemes of Myers-Scotton (2002), there are no such morphemes in Chinese or English. Given that the one function of late system morphemes is to construct sentences, this study aims to examine the associated costs by selecting the word order of the relative clause as a means of testing the effects of switching which involves movement in sentence completion. The Morpheme Order Principle also suggests that the word order of the embedded language should be consistent with that of the matrix language. The result from this study, the switching on relative clauses producing cost, shows that the sentence completing stage is the source of cost. This result conforms to the prediction of the Morpheme Order Principle in Matrix Language Framework as well as the Unified Structure Principle and Different Access Hypotheses (Myers-Scotton and Jake, 2009) and reflects the process of sentence-level code-switching.

However, the results in these two experiments are not consistent, indicating some factors influencing the production of costs. Matrix languages and the difficulty of the task could be two major determining factors. First, CS costs are influenced by the matrix language's proficiency. When the matrix languages change, bilinguals between Chinese and English show varying effect sizes in various regions. Bilinguals have different control abilities when processing languages because their language proficiency is different (Schilling et al., 2014). Participants in this study process language, code-switching, and movement together. Native speakers of matrix languages are better at processing language and movement. That is why Chinese native speakers exhibit a sensitive perception of consistency differences when Chinese is the matrix language. Also, the costs are impacted by the complexity of the task. When the structure is simple (Ex1), bilinguals have enough cognitive resources to process the switching quickly online. However, bilinguals hardly understand sentences automatically when dealing with a more complicated relative clause nested inside the indirect speech clause (Ex2). Therefore, they may ignore the inconsistency during online processing. When they have enough time to think about linguistic structure and the built-in grammar of code-switching, they notice inconsistency and make better decisions offline.

4.2. The mechanism of relative clauses processing

Researchers have proposed different theoretical models to explain the generation of relative clauses, among which the Relativized Minimality Theory and the Dependency Locality Theory are the two most widely discussed (Zhang, 2015; Hu et al., 2016). Many empirical studies on relative clauses provide evidence for these two theories. This study provides evidence from Chinese-English bilinguals and supports Dependency Locality Theory.

The word orders of relative clauses are manipulated in two experiments. In the sentences, the words "always" "wear" "*yanjing* (glasses)" and "gentleman" are used to make relative clauses "gentleman who always wears *yanjing* (glasses)" and "always wears *yanjing* (glasses) *de* (aux) gentleman" respectively in the Chinese or English relative clause word order. During statistics, we draw the processing diagram according to the underlying structure order and the time course, respectively (see Figure 3).

Based on the underlying structure (the figures above), the response patterns of relative clauses are relatively consistent. Although the sentences from the two conditions are not the same in the linear with different movement directions, the corresponding components have similar processing trends. However, based on the surface structures order (the figures below), the reaction patterns of relative clauses are generally different. It can be inferred that the time course may not be the most important point for the processing. The overall trend is that the processing of switched inner N (glasses) becomes slower, reflecting the code-switching cost. At the same time, the patterns may also be related to the matrix language. We also control the frequency of words corresponding to the same position, which eliminated the influence of vocabulary and language differences to a certain extent.

According to the patterns in the Experiments, we find out that the process of the relative clause may be more related to the underlying structure rather than to the surface time course. Though the experiments use self-paced reading in which the participants read word by word according to the time course, the results show a more consistent underlying reading pattern, indicating that when understanding relative clauses, bilinguals tend to use the underlying structure. Reading comprehension starts from the underlying grammatical structure (Peng, 1991). It also indicates that noun movements are in the processing of relative clauses. This study supports that the underlying structure plays a role in relative clause processing, consistent with the predictions of the Dependency Locality Theory (Rizzi, 1990). These experiments also prove the rationales for the direct comparisons of corresponding components (Inner-n, Outer-n, V) during data analysis.

5. Conclusion

This study uses relative clause order as a maker to look at the code-switching cost in syntactic processing. The findings imply that matrix language should be consistent with the relative clause order. If not, the understanding of code-switching produces cost. The syntactic operation, which includes word order processing and movement, is one of the sources of code-switching costs. Our finding is consistent with the predictions of the Morpheme-Order Principle in the Matrix Language Framework, the Uniform Structure Principle, and the Differential Access Hypothesis (Myers-Scotton and Jake, 2009).

Experiment 2 is a modification of Experiment 1. Comparing the two, we believe that the matrix language is mainly decided by the core component of the clause, rather than the learner's mother tongue. In addition, Experiment 2 also eliminated the sentence end effect. Previous studies on sentence end effects often focus on the end of the sentences, rather than the punctuation in the middle. The results of

10.3389/fpsyg.2023.1144530

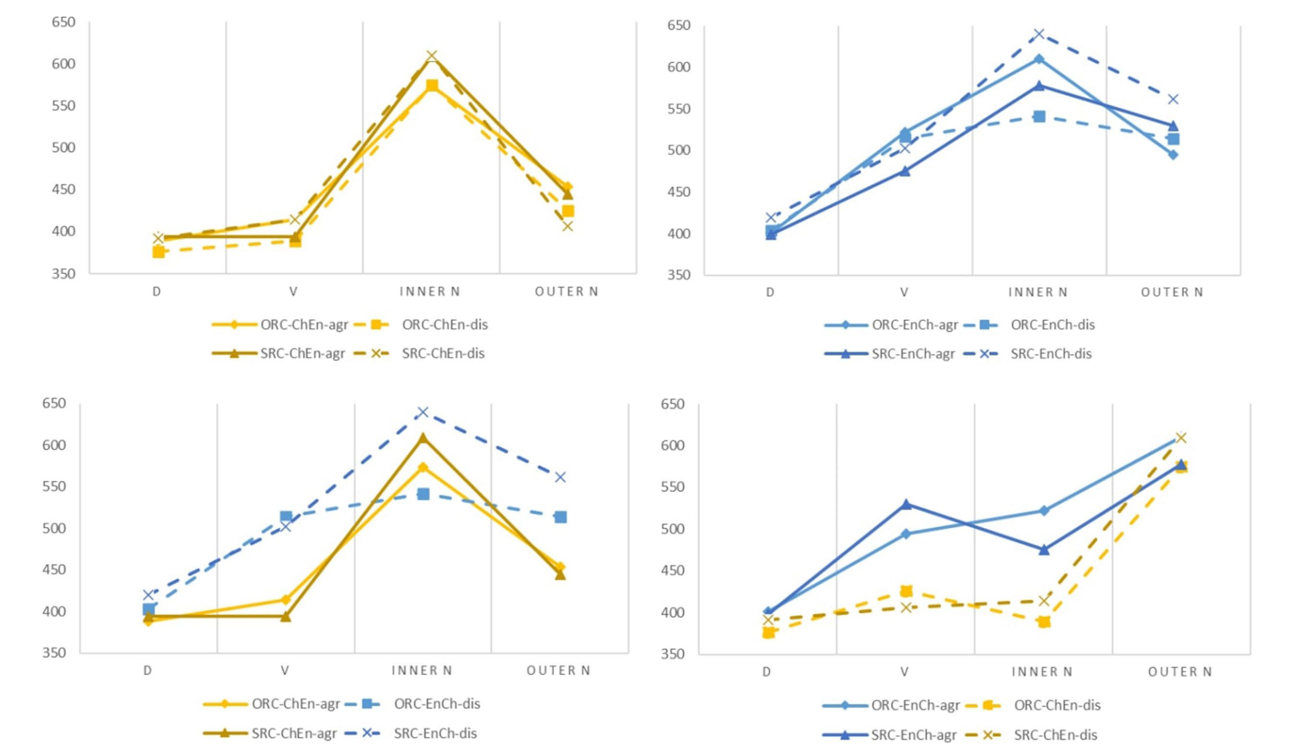


FIGURE 3
Processing diagram of relative clause (The two graphs above are in the order of the underlying structure. The two graphs below are in the order of time course).

this experiment also suggest that researchers need to pay more attention to the processing of punctuation.

The processing patterns in the Experiments show that Chinese and English bilinguals rely on the underlying structure when understanding relative clauses. Even when reading relative clauses word by word, bilinguals tended to process the relative clauses from the bottom-up pattern. The finding of this pattern provides bilingual evidence for the Dependency Locality Theory. Furthermore, the experiments also indicate that the subjects' language proficiency, as well as the complexity of the structures and tasks involved, may affect the understanding of code-switching.

This study investigates code-switching on the syntactic level, which gives new evidence on the source of code-switching costs. At the same time, this study provides new clues for several theoretical questions, such as the definition of matrix language in clauses, and the mechanism of relative clause processing from a bilingual perspective. Finally, this study finds several linguistic factors may have influence on code-switching costs. Some of them are with less concern before.

This work presented several limitations. First, there are not enough participants, particularly En-Cn bilinguals. The majority of them carried out the experiments online, where there may have been more disturbance than in the lab. Second, because only Chinese and English were used to test the source in this study, the findings need to be verified in more language conditions. Therefore, further investigation is needed to examine the conclusion with more languages and grammatical structures.

In conclusion, this study investigates code-switching at the syntactic level and provides fresh evidence regarding the origin of code-switching costs. In addition, this study offers new insights into some theoretical issues, including how relative clauses are processed from a bilingual perspective and the definition of matrix language in clauses. Finally, this study discovers that some linguistic variables may affect code-switching costs. Some of them showed less concern before.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by YZ, Peking University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YZ is WH's Ph.D. supervisor. This research is a part of WH's doctoral dissertation. WH designed and conducted the experiments, analyzed the data, and wrote the paper. YZ offered valuable guidance and assistance and revised every manuscript. WH and YZ discussed and finished the research together. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by China's National Social Sciences Project "Application of Chinese Proficiency Grading Standards and Strategies of Textbook Research and Development" (21STA032).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1144530/full#supplementary-material

References

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/JSS.V067.I01

Bosma, E., and Pablos, L. (2020). Switching direction modulates the engagement of cognitive control in bilingual reading comprehension: an ERP study. *J. Neurolinguistics* 55:100894. doi: 10.1016/j.jneuroling.2020.100894

Casaponsa, A., Thierry, G., and Dunabeitia, J. A. (2020). The role of orthotactics in language switching: an ERP investigation using masked language priming. *Brain Sci.* 10:22. doi: 10.3390/brainsci10010022

Chang, X., Bai, H., and Pei, W. (2017). Factors influencing the cost of language switching for bilinguals. *Adv. Psychol. Sci.* 25, 1469–1478. doi: 10.3724/SP.J.1042.2017.01469

Chang, X., Zhang, G., and Wang, P. (2009). The psychological mechanism of English sentence processing in Chinese second language learners: a case study of active sentences. *Aust. J. Psychol.* 41, 471–480. doi: 10.3724/SP.J.1041.2009.00471

Costa, A., and Santesteban, M. (2004). Lexical access in bilingual speech production: evidence from language switching in highly proficient bilinguals and L2 learners. *J. Mem. Lang.* 50, 491–511. doi: 10.1016/j.jml.2004.02.002

Cui, Z., and Zhang, J. (2010). The mechanism of code switching in Chinese-English bilingual speech comprehension: evidence from sublexical level. *J. Psychol. Sci.* 42, 173–184. doi: 10.3724/SP.J.1041.2010.00173

Dijkstra, T., and van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Bilingual. Lang. Cogn.* 5, 175–197. doi: 10.1017/S1366728902003012

Dryer, M. S. (2013). "Relationship between the order of object and verb and the order of relative clause and noun" in *The World Atlas of Language Structures Online*. eds. M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology)

Feng, L., and Wang, Y. (2013). Difficulty in processing Chinese relative clauses and its constraints for native English speakers. *Lang. Teach. Res.* 5, 1–9.

Grainger, J., and Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *Q. J. Exp. Psychol. Hum. Exp. Psychol.* 39, 295–319. doi: 10.1080/14640748708401788

Gutierrez-Mangado, M. J., and Ezeizabarrena, M. J. (2012). "Asymmetry in child comprehension and production of Basque subject and object relative clauses." in *BUCLD 36 Online Proceedings Supplement*.

He, Z., and Yu, G. (2001). A review of code-switching research. *J. Mod. Foreign Lang.* 24, 85–95. doi: 10.3969/j.issn.1003-6105.2001.01.009

Hu, S., Gavarro, A., Vernice, M., and Guasti, M. T. (2016). The acquisition of Chinese relative clauses: contrasting two theoretical approaches. *J. Child Lang.* 43, 1–21. doi: 10.1017/S0305000914000865

Huang, S., and Lin, C. (2010). The effect of task stimulus type on task switching. *Psychol. Sci.* 33, 1334–1337.

Li, R., Zhang, Z., Ni, C., Xiao, W., Wei, J., Wei, D., et al. (2018). Examining the functional category in Chinese-English code-switching: evidence from the eye-movements. *J. Psycholinguist. Res.* 47, 1–28. doi: 10.1007/s10936-017-9513-z

Linck, J. A., Schwieter, J. W., and Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Biling. Lang. Congn.* 15, 651–662. doi: 10.1017/S136672891100054X

Liu, X., He, C., Chen, J., and Deng, Q. (2015). Bilingual cognitive control mechanisms in fluent Cantonese-general bilingual speakers: evidence from a dual-task switching paradigm. *J. Psychol. Sci.* 47, 439–454. doi: 10.3724/SP.J.1041.2015.00439

Liu, T., Zhou, T., and Yang, Y. (2011). The universality of the processing advantages of relative clauses: evidence from an ERP study of relative clauses in Chinese. *J. Lang. Sci.* 10, 1–20.

Ma, Z. (2012). An empirical study on the second language acquisition of the central language orientation parameter in Chinese relative clause: the hypothesis of the juxtaposition of interlanguage identical classes with rigid central language parameters. *Int. J. Chin. Lang. Teach.* 26, 220–232.

Meuter, R. F. I., and Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *J. Mem. Lang.* 40, 25–40. doi: 10.1006/jmla.1998.2602

Monsell, S. (2003). Task switching. *Trends Cogn. Sci.* 7, 134–140. doi: 10.1016/S1364-6613(03)00028-7

Myers-Scotton, C. (1993). *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford, Eng: Clarendon Press

Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press

Myers-Scotton, C., and Jake, J. L. (2009). "A universal model of code-switching and bilingual language processing and production" in *The Cambridge Handbook of Linguistic Code-Switching*. eds. B. E. Bullock and A. J. Toribio (Cambridge: Cambridge University Press), 336–357.

Ni, W., Fodor, J. D., Crain, S., and Shankweiler, D. (1998). Anomaly detection: eye movement patterns. *J. Psycholinguist. Res.* 27, 515–539. doi: 10.1023/A:1024996828734

Peng, R. (1991). *Psychology of Language*. Beijing: Beijing Normal University Press.

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: The MIT Press.

Schilling, C. J., Storm, B. C., and Anderson, M. C. (2014). Examining the costs and benefits of inhibition in memory retrieval. *Cognition* 133, 358–370. doi: 10.1016/j.cognition.2014.07.003

Tarłowski, A., Wodniecka, Z., and Marzecova, A. (2013). Language switching in the production of phrases. *J. Psycholinguist. Res.* 42, 103–118. doi: 10.1007/s10936-012-9203-9

Thomas, M. S. C., and Allport, A. (2000). Language switching costs in bilingual visual word recognition. *J. Mem. Lang.* 43, 44–66. doi: 10.1006/jmla.1999.2700

Timmermeister, M., Leseman, P., Wijnen, F., and Blom, E. (2020). No bilingual benefits despite relations between language switching and task switching. *Front. Psychol.* 11:1832. doi: 10.3389/fpsyg.2020.01832

Toribio, A. J., and Bullock, B. E. (2009). *Cambridge Handbook of Linguistic Code-Switching*. Cambridge, UK: Cambridge University Press

Wang, H., and Bing, W. (2013). Inconsistency between Chinese relative clause frequency and processing difficulty: evidence from ERP. *Foreign Lang. Stud.* 3, 13–22.

Wang, J., Cong, F., Zhu, F., Guan, Y., and Cui, L. (2018). The Influence of pre-visual processing on punctuation skipping. *Appl Psychol.* 175–182+192. doi: 10.3969/j.issn.1006-6020.2018.02.008

Wei, X., and Chen, Z. (2020). The asymmetry of the influence of clause type and embedding position on the acquisition of relative clauses. *World For. Lang.* 4, 63–70. doi: 10.3969/j.issn.1005-7242.2013.03.002

Xiong, Z. (2005). DP structure with "de" as the core. *J. Contemp. Linguist.* 7:18.

Yim, O., and Bialystok, E. (2012). Degree of conversational code-switching enhances verbal task switching in Cantonese-English bilinguals. *Biling. Lang. Congn.* 15, 873–883. doi: 10.1017/S1366728912000478

Zeller, J. P. (2020). Code-switching does not equal code-switching. An event-related potentials study on switching from L2 German to L1 Russian at prepositions and nouns. *Front. Psychol.* 11:1387. doi: 10.3389/fpsyg.2020.01387

Zhang, Q. (2015). Noun phrase accessibility hierarchy hypothesis and the absolutive hypothesis from the perspective of Chinese relative clauses. *J. Foreign Lang.* 6, 69–74.

Zhang, J., and Cui, Z. (2008). Code switching and its cost in word recognition among Tibetan-Chinese-English bilinguals. *J. Psychol. Sci.* 2, 136–147. doi: 10.1007/BF02656947

Zhang, M., Li, S., Ma, F., Kang, C., and Guo, T. (2020). Mixed language training enhances proactive language control in bilingual language production. *Lang. Cogn. Neurosci.* 36, 152–166. doi: 10.1080/23273798.2020.1803374

Zhang, J., and Wang, Y. (2012). Code-switching mechanisms in proficient Chinese-English bilinguals: phrase-level evidence. *J. Psychol. Sci.* 44, 166–178. doi: 10.3724/SP.J.1041.2012.00166

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 15 June 2023
DOI 10.3389/fpsyg.2023.1142128

OPEN ACCESS

EDITED BY
Linjun Zhang,
Peking University, China

REVIEWED BY
Yingyi Liu,
University of Chester, United Kingdom
Ziyin Mai,
City University of Hong Kong, Hong Kong SAR, China

*CORRESPONDENCE
Stephanie W. Y. Chan
✉ swychan@hku.hk

RECEIVED 11 January 2023
ACCEPTED 17 May 2023
PUBLISHED 15 June 2023

CITATION
Chan SWY, Cheung WM and Marton F (2023) The early language gap between first- and second-language learners: acquisition of Chinese characters among preschoolers. *Front. Psychol.* 14:1142128.
doi: 10.3389/fpsyg.2023.1142128

The early language gap between first- and second-language learners: acquisition of Chinese characters among preschoolers

Stephanie W. Y. Chan[1]*, Wai Ming Cheung[1] and Ference Marton[2]

[1]Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong SAR, China, [2]Department of Pedagogical, Curricular and Professional Studies, University of Gothenburg, Gothenburg, Sweden

For culturally and linguistically diverse children, early second language (L2) development is important for school achievement and social inclusion. These children face challenges in acquiring L2, especially in Hong Kong, where the dominant Chinese language contrasts strongly with their home languages. Studies that compared the language abilities of first language (L1) and L2 students in English-speaking contexts have reported young L2 learners' disadvantage in using the dominant language in oral language and comprehension at school entry. The findings raise the question of whether L2 learners who fall behind their L1 peers in language abilities will be further disadvantaged, showing a weaker development gradient. This study used the Chinese Character Acquisition Assessment (CCAA) to compare character acquisition of 491 L2 children aged from 3 to 6 years against that of 240 of their L1 peers from Hong Kong kindergartens. The CCAA is comprised of six subtests and assesses children's abilities to make associations among character written form (orthography), sound, and meaning. Results showed that L2 learners had greater development in meaning and sound associations across class levels, implying that they may first develop oral language related abilities. In addition, results indicate that diverging gaps between L1 and L2 learners' Chinese character acquisition existed across class levels for the associations involving written character form, but not in regard to associations between character meaning and sound. This study highlights the Chinese learning needs of L2 preschoolers and provides understanding of their abilities in mapping among character written forms, sounds, and meanings. The findings suggest the importance of supporting L2 children's oral language at earlier stages of Chinese learning, and the need to provide instructional support to compensate for their relative weakness in literacy at school entry.

KEYWORDS

Chinese character acquisition, second language (L2), Chinese as a second language (CSL), Chinese character acquisition assessment, culturally and linguistically diverse children, second language acquisition, preschool, early childhood development

1. Introduction

Amid growing international mobility, it is increasingly necessary to understand second language (L2) acquisition of culturally and linguistically diverse children to promote their academic achievement and social integration. In Hong Kong, where Cantonese is spoken by 93.7% of the local population, the ethnic minority population has risen steadily in the past two decades, making up 8.4% of the population (Hong Kong Census and Statistics Department, 2022). Currently, more than half of the local kindergartens have admitted non-native Chinese speaking students (Hong Kong Legislative Council, 2021). These L2 Chinese learners mainly comprise of South Asians and Southeast Asians, including Pakistanis, Nepalese, Filipinos, and Indians, whose home languages include Urdu, Nepali, Hindi, and English. Research reports have documented that South Asians encounter tremendous challenges in acquiring the Chinese language, and the language barriers have limited their educational and career opportunities (e.g., Equal Opportunities Commission, 2012; Bhowmik and Kennedy, 2016). In addition, South Asian households with children experience a higher poverty rate (30.8%) than the local population's average (15.2%) (Research Office of the Hong Kong Legislative Council Secretariat, 2017). The Hong Kong Special Administrative Region government prioritizes L2 students' early Chinese learning as a pathway to adaptation to the Hong Kong education system and social integration (Hong Kong Legislative Council, 2012; Curriculum Development Council, 2017).

Many studies have compared the language abilities of L1 students with their L2 counterparts in school-age populations in contexts speaking English and other alphabetic languages. However, very few efforts have been made to provide empirical evidence on these L2 children's Chinese language development systematically. In reviews of studies on Chinese as second language or foreign language (e.g., Ma et al., 2017; Chan et al., 2022), it has often been raised that the majority of studies are conducted based on school-aged or adult populations, and very few provided understanding on L2 preschoolers' Chinese-language proficiency. To contribute to such understanding, this study explores L2 preschoolers' Chinese character acquisition with reference to L1 Chinese preschoolers in Hong Kong.

1.1. Early language skills of first- and second-language children

In studies of young children in English-speaking contexts, as compared to their first language (L1) counterparts, L2 English learners have been found to have weaker receptive vocabulary (Bialystok et al., 2010; Yesil-Dagli, 2011; Jackson et al., 2014), expressive vocabulary (Limbos and Geva, 2001; Geva and Yaghoub Zadeh, 2006), reading comprehension (Droop and Verhoeven, 2003), and listening comprehension (McKendry and Murphy, 2011). Dutch L2 learners also have smaller vocabulary size than their L1 peers (Verhoeven, 2000). England's large-scale National Pupil Database likewise reveals that among 5-year-olds at the end of reception year, considerably fewer L2 learners than L1 learners demonstrated good developmental reading levels (Strand et al., 2015). When it comes to English letter recognition, however, L2 learners at the same age were able to perform similarly to their L1 peers (Chiappe et al., 2002) or even better than them (Lipka and Siegel, 2007). L2 and L1 learners also do not appear to differ significantly in their ability to read words (e.g., Verhoeven, 2000; Limbos and Geva, 2001; Chiappe et al., 2002) or spell them (for review, see Lesaux et al., 2006), though Grade Two L2 English learners have been found to read isolated words and letters faster than their L1 counterparts (Geva and Yaghoub Zadeh, 2006). These comparative studies have shown that young L2 learners' disadvantage in using the dominant language lies in oral language and comprehension, rather than in word-decoding or spelling abilities.

This raises the question of whether L2 learners who fall behind their L1 peers in certain language abilities will be further disadvantaged, showing a weaker developmental gradient. Empirical studies show various patterns. Supporting the possibility of divergence between these two groups over time, Verhoeven (2000) found that L1–L2 gaps in Dutch vocabulary and reading widened between first and second grade; and when Appel and Vermeer (1998) reviewed several Dutch studies conducted in primary schools, they found a diverging trajectory in vocabulary, even between L2 learners and low-income L1 learners. This is in line with Stanovich's (1986) Matthew Effect on reading and vocabulary, named after a verse in the biblical Book of Matthew that notes that the rich get richer, while the poor get poorer. It thus seems reasonable to speculate that L2 learners' initial vocabulary delay may put them at ever-higher risk of literacy and school-achievement disadvantages with schooling.

However, other studies have identified narrowing gaps, but this often occurred in higher grades. For instance, convergence in English reading achievement among U.S. Hispanic and White children has been observed from kindergarten to fifth grade (Reardon and Galindo, 2009) and from fifth grade to eighth grade (Kieffer, 2012). This narrowing pattern was also found in national datasets covering the same nine grades (Halle et al., 2012). England's National Pupil Database also investigated the differences between L1 and L2 English learners as of 2013, and concluded that the overall reading gap was smaller among 16-year-olds than among 5-year-olds (Strand et al., 2015). Though that comparison was based on the percentages of students attaining an expected level, rather than actual performances, the balance of evidence suggests that L2 learners are more able to develop certain capabilities over time, even though they still rank behind their L1 peers in absolute terms. Nevertheless, there is some evidence that English L2 learners under age 10 are able to catch up with their L1 peers on specific language measures, including decoding, spelling (Lesaux et al., 2006), word reading, word-reading fluency, and phonological processing (Lipka and Siegel, 2007). Similarly, Geva and Yaghoub Zadeh (2006) reported that L1 and L2 English-learners' reading skills were not significantly different by Grade Two. Whether individual L2 children became proficient in the target language at a young age also affected whether they could catch up later, not only in reading itself but in other subjects (e.g., mathematics; Halle et al., 2012). English L2 learners have even been found to surpass their L1 peers in spelling after just a few years of schooling (Lipka and Siegel, 2007).

It cannot be concluded, based on the above-cited inconsistent findings of studies involving different ages of schoolchildren and various school contexts, whether L2 learners' relatively delayed language acquisition is exaggerated or reduced by education in the early years. Thus, it would appear worthwhile to extend these explorations to include the early acquisition of non-alphabetic languages such as Chinese.

1.2. Theoretical framework underlying Chinese word reading and character acquisition

In fact, our understanding of learning to read languages is largely based on studies of alphabetic languages, especially in English (Perfetti et al., 2013; Share, 2021). As research on reading Chinese and other non-alphabetic scripts emerges, there has been increased interest in examining language learning processes across scripts that are different in their correspondence between orthography and phonology (Perfetti et al., 2005; Share, 2021). Although English is considered an opaque script because of its irregularity over transparent alphabetic scripts like Finnish and German, readers can follow alphabetic principles to decode words based on the combination of letters. On the contrary, the Chinese script is morphosyllabic; and most characters map to a morpheme and syllable simultaneously. Chinese learners need to identify the visual information represented within each character, these include phonetic and semantic radicals, orthographical structure, and stroke positions. Research on L1 Chinese preschoolers suggests that unique skills are involved in Chinese character and word reading. According to McBride and Wang (2015), phonological sensitivity, rapid automatized naming, morphological awareness, and visual-orthographic abilities are the core cognitive abilities for Chinese learning and literacy. Studies collectively found that morphological awareness and orthographic knowledge predict later Chinese word reading (e.g., Tong et al., 2009; Lin et al., 2018). In particular, in studies that compared the contributions of multiple cognitive-linguistic skills, orthographical knowledge was found to be the strongest predictor (e.g., McBride-Chang and Ho, 2009; Yang et al., 2019). As studies of early Chinese learning suggest that the underlying abilities may be different as compared to alphabetic scripts owing to the characteristics of the Chinese writing system, understanding of Chinese language development is needed to expand our understanding of the processes that are language specific and universal.

The Lexical Constituency Model (Perfetti et al., 2005) provides a framework to understand word reading across scripts, including Chinese. According to the Lexical Constituency Model (Perfetti et al., 2005), word representations are comprised of three interrelated constituents, namely, orthography, phonology, and semantics. The identity of a word is specified at the value of its pronunciation, orthographic form, and meaning range. Successful word identification involves simultaneous retrieval of all three constituents, such that missing any of the values (pronunciation, orthographic form, or meaning range), including incomplete retrievals, would result in failing to word reading failures.

This understanding of word representation is consistent with theories that conceptualize Chinese character acquisition as mastering the relationships among character written form, sound, and meaning, the three constituents of Chinese characters (Ai, 1949; Tse and Zhu, 2001; Chan et al., 2020). To have acquired a Chinese character, one has to fulfill three conditions: First, when presented with the written form of the character, one is able to produce the sound and meaning of the character; second, when hearing the character sound, one is able to visualize the character form and understand the meaning; third, when having a meaning in mind, one is able to map to the correct character form and produce the sound (Ai, 1949; Tse and Zhu, 2001; Chan et al., 2020). This conceptualization aligns with the Lexical Constituency Model (Perfetti et al., 2005) in the way that semantic, phonological, and orthographic processes are involved in the process of character acquisition; but as McBride (2016) suggested, the processes of character and word reading greatly overlap, but do not constitute identical processes as they involve different skills.

1.3. Early language development of second language Chinese learners in Hong Kong

Preschool experiences have been identified as crucially important to L2 learners' subsequent language acquisition (Hau, 2008; Li and Chuk, 2015), despite their generally low Chinese-language proficiency while in kindergartens in Hong Kong (Hong Kong Unison, 2012). We also know that school-level factors, such as the medium of instruction, relate to L2 Chinese preschoolers' Chinese language proficiency (Tse et al., 2020). The Education Bureau of Hong Kong Special Administrative Region emphasizes the role of kindergartens in preparing children's holistic development for formal schooling. Kindergarten education is offered in three class levels: nursery class (K1) for 3-year-olds, lower kindergarten (K2) for 4-year-olds, and upper kindergarten (K3) for 5-year-olds. The objectives of preschool education within language development include inculcating an interest in Chinese learning, a good communication attitude, and a foundation for language usage (Curriculum Development Council, 2017). Chinese character acquisition is fundamental to achieving these objectives and to language development more generally and is highly prioritized from preschool onward. Given the different metalinguistic skills underlying reading Chinese as compared to reading alphabetic languages, the non-native speakers may experience challenges in acquiring Chinese characters. Specifically, the Guide identified enunciation of character tones, vocabulary, and character recognition and writing as the typical challenges faced by non-native Chinese speaking children, who mostly speak alphabetic or alphasyllabic home languages. In addition, the standard written Chinese language does not map directly to the spoken language of Cantonese (the main spoken dialect in Hong Kong) as words and phrases are represented differently. The inconsistency across forms may pose additional challenges to L2 Chinese learners in Hong Kong.

There is some evidence that language skills related to character acquisition predict later literacy development in L2 Chinese learners. In studies of L2 Chinese children in grades four to six, radical awareness predicted their later character reading (Wong, 2017) and character recognition and orthographic knowledge predicted later reading comprehension (Wong, 2019). Orthographic awareness and morphological awareness are associated with character reading and spelling (Wong and

Zhou, 2022). In addition, listening and reading comprehension skills predicted each other over 2 years (Wong, 2018). In one of the few studies comparing L1 and L2 Chinese learners, Hau (2008) found that L2 learners who were rated better in Chinese in first grade maintained this status through the end of third grade, indicating that early language skills predict later L2 proficiency to an extent. Nearly half of the L2 learners made greater gains in Chinese language than the average L1 learners in the same classes did from first to third grade, possibly implying a narrowing of the L1–L2 gap over those grades. However, around a fifth of the L2 learners continued to score well below the class average throughout the three grades. Intervention studies involving L1 and L2 Chinese children also suggest language development and acquisition processes may be different between the two groups. For example, in a study of second- and third-grade L1 and L2 learners, Wang et al. (2018) found commonalities and differences in the intervention effects on the students' abilities to write Chinese characters: the copying condition was effective for both groups, but the radical knowledge approach only benefited the L1 Chinese group, and the phonological approach supported the L2 Chinese group.

Direct comparisons between L1 and L2 learners of Chinese are rare, even in studies that make claims about the latter's Chinese-language proficiency. Thus, it is not yet possible to deduce whether these two groups' proficiency levels are generally diverging, parallel, or narrowing across age. Identifying language abilities at the earliest stages can build a foundation for the understanding L2 learners' long-term weaknesses and needs. It will also be useful to identify this group's fundamental language abilities through systematic direct assessments, to complement the existing body of evidence based primarily on self-reports, teachers' perceptions (e.g., Loper, 2004; Ku et al., 2005; Shum et al., 2011), school examination scores (e.g., Hau, 2008), or researcher-developed tests (e.g., Tsung et al., 2010).

1.4. The current study

It has often been reported in studies in English-speaking contexts that L2 learners lag behind their L1 peers in emergent language abilities. However, our understanding of L2 learners' Chinese language skills is very limited, even though policies have been directed toward the need to promote these skills as early as possible. To characterize L2 Chinese learners and to identify effective approaches of teaching and learning, knowledge of their early language developmental trajectory is needed. Based on theories that are language-universal and specific to early Chinese character acquisition, comparisons of early language skills should include children's knowledge of orthography, phonology, and semantics of Chinese characters. Do young L2 Chinese children in Hong Kong lag behind L1 children on their Chinese character acquisition? If these L2 children in Hong Kong are delayed at kindergarten entry, will the gap between the two groups of learners be further exaggerated (the Matthew effect), or will the gap be narrowed, with L2 children catching up with their peers as suggested in some studies?

We aim to understand Chinese character acquisition in L2 children in Hong Kong; and to investigate how L2 children compare with L1 children in the 3 years of kindergarten education in terms of Chinese character acquisition in a cross-sectional study. The research questions are as follows: (1) What are the Chinese character acquisition abilities of L2 learners across the three kindergarten class levels (K1, K2, and K3) in Hong Kong? (2) Are there any differences in Chinese character acquisition abilities between L1 and L2 learners across K1, K2, and K3? If so, what are they?

2. Materials and methods

2.1. Participants

Participants comprised 491 L2 and 240 L1 learners from 12 kindergartens in Hong Kong. The L2 learners included 142 children from K1 (75 girls, M_{age} = 44.78 months, SD = 6.74), 176 from K2 (96 girls, M_{age} = 56.89 months, SD = 6.07), and 173 from K3 classes (95 girls, M_{age} = 66.60 months, SD = 7.14). They all spoke non-Chinese home languages and had diverse ethnic backgrounds (166 Pakistani, 148 Nepali, 59 Filipino, and 71 Indian). The Chinese L1 learners included 55 from K1 (38 girls, M_{age} = 45.79 months, SD = 5.95), 102 from K2 (47 girls, M_{age} = 56.38 months, SD = 5.28), and 83 from K3 classes (45 girls, M_{age} = 66.90 months, SD = 5.92) from the same kindergartens as the L2 participants. The L1 and L2 participants within each class level did not differ significantly in age in months ($ps > 0.40$).

The 12 kindergartens were drawn from the three territories of Hong Kong. Among these kindergartens, two were located in relatively higher poverty rate districts and two were located in relatively lower poverty rate districts (Hong Kong Census and Statistics Department, 2021). The L2 learners enrolled in the 12 kindergartens accounted for between 4.2 and 100% (M = 37.47%) of their total student populations. Five of these kindergartens had a low proportion of L2 learners (below 20%), three had a high proportion (above 80%), with the remaining four had a moderate proportion. All participating were local kindergartens in Hong Kong that followed the curriculum guidelines issued by the Hong Kong Education Bureau for kindergarten education. Under the Biliteracy and Trilingualism language policy, local kindergartens promote children's development in Chinese language (Cantonese and Mandarin), and English as a second language. The local curriculum guide advocates a child-centered integrated approach that promotes children's learning and development through play, such that languages are not taught at subject level but are promoted through an integrated curriculum alongside other learning areas (Curriculum Development Council, 2006, 2017).

2.2. Instrument

We used the six-subtest Chinese Character Acquisition Assessment (CCAA; Chan et al., 2020), a validated measure of Chinese character acquisition for L2 preschoolers. The measure operationalized character acquisition as six abilities, in line with Tse and Zhu's (2001) conceptualization. Each CCAA task involves one association between two of the three constituents (written character form, sound, or meaning; see **Figure 1**) of 36 Chinese characters.

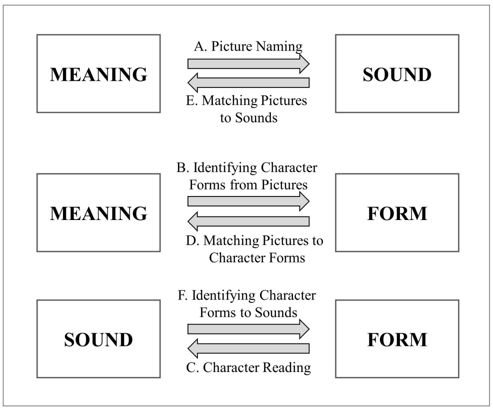


FIGURE 1
The six Chinese Character Acquisition Assessment (CCAA) tasks.

In subtest A: Picture naming (meaning to sound), children are presented with individual pictures, and are asked to say out the characters represented. In subtest B: Identifying character forms from pictures (meaning to form), children select one of four written characters that they feel best corresponds to a picture shown. In subtest C: Character reading (form to sound), children read out written character forms individually. In subtest D: Matching pictures to character forms (form to meaning) involves children choosing one picture from among four that carries the same meaning as a written character form they are shown. In subtest E Matching pictures to sounds (sound to meaning), they choose one of four pictures corresponding to the meaning of an audio-recorded character they hear; and in subtest F: Identifying character forms from sounds (sound to form), they choose one from four written characters that corresponds to the audio-recorded character. The tasks for subtests A and C were open-ended, with a partial-scoring scheme (0 incorrect, 0.5 = partially correct, 1 = correct), whereas subtests B, D, E, and F were multiple-choice tasks and scored dichotomously (0 = incorrect; 1 = correct). The tested characters were selected systematically from a list of 200 characters that frequently appeared in the teaching plans of Hong Kong preschools attended by L2 learners. To ensure the children understand each task, children were given one trial item per subtest as part of the instructions, with additional Cantonese and English instructions where necessary. The internal consistency reliabilities (Cronbach's alpha) of each subtest for the L1 and L2 participants ranged from 0.92 to 0.98 and from 0.75 to 0.97, respectively. The development and validation of the measure is detailed in Chan et al. (2020).

2.3. Procedures

Ethical approval was obtained from the Human Research Ethics Committee of the University of Hong Kong. The six CCAA subtests were individually administered to each child over two consecutive days by trained researchers with their parents' and school principals' informed consent; each session lasted for around 15 min. Instructions were provided in Cantonese and English to ensure both L1 and L2 children understood the tasks. The trained researchers included graduates and undergraduates of psychology and education majors who were bilingual in Cantonese and English and attended a training session on test administration. Among the tests conducted, the performance of around 10% (50 tests) randomly selected children were rated by two independent assessors in order to ensure inter-rater reliability. Pearson product-moment correlation coefficients were significant, ranging from $r = 0.99$ to 1.00, reflecting a high degree of agreement among assessors.

2.4. Data treatment

To facilitate comparison among all six subtests involving the two response types (open-ended response and multiple-choice response), we adjusted the scores of the four multiple-choice subtests to account for random guessing by deducting from the total score [the number of incorrect responses divided by three], as recommended by Lord (1975). Then, we computed each child's composite score out of 216 (36 characters in each of the six subtests) by summing his/her six subtest scores.

3. Results

Table 1 presents descriptive statistics for CCAA scores for L1 and L2 learners. We conducted a series of one-way Analyses of Variance (ANOVAs) to determine the main effects of demographic variables on L2 learners' scores on each subtest. The main effect of ethnicity was significant for subtest A ($p = 0.033$), and the main effects of school type (proportion of L2 learners) were significant for all subtests except B ($ps < 0.037$). There were no main effects of gender for all six subtests ($ps > 0.081$). For the L1 learners, no main effects were observed for gender ($ps > 0.159$), and school type was significant only for subtest F, and only marginally ($p = 0.049$).

We conducted two-way mixed Analysis of Covariance (ANCOVA) to compare L2 learners' scores by class level, with ethnicity and school type controlled as covariates. After Greenhouse-Geisser correction, the Level × Subtest interaction was significant [$F_{(6.18,1295.31)} = 19.81$, $p < 0.001$, partial $\eta^2 = 0.086$, $\iota\varepsilon = 0.62$]. Subsequent one-way ANOVAs showed significant main effects of class level on all six subtests, and Bonferroni's *post-hoc* analyses indicated that the differences between K1 and K2, and between K2 and K3, were significant across all subtests. However, main effects of subtest were significant only for two levels, K2 [$F_{(2.82,444.77)} = 27.47$, $p < 0.001$, partial $\eta^2 = 0.15$, $\iota\varepsilon = 0.58$] and K3 [$F_{(3.05,466.94)} = 42.58$, $p < 0.001$, partial $\eta^2 = 0.22$, $\iota\varepsilon = 0.63$]. In K2 and K3, scores were higher on subtest A (Picture naming) than on any other subtest ($ps < 0.029$), and on subtest E (Matching Pictures to Sounds) than on subtests B, C, D, and F ($ps < 0.001$). Additionally, K3 children's scores were higher on subtest B (Identifying Character Forms from Pictures) than on subtests C, D, and F ($ps < 0.001$).

We conducted a 3 (level) × 2 (language-group) × 6 (subtest) three-way mixed ANCOVA to examine the relations between class level and subtest in the two language groups, controlling for school type. A significant interaction effect of class level, language group, and subtest was detected using the Greenhouse-Geisser correction method [$F_{(5.49,1784.88)} = 58.22$, $p < 0.001$, partial $\eta^2 = 0.152$,

TABLE 1 Means and standard deviations of L1 and L2 participants' CCAA scores, by class level.

Subtest	Association	L2 prticipants (n = 491)							L1 participants (n = 240)						
			K1 (n = 142)		K2 (n = 176)		K3 (n = 173)			K1 (n = 54)		K2 (n = 102)		K3 (n = 82)	
		Rel (α)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Rel	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)
(A) Picture naming	Meaning to sound	0.97	0.00–16.00	1.42 (2.79)	0.00–34.00	8.14 (8.92)	0.00–36.00	15.51 (10.41)	0.93	16.00–36.00	29.37 (5.60)	19.50–36.00	33.51 (3.09)	23.50–36.00	34.96 (1.92)
(B) Identifying character forms from pictures	Meaning to form	0.79	−6.67–16.00	0.46 (3.21)	−6.67–28.00	3.01 (6.38)	−5.33–30.00	9.14 (8.95)	0.94	−5.33–34.67	5.28 (7.63)	−1.33–36.00	17.87 (10.40)	4.00–36.00	28.05 (8.57)
(C) Character reading	Form to sound	0.95	0.00–6.00	0.16 (0.65)	0.00–25.00	2.58 (4.97)	0.00–27.50	7.16 (7.29)	0.98	0.00–35.50	5.09 (7.30)	0.00–36.00	16.74 (10.60)	4.00–36.00	27.66 (8.27)
(D) Matching pictures to character forms	Form to meaning	0.80	−6.67–6.67	−0.96 (2.77)	−9.33–24.00	2.32 (6.89)	−6.67–30.67	7.48 (8.66)	0.95	−6.67–36.00	5.33 (8.86)	−5.33–36.00	16.33 (11.97)	4.00–36.00	28.14 (9.16)
(E) Matching pictures to sounds	Sound to meaning	0.86	−6.67–10.67	0.78 (3.45)	−5.33–26.67	6.76 (7.90)	−5.33–33.33	14.05 (9.67)	0.92	−4.00–34.67	19.04 (10.07)	−2.67–36.00	27.07 (9.12)	20.67–36.00	31.22 (3.81)
(F) Identifying character forms to sounds	Sound to form	0.75	−8.00–8.00	0.19 (2.85)	−8.00–28.00	3.02 (6.24)	−6.67–28.00	7.17 (8.48)	0.94	−5.33–34.67	5.04 (8.35)	−4.00–36.00	16.25 (10.81)	5.33–36.00	26.96 (8.00)
		0.97	−18.67–59.33	2.44 (8.91)	−13.00–147.33	26.32 (35.39)	−9.00–184.50	60.51 (47.35)	0.99	1.00–208.17	70.05 (39.95)	35.00–212.00	129.03 (49.86)	81.83–214.67	176.24 (35.87)

Data for subtests A and C (open-ended) were raw scores and data for subtests B, D, E, and F (multiple choice) were adjusted for random guessing. Subtest maximum score was 36, and the maximum composite score was 216. L1, first language; L2, second language; Rel, reliabilities.

$\iota\varepsilon$ = 0.55]. The relations between class level and language group in each subtest were also explored using two-way ANCOVAs. Interaction effects between class level and language group were significant for all subtests except E. Pairwise *post-hoc* analyses with Bonferroni's adjustment were performed within each level and subtest and indicated significant differences between the two language groups at each level. The magnitude of L1–L2 gap for K2 was larger than K1's for subtests B, C, D, and F, but not different for subtests A and E. K3's L1–L2 gap, meanwhile, was larger than K2's for subtests B, C, D, and F, but smaller for subtests A and E. The subtest score comparisons and patterns are presented in **Table 2** and **Figure 2**, respectively.

As reported in the demographic analysis, we found that the effect of ethnicity was significant for subtest A (Picture naming). Hence, we conducted *post-hoc* analyses with Bonferroni's corrections and found ethnic group differences within the subtest. Specifically, Pakistani children ($M_{adjusted}$ = 10.62, SE = 0.64, 95% CI [9.36, 11.88]) had higher scores than Filipino ($M_{adjusted}$ = 5.36, SE = 1.05, 95% CI [3.30, 7.42], p < 0.001) and Indian children ($M_{adjusted}$ = 6.96, SE = 1.09, 95% CI [4.83, 9.09], p = 0.02).

4. Discussion

In this study, we described L2 learners' Chinese character acquisition and compared their performance against their L1 peers for K1, K2, and K3 children. The study contributes to our understanding of L2 learners' acquisition by providing a systematic comparison of the character acquisition of L2 and L1 Chinese children in terms of character orthography, phonology, and semantics.

4.1. Second language learners' Chinese character acquisition across class levels

Regarding L2 learners' performance on the test, we found low average scores across all CCAA subtests taken by the K1 L2 group (M = −0.96 to 1.42). Specifically, the average K1 L2 child had acquired fewer than two of the 36 tested characters, suggesting that most of them had yet to begin acquiring Chinese characters when they entered preschool, and perhaps that they had received very little Chinese-language input outside school. Their parents' limited Chinese language proficiency may be an additional explanation, as reported in a prior interview-based study (Lisenby, 2011). However, L2 learners in K2 and K3 scored higher on all six subtests than learners in K1, supporting the idea that L2 learners may have progressed in all aspects of Chinese character acquisition upon receiving preschool education. Meanwhile, our findings also suggest that there may be differences among L2 learners, as reflected by the ethnic group differences within the Picture Naming subtest.

However, class-level differences were not uniform across subtests, as evidenced by the interaction effect between the subtest and class level that we found. Within-level explorations also revealed a possible developmental sequence for acquiring Chinese characters. That is, while there were no inter-subtest differences in the K1 L2 group, subtest A (meaning to sound) was the highest-scored subtest for both the K2 and K3 groups, followed

TABLE 2 Inter-group subtest score comparisons, by class level (N = 731).

		K1		K2		K3		Level × Lang interaction	Interaction contrast estimate between level	
Subtest		*M*	95% CI	*M*	95% CI	*M*	95% CI	*F*-value	K2−K1	K3−K2
(A) Picture naming (Meaning to Sound)	L2	1.84 (0.59)	0.67, 3.00	8.35 (0.53)	[7.30, 9.39]	15.71 (0.54)	[14.65, 16.77]	$F_{(2,711)} = 18.76$*** $p < 0.001$ $\eta_p^2 = 0.050$	−2.22 (1.42) [−5.00, 0.56] $F_{(1,711)} = 2.45$ $p = 0.118$ $\eta_p^2 = 0.003$	−6.04 (1.28) [−8.56, −3.53] $F_{(1,711)} = 22.27$*** $p < 0.001$ $\eta_p^2 = 0.030$
	L1	28.77 (0.95)	[26.89, 30.64]	33.06 (0.70)	[31.68, 34.43]	34.37 (0.78)	[32.85, 35.90]			
	L1-L2	26.93 (1.13)	[24.71, 29.15]	24.71 (0.88)	[22.98, 26.53]	18.67 (0.95)	[16.80, 20.53]			
		$p < 0.001$***		$p < 0.001$***		$p < 0.001$***				
(B) Identifying character forms from pictures (meaning to form)	L2	0.44 (0.70)	[−0.91, 1.81]	3.00 (0.59)	[1.83, 4.16]	9.13 (0.60)	[7.95, 10.30]	$F_{(2,694)} = 37.38$*** $p < 0.001$ $\eta_p^2 = 0.097$	10.03 (1.60) [6.88, 13.17] $F_{(1,694)} = 39.13$*** $p < 0.001$ $\eta_p^2 = 0.053$	4.05 (1.42) [1.27, 6.84] $F_{(1,694)} = 8.16$** $p = 0.004$ $\eta_p^2 = 0.012$
	L1	5.32 (1.08)	[3.20, 7.44]	17.90 (0.77)	[16.38, 19.41]	28.08 (0.86)	[26.40, 29.77]			
	L1-L2	4.88 (1.30)	[2.34, 7.42]	14.90 (0.98)	[12.99, 16.82]	18.96 (1.05)	[16.90, 21.02]			
		$p < 0.001$***		$p < 0.001$***		$p < 0.001$***				
(C) Character reading (form to sound)	L2	0.19 (0.58)	[−0.96, 1.34]	2.59 (0.52)	[1.57, 3.62]	7.18 (0.53)	[6.14, 8.21]	$F_{(2,704)} = 58.77$*** $p < 0.001$ $\eta_p^2 = 0.143$	9.26 (1.40) [6.51, 12.00] $F_{(1,704)} = 43.80$*** $p < 0.001$ $\eta_p^2 = 0.059$	6.33 (1.25) [3.87, 8.79] $F_{(1,704)} = 25.53$*** $p < 0.001$ $\eta_p^2 = 0.035$
	L1	5.05 (0.95)	[3.19, 6.91]	16.71 (0.68)	[15.37, 18.05]	27.63 (0.76)	[26.14, 29.12]			
	L1–L2	4.86 (1.12)	[2.66, 7.06]	14.12 (0.86)	[12.42, 15.81]	20.45 (0.93)	[18.62, 22.27]			
		$p < 0.001$***		$p < 0.001$***		$p < 0.001$***				
(D) Matching pictures to character forms (form to meaning)	L2	−0.94 (0.72)	[−2.36, 0.48]	2.33 (0.63)	[1.10, 3.57]	7.49 (0.64)	[6.24, 8.75]	$F_{(2,690)} = 34.48$*** $p < 0.001$ $\eta_p^2 = 0.091$	7.73 (1.69) [4.42, 11.05] $F_{(1,690)} = 21.01$*** $p < 0.001$ $\eta_p^2 = 0.030$	6.65 (1.52) [3.67, 9.63] $F_{(1,690)} = 19.21$*** $p < 0.001$ $\eta_p^2 = 0.027$
	L1	5.30 (1.13)	[3.07, 7.52]	16.30 (0.83)	[14.68, 17.92]	28.11 (0.92)	[26.31, 29.91]			
	L1-L2	6.23 (1.35)	[3.58, 8.89]	13.97 (1.04)	[11.92, 16.02]	20.62 (1.12)	[18.41, 22.83]			
		$p < 0.001$***		$p < 0.001$***		$p < 0.001$***				
(E) Matching pictures to sounds (sound to meaning)	L2	1.02 (0.68)	[−0.31, 2.36]	6.89 (0.59)	[5.73, 8.06]	14.19 (0.60)	[13.00, 15.37]	$F_{(2,699)} = 2.65$ $p = 0.071$ $\eta_p^2 = 0.008$	2.24 (1.59) [−0.88, 5.35] $F_{(1,699)} = 1.99$ $p = 0.161$ $\eta_p^2 = 0.003$	−3.21 (1.43) [0.40, 6.01] $F_{(1,699)} = 5.04$* $p = 0.025$ $\eta_p^2 = 0.007$
	L1	18.67 (1.06)	[16.59, 20.76]	26.78 (0.78)	[25.25, 28.32]	30.87 (0.86)	[29.17, 32.56]			
	L1-L2	17.65 (1.27)	[15.16, 20.15]	19.89 (0.99)	[17.96, 21.82]	16.68 (1.06)	[14.60, 18.76]			
		$p < 0.001$***		$p < 0.001$***		$p < 0.001$***				
(F) Identifying character forms to sounds (sound to form)	L2	0.20 (0.67)	[−1.12, 1.52]	3.02 (0.58)	[1.89, 4.16]	7.17 (0.85)	[6.02, 8.32]	$F_{(2,698)} = 43.79$*** $p < 0.001$ $\eta_p^2 = 0.111$	8.38 (1.56) [5.33, 11.44] $F_{(1,698)} = 28.98$*** $p < 0.001$ $\eta_p^2 = 0.040$	6.57 (1.40) [3.82, 9.31] $F_{(1,698)} = 22.06$*** $p < 0.001$ $\eta_p^2 = 0.031$
	L1	5.04 (1.04)	[3.00, 7.08]	16.25 (0.77)	[14.74, 17.75]	26.96 (0.85)	[25.30, 28.62]			
	L1-L2	4.84 (1.25)	[2.39, 7.29]	13.22 (0.96)	[11.33, 15.11]	19.79 (1.04)	[17.75, 21.82]			
		$p < 0.001$***		$p < 0.001$***		$p < 0.001$***				

L1–L2 = pairwise comparisons between L1 and L2 learners; Level × Lang = interaction between level and language group; school types were evaluated as model covariates; mean differences were based on estimated marginal means; 95% confidence intervals are in square brackets; standard errors are in parentheses. CI, confidence interval; η_p^2, partial eta squared.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

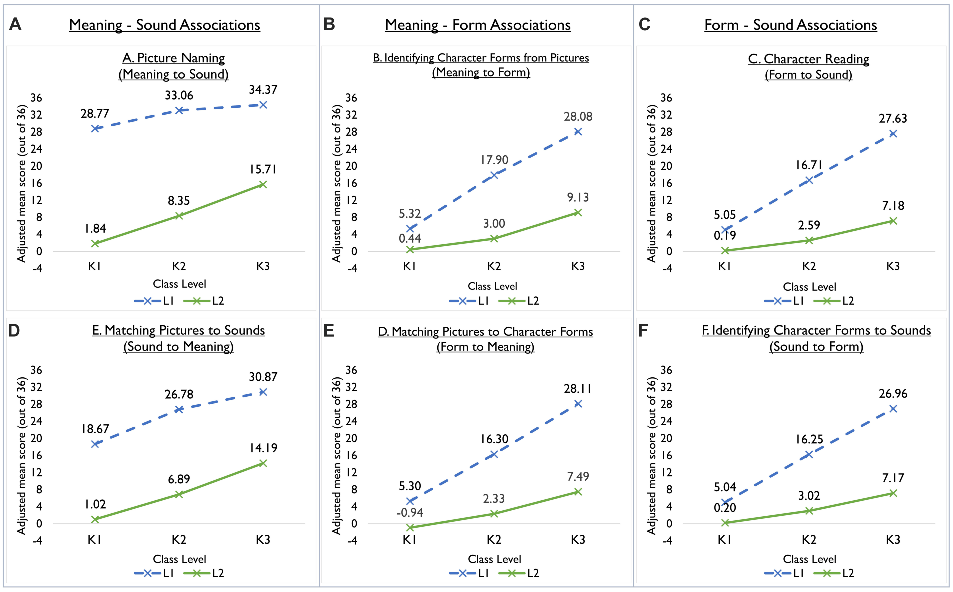


FIGURE 2
Adjusted means of L2 and L1 learners in three class levels for the six subtests.

by subtest E (sound to meaning). Second-language learners also exhibited sharper score differences across K1 and K2 in subtests A and E (meaning-sound associations), than in subtests B, C, D, and F (character-form associations), implying that beginning L2 learners first develop the oral aspects underlying Chinese character acquisition. This is in line with the consensus that literacy development relies, in the first instance, on listening and speaking skills (e.g., Berninger, 2000).

4.2. The early language gap between first- and second-language learners

More importantly, following our speculation that L2 learners might experience character acquisition delays throughout their preschool years, we indeed found substantial gaps between L1 and L2 learners' performance on all subtests at all class levels. Regarding the hypothesized patterns of divergence and convergence or the lack thereof, we found that—although L2 learners had lower scores than L1 learners at all points—the patterns of the significant interaction effects among class level, language group, and subtest indicated that L1-L2 group differences in these three levels were not uniform for all abilities.

Further examination based on the Chinese character acquisition framework (Ai, 1949; Tse and Zhu, 2001; Chan et al., 2020) led us to identify two patterns, a unique contribution of the study. The first pattern relates to children's knowledge of associations involving character written forms (orthography), and is displayed in subtests B, C, D, and F. At K1, the L1-L2 gaps for these four subtests were smaller than those of subtests A and E (both of which tested meaning-sound associations), with adjusted mean differences of between 4.52 and 5.77. Gap sizes in K2 and K3 were also found to be larger than those in K1 and K2, respectively, indicating that L1 learners may benefit much more than L2 learners at each level, and that the latter group is likely to fall further behind as primary-school entry approaches. These results, illustrated in **Figure 2**, resemble the diverging pattern in Stanovich's (1986) discussion of the Matthew effect. Initially, our sampled L1 and L2 learners were relatively close on the written form of characters, but the L1 learners' slight advantage left them "richer"; and as this advantage continued to grow, the L2 learners became relatively "poorer." This effect may threaten L2 learners' development of literacy-related abilities in formal schooling. Notably, the L1–L2 gaps for subtests B, C, D and F widened more between K1 and K2 than between K2 and K3, as reflected by the effect sizes in our interaction-contrast estimates. It is possible that the measure was unable to fully capture L1 learners' acquisition abilities beyond the specific range of characters tested, and hence we may have underestimated the gap. Or conversely, the gap increase could have been more intense from K1 to K2 than from K2 to K3 among those characters taught in preschools.

The second main pattern we observed was in the L1–L2 score gaps for subtests A and E. These subtests tested associations between meaning and sound, which inclines more to oral abilities than the other four. At the word level, such abilities resemble expressive and receptive vocabulary. Among all subtests, the mean differences between K1's L1 and L2 learners were particularly large in the cases of subtests A (M_{L1-L2} = 25.55) and E (M_{L1-L2} = 16.85), indicating that L2 learners' development of the oral aspects of Chinese character acquisition was markedly delayed. This finding is consistent with previous findings that L2 kindergartners and first graders were weaker in vocabulary than their L1 counterparts (e.g., Geva and Yaghoub Zadeh, 2006; Lipka and Siegel, 2007; Bialystok et al., 2010; Yesil-Dagli, 2011). As observed in English-speaking contexts, L1 learners have already started to develop sound-meaning and meaning-sound associations before preschool, possibly through home literacy activities (McCoy and Cole, 2011); and Chinese infants as young as 6 months start to develop tones (Mattock and Burnham, 2006). As such, the largest component of the L1-L2 gap in oracy at preschool entry may be attributable to L1 children having already picked up these abilities at home. A diverging L1–L2 pattern was not observed in subtests A and E. Instead, gap sizes were maintained from K1 to K2, and reduced slightly from K2 to K3. This may imply a widening of abilities, or our measures may have underestimated the abilities of K2 and K3 L1 learners because of ceiling effects. Nevertheless, the sampled L2 learners' oral abilities around the tested Chinese characters progressed noticeably from level to level.

To sum up, diverging gaps were observed for abilities related to the character written form, but not for oracy-related ones, seemingly because it was more difficult for L2 learners to keep up with their L1 peers on measures related to the former. Our findings differ from Lesaux et al.'s (2006) findings that L2 learners' decoding skills can catch up with those of their L1 counterparts via just one or 2 years of preschool and/or first-grade education. In this study, widening gaps were found across class levels for character reading (subtest C) and matching character forms to sounds (subtest F). This echoes the suggestion that it takes longer for L2 learners to acquire Chinese, as compared to alphabetic languages (McBride, 2016), and indicates that this may be due to the difficulty of acquiring orthography-phonology correspondence.

4.3. What advantage do L1 learners have over L2 learners in learning written language?

Our identification of diverging gaps in character knowledge of character forms raises the question, "What advantage do L1 learners have over L2 learners in learning written language?" One possibility may be deduced from the pattern we observed regarding subtests A and E: L1 learners' advantage may lie in their initial superiority in the spoken aspect of Chinese character acquisition. The orthography-related subtests present a pattern resembling the Matthew effect, whereas in the spoken aspect of character acquisition, L2 learners are less disadvantaged. This echoes the findings of alphabetic language studies, that L2 learners need to develop oral language to a certain extent to acquire literacy, much as their L1 peers do (Verhoeven, 2000; Geva and Yaghoub Zadeh, 2006). Thus, support for L2 learners' early childhood oral-language acquisition will be crucial to closing the L1–L2 literacy gap.

The two patterns we identified may provide some hints about the relationship between the processing of spoken and written Chinese. As compared to their L1 peers, our L2 learners appeared more able to keep up level by level with changes in the range of characters tested when it came to oracy, but less able when it came to literacy aspects of the same characters. While we are unable to draw any firm conclusions, it is conceivable that, for young L2 learners, the oracy-literacy relationship in Chinese may not be as strong as it is in alphabetic languages.

We can also speculate that the widening gaps we observed, especially in the orthographic aspect, may further disadvantage L2 learners in reading comprehension when they enter first grade. According to the Simple View of Reading (Gough and Tunmer, 1986), reading comprehension is a product of word decoding and linguistic comprehension, but the roles of these two components vary, both across stages of language learning and orthographies (Tobai and Bonifacci, 2015; Cadime et al., 2017). For beginning learners of opaque orthographies like English, the role of word decoding is stronger than listening comprehension, unlike with transparent orthographies (Florit and Cain, 2011). Thus, it is highly likely that the role of decoding in reading comprehension will be stronger still when learning Chinese orthography, which has an even more opaque orthography than English. As noted earlier, Wong (2019) found that Chinese-decoding strongly predicted reading, as well as improvement thereof 2 years later, later among ethnic minority fourth-graders in Hong Kong. Decoding involves associating word forms with sounds, so if L2 learners increasingly fall behind L1 learners in form-sound and form-meaning associations between K1 and K3, such a disadvantage is likely to be carried forward to first grade, and negatively impact their reading comprehension. Given L2 learners' potential constraints to decoding when entering first grade, the findings support the design of interventions targeted for oral language to support their acquisition of the spoken and written abilities underlying Chinese character acquisition at the kindergarten level. In addition, reading-comprehension strategies could be developed specifically for L2 learners to compensate for their relative weaknesses in decoding at first grade.

4.4. Limitations and future research directions

The present study has several limitations. First, as L2 learners are highly diverse, our sample may not represent all L2 learners. Second, our cross-sectional design limited us from drawing firm conclusions regarding trends in Chinese character acquisition over time. Future research should therefore follow the longitudinal development of individuals' Chinese character acquisition across levels or ages. Third, the potential ceiling effect for the two oracy-related subtests (subtests A and E) may have underestimated the L1 Chinese children's abilities. Future iterations of the measure include additional items for children who score close to full scores on these subtests. Fourth, we focused solely on Chinese character acquisition, at the expense of other important abilities and factors including early language skills in Chinese (e.g., phonological awareness and orthographic awareness) and other languages (e.g., English and home languages); language proficiency in other languages; cognitive abilities; home factors (e.g., socioeconomic status, parental language proficiency and involvement); and demographic attributes. These factors have been identified in the emerging studies of L2 learners to be related to their language abilities. It is also worth highlighting that the variability between L1 and L2 learners was greater at higher class levels. Standard deviations of the subtest scores were 0.65 to 3.45 in K1; 4.97 to 8.92 in K2; and 7.29 to 10.41 in K3. Our study has described *what* the L1-L2 differences were but not *how* these differences emerged, and it will be useful in future to investigate the underlying factors or reasons that account for these differences.

5. Concluding remarks

Due to globalization, learning languages other than one's own—notably, Chinese—is increasingly important and popular. However, the unique orthographic, phonological, and semantic features of Chinese script render it necessary to measure and describe the language abilities of L2 Chinese children in new ways that differ from the body of the existing literature on early L2 language development. This study has helped to address that problem, and extended explorations of L2 Chinese learners' language abilities to include the preschool stage. Findings call for educators' and researchers' attention to, and further exploration of, L2 Chinese learners' distinctive characteristics, and the development of teaching practices that address their needs.

Data availability statement

The datasets presented in this article are not readily available because consent obtained on data usage was limited to the research team. Requests to access the datasets should be directed to SC, swychan@hku.hk.

Ethics statement

The studies involving human participants were reviewed and approved by the Human Research Ethics Committee of the University of Hong Kong. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

WC and SC collected the data. SC performed the data analysis and wrote the first draft. All authors conceptualized the study, reviewed and commented on the manuscript.

Funding

This research was supported by the Hong Kong General Research Fund (General Research Fund under file no. 1760615). Publication made possible in part by support from the HKU Libraries Open Access Authors Fund sponsored by the HKU Libraries.

Acknowledgments

We wish to thank the participating schools, families, and children, and the research assistants who supported the data collection of the study. Part of the study was included in a doctoral thesis submitted to the University of Hong Kong by SC.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ai, W. (1949). *Yue Du Xin Li; Han Zi Wen Ti [The Psychology of Reading; Issues Related to Chinese Characters]. Guo Li Bianyi Guan*

Appel, R., and Vermeer, A. (1998). Speeding up second language vocabulary acquisition of minority children. *Lang. Educ.* 12, 159–173. doi: 10.1080/09500789808666746

Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Top. Lang. Disord.* 20, 65–84. doi: 10.1097/00011363-200020040-00007

Bhowmik, M. K., and Kennedy, K. J. (2016). *Out of School Ethnic Minority Young People in Hong Kong*. Singapore: Springer.

Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Biling* 13, 525–531. doi: 10.1017/S1366728909990423

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., Cosme, M. D. C., et al. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Read. Writ.* 30, 591–611. doi: 10.1007/s11145-016-9691-3

Chan, J., Woore, R., Molway, L., and Mutton, T. (2022). Learning and teaching Chinese as a foreign language: A scoping review. *Rev. Educ.* 10:e3370. doi: 10.1002/rev3.3370

Chan, S. W. Y., Cheung, W. M., Huang, Y., Lam, W.-I., and Lin, C.-H. (2020). Development and validation of a Chinese character acquisition assessment for second-language kindergarteners. *Lang. Test.* 37, 215–234. doi: 10.1177/0265532219876527

Chiappe, P., Siegel, L. S., and Gottardo, A. (2002). Reading-related skills of kindergarteners from diverse linguistic backgrounds. *Appl. Psycholinguist.* 23, 95–116. doi: 10.1017/S014271640200005X

Curriculum Development Council, (2006). *Guide to the Pre-primary Curriculum*. Available online at: https://www.edb.gov.hk/attachment/en/curriculum-development/major-level-of-edu/preprimary/pre-primaryguide-net_en.pdf (accessed April 26, 2023).

Curriculum Development Council (2017). *Kindergarten Education Curriculum Guide: Joyful Learning Through Play, Balanced Development All The Way*. Available online at: https://www.edb.gov.hk/attachment/en/curriculum-development/major-level-of-edu/preprimary/ENG_KGECG_2017.pdf (accessed January 6, 2023).

Droop, M., and Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Read. Res. Q.* 38, 78–103. doi: 10.1598/RRQ.38.1.4

Equal Opportunities Commission (2012). *Study on Racial Encounters and Discrimination Experienced by South Asians*. Available online at: https://www.eoc.org.hk/EOC/Upload/UserFiles/File/ResearchReport/201203/Race_eFull%20Report.pdf (accessed January 6, 2023).

Florit, E., and Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educ. Psychol. Rev.* 23, 553–576. doi: 10.1007/s10648-011-9175-6

Geva, E., and Yaghoub Zadeh, Z. (2006). Reading efficiency in native English-speaking and English-as-a-second-language children: The role of oral proficiency and underlying cognitive-linguistic processes. *Sci. Stud. Read.* 10, 31–57. doi: 10.1207/s1532799xssr1001_3

Gough, P. B., and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *RASE* 7, 6–10. doi: 10.1177/074193258600700104

Halle, T., Hair, E., Wandner, L., McNamara, M., and Chien, N. (2012). Predictors and outcomes of early versus later English language proficiency among English language learners. *Early Child. Res. Q.* 27, 1–20. doi: 10.1016/j.ecresq.2011.07.004

Hau, K. T. (2008). *Tracking the Adaptation and Development of Non-Chinese Speaking Children (NCS) in Mainstream Schools*. Hong Kong: Education Bureau.

Hong Kong Census and Statistics Department (2021). *Hong Kong Poverty Situation Report 2020*. Availble online at: https://www.censtatd.gov.hk/en/data/stat_report/product/B9XX0005/att/B9XX0005E2020AN20E0100.pdf (accessed April 26, 2023).

Hong Kong Census and Statistics Department (2022). *Population Census 2021 – Thematic Report 2021: Ethnic Minority*. Availble online at: https://www.censtatd.gov.hk/en/data/stat_report/product/B1120111/att/B11201112021XXXXB0100.pdf (accessed January 6, 2023).

Hong Kong Legislative Council (2012). *LCQ18: Chinese Learning Support for Ethnic Minority Children*. Hong Kong: Hong Kong Legislative Council.

Hong Kong Legislative Council (2021). *Legislative Council Panel on Education: Progress on Supporting Chinese Learning and Teaching for Non-Chinese Speaking Students: LC Paper No. CB(4)1455/20-21(05)*. Availble online at: https://www.legco.gov.hk/yr20-21/english/panels/ed/papers/ed20210903cb4-1455-5-e.pdf (accessed January 6, 2023).

Hong Kong Unison (2012). *Report: Survey on Kindergarten Education for Ethnic Minority Students in Hong Kong*. Hong Kong: Hong Kong Unison.

Jackson, C. W., Schatschneider, C., and Leacox, L. (2014). Longitudinal analysis of receptive vocabulary growth in young Spanish English–speaking children from migrant families. *Lang. Speech. Hear. Serv. Schl.* 45, 40–51. doi: 10.1044/2013_LSHSS-12-0104

Kieffer, M. J. (2012). Early oral language and later reading development in Spanish-speaking English language learners: Evidence from a nine-year longitudinal study. *J. Appl. Dev. Psychol.* 33, 146–157. doi: 10.1016/j.appdev.2012.02.003

Ku, H. B., Chan, K. W., and Sandhu, K. K. (2005). *A Research Report on the Education of South Asian Ethnic Minority Groups in Hong Kong*. Hong Kong: The Hong Kong Polytechnic University.

Lesaux, N., Koda, K., Siegel, L., and Shanahan, T. (2006). "Development of literacy," in *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*, eds D. August and T. Shanahan (London: Routledge), 75–122.

Li, D. C., and Chuk, J. Y. (2015). South Asian students' needs for Cantonese and written Chinese in Hong Kong: A linguistic study. *Int. J. Multiling.* 12, 210–224. doi: 10.1080/14790718.2015.1009375

Limbos, M. M., and Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *J. Learn. Disabil.* 34, 136–151. doi: 10.1177/002221940103400204

Lin, D., Sun, H., and McBride, C. (2018). Morphological awareness predicts the growth rate of Chinese character reading. *Dev. Sci.* 22:e12793. doi: 10.1111/desc.12793

Lipka, O., and Siegel, L. S. (2007). The development of reading skills in children with English as a second language. *Sci. Stud. Read.* 11, 105–131. doi: 10.1080/10888430709336555

Lisenby, B. E. (2011). Early childhood education for ethnic minorities in Hong Kong: Parent experience and policy support. *Asia Pac. J. Multidiscip. Res.* 5, 69–90.

Loper, K. (2004). *Race and Equality: A Study of Ethnic Minorities in Hong Kong's Education System*. Hong Kong: The University of Hong Kong.

Lord, F. M. (1975). Formula scoring and number-right scoring. *J. Educ. Meas.* 12, 7–11. doi: 10.1111/j.1745-3984.1975.tb01003.x

Ma, X., Gong, Y., Gao, X., and Xiang, Y. (2017). The teaching of Chinese as a second or foreign language: A systematic review of the literature 2005-2015. *J. Multiling Multicult. Dev.* 38, 815–830. doi: 10.1080/01434632.2016.1268146

Mattock, K., and Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy* 10, 241–265. doi: 10.1207/s15327078in1003_3

McBride, C. (2016). *Children's Literacy Development: A Cross-Cultural Perspective on Learning to Read and Write (Second Edition)*. New York, NY: Routledge.

McBride, C., and Wang, Y. (2015). Learning to read Chinese: Universal and unique cognitive cores. *Child Dev Perspect* 9, 196–200. doi: 10.1111/cdep.12132

McBride-Chang, C., and Ho, C. S.-H. (2009). Predictors of beginning reading in Chinese and English: A 2-year longitudinal study of Chinese kindergarteners. *Sci. Stud. Read.* 9, 117–144. doi: 10.1207/s1532799xssr0902_2

McCoy, E., and Cole, J. (2011). *A Snapshot of Local Support for Literacy: 2010 Survey*. London: National Literacy Trust.

McKendry, M. G., and Murphy, V. A. (2011). A comparative study of listening comprehension measures in English as an additional language and native English-speaking primary school children. *Int. J. Eval. Res.* 24, 17–40. doi: 10.1080/09500790.2010.531702

Perfetti, C., Cao, F., and Booth, J. (2013). Specialization and universals in the development of reading skill: How Chinese reading informs a universal science of reading. *Sci. Stud. Read.* 17, 5–21. doi: 10.1080/10888438.2012.689786

Perfetti, C. A., Liu, Y., and Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychol. Rev.* 112, 43–59. doi: 10.1037/0033-295x.112.1.43

Reardon, S. F., and Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *Am. Educ. Res. J.* 46, 853–891. doi: 10.3102/0002831209333184

Research Office of the Hong Kong Legislative Council Secretariat (2017). *Fact Sheet: Poverty of Ethnic Minorities in Hong Kong*. Availble online at: https://www.legco.gov.hk/research-publications/english/1617fs08-poverty-of-ethnic-minorities-in-hong-kong-20170608-e.pdf (accessed January 6, 2023).

Share, D. L. (2021). Is the science of reading just the science of reading english? *Read. Res. Q.* 56, S391–S402. doi: 10.1002/rrq.401

Shum, M. S. K., Gao, F., Tsung, L., and Ki, W. (2011). South Asian students' Chinese language learning in Hong Kong: Motivations and strategies. *J. Multiling Multicult. Dev.* 32, 285–297. doi: 10.1080/01434632.2010.539693

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Read. Res. Q.* 21, 360–407.

Strand, S., Malmberg, L., and Hall, J. (2015). *English as an Additional Language (EAL) and Educational Achievement in England: An Analysis of the National Pupil Database*. Available online at: https://ore.exeter.ac.uk/repository/bitstream/handle/10871/23323/EAL_and_educational_achievement2.pdf?sequence=1&isAllowed=y (accessed January 6, 2023).

Tobai, V., and Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Read. Writ.* 28, 939–957. doi: 10.1007/s11145-015-9556-1

Tong, X., McBride-Chang, C., Shu, H., and Wong, A. M.-Y. (2009). Morphological awareness, orthographic knowledge, and spelling errors: Keys to understanding early Chinese literacy acquisition. *Sci. Stud. Read.* 13, 426–452. doi: 10.1080/10888430903162910

Tse, S. K., Pang, E. Y. W., To, H., Tsui, P. F., and Lam, L. S. (2020). The impact of medium of instruction on Hong Kong non-Chinese speaking kindergarten children's learning Chinese. *Int. J. Early Years Educ.* 30, 1–14. doi: 10.1080/09669760.2020.1863191

Tse, S. K., and Zhu, X. H. (2001). "Shizi jiaoyuxue de xinli [Psychology of the teaching and learning of Chinese characters]," in *Gao xiao han zi jiao yu xue [Highly efficient teaching and learning of characters]*, ed. S. K. Tse (Hong Kong: Greenfield Educational Center), 44–96.

Tsung, L., Zhang, Q., and Cruickshank, K. (2010). Access to majority language and educational outcomes: South Asian background students in postcolonial Hong Kong. *Diaspora Indig. Minor. Educ.* 4, 17–32. doi: 10.1080/15595690903442256

Verhoeven, L. (2000). Components in early second language reading and spelling. *Sci. Stud. Read.* 4, 313–330. doi: 10.1207/S1532799XSSR0404_4

Wang, Y., McBride, C., Zhou, Y., Joshi, R. M., and Farver, J. A. M. (2018). Becoming literate in Chinese: A comparison of native-speaking and non-native-speaking children. *J. Res. Read.* 41, 511–524. doi: 10.1111/1467-9817.12122

Wong, Y. K. (2017). The role of radical awareness in Chinese-as-a-second-language learners' Chinese character reading development. *Lang. Aware.* 26, 211–225. doi: 10.1080/09658416.2017.1400039

Wong, Y. K. (2018). Developmental relations between listening and reading comprehension in young Chinese language learners: A longitudinal study. *J. Psycholinguist. Res.* 50, 1–13. doi: 10.1007/s10936-018-9619-y

Wong, Y. K. (2019). Role of decoding competence in the Chinese reading comprehension development of ethnic minority students in Hong Kong. *Int. J. Biling. Educ. Billing.* 22, 1016–1029. doi: 10.1080/13670050.2017.1329273

Wong, Y. K., and Zhou, Y. (2022). Effects of metalinguistic awareness on Chinese as a second language spelling through the mediation of reading and copying. *Read. Writ.* 35, 853–875. doi: 10.1007/s11145-021-10167-0

Yang, X., Peng, P., and Meng, X. (2019). How do metalinguistic awareness, working memory, reasoning, and inhibition contribute to Chinese character reading of kindergarten children? *Infant Child Dev.* 28:e2122. doi: 10.1002/icd.2122

Yesil-Dagli, U. (2011). Predicting ELL students' beginning first grade English oral reading fluency from initial kindergarten vocabulary, letter naming, and phonological awareness skills. *Early Child Res. Q.* 26, 15–29. doi: 10.1016/j.ecresq.2010.06.001

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 25 July 2023
DOI 10.3389/fpsyg.2023.1145493

Check for updates

OPEN ACCESS

EDITED BY
Yang Zhang,
University of Minnesota, Twin Cities,
United States

REVIEWED BY
Vittorio Tantucci,
Lancaster University, United Kingdom
Jing Yang,
Zhejiang University, China

*CORRESPONDENCE
Yang Zhao
✉ zhaoyang@pku.edu.cn

RECEIVED 16 January 2023
ACCEPTED 05 July 2023
PUBLISHED 25 July 2023

CITATION
Yan S, Mai Z and Zhao Y (2023) Positive cross-linguistic influence in the representation and processing of sentence-final particle *le* by L2 and heritage learners of Chinese.
Front. Psychol. 14:1145493.
doi: 10.3389/fpsyg.2023.1145493

Positive cross-linguistic influence in the representation and processing of sentence-final particle *le* by L2 and heritage learners of Chinese

Shanshan Yan[1], Ziyin Mai[2] and Yang Zhao[1]*

[1]School of Chinese as a Second Language, Peking University, Beijing, China, [2]Department of Linguistics and Translation, City University of Hong Kong, Hong Kong, China

This study investigates the representation and processing of written Chinese sentences subject to a semantic condition (i.e., "direction of change") attached to the sentence-final particle (SFP) *le* in Mandarin Chinese. Three groups of bilingual speakers of Chinese and English who differ in their onset age of bilingualism and proficiency of English were studied. It was anticipated that there would be a positive cross-linguistic influence (CLI) from English due to similarities between the SFP *le* and the English adverb *already* in terms of direct semantic transfer. An acceptability judgment (AJ) task and a self-paced reading (SPR) task were conducted to elicit judgment and processing difficulty with and without semantic violations. The participants included English-dominant second language (L2) learners ($n = 18$) and heritage learners ($n = 19$) who had advanced proficiency in Chinese, as well as monolingually raised Mandarin speakers from China as a baseline control group ($n = 18$). The results indicated that sensitivity to violations of the semantic condition varied depending on factors such as the specific structure (noun vs. verb phrase), the task type (offline vs. online), and the type of bilingual speaker (early vs. late). Among the three groups of bilinguals, the heritage learners demonstrated a representation of the semantic condition that resembled the target language across different sentence structures, whereas the L2 learners did not. Furthermore, the heritage learners exhibited earlier sensitivity to violations during online processing compared to the baseline control group. These exceptional results can be attributed to the heritage learners' early exposure to and positive CLI between the SFP *le* in Mandarin and the English adverb *already*.

KEYWORDS

positive cross-linguistic influence, bilingual co-activation, SFP *le*, L2 learners, heritage learners

1. Introduction

Cross-linguistic influence (CLI), in which grammatical properties of prior or simultaneously acquired language(s) influence those of subsequently acquired ones, is widely reported (see Lago et al., 2021 for a review). Moreover, first language (L1) properties, if transferred into the second language (L2), may persist into very advanced stages of L2 acquisition (Inagaki, 2001; Rankin, 2014; Bauke, 2020; among others). In sentence reading tasks which necessarily involve real-time information processing, coordination of grammatical properties from various linguistic domains such as syntax and semantics can be cognitively taxing (Hopp, 2006). Most studies on CLI have focused on the negative effects of CLI (Odlin, 2001; Yuan, 2004; Yuan and Zhao, 2005; Jackson and Dussias, 2009; among others). However, much less emphasis has been placed on its positive side, which should facilitate both the acquisition and processing of the target language (Odlin, 1989; Yip et al., 2018). Furthermore, for structures that pose similar constraints cross-linguistically, the syntactic differences between them, for example whether less embedded ones exert larger positive CLI effects, are unknown. For L2 learners (L2ers), online reading tasks are more challenging than offline ones as explicit knowledge is less accessible by the parser in online tasks due to time pressure (e.g., Jiang, 2004; Andersson et al., 2019; Mickan and Lemhöfer, 2020). However, for heritage language learners (HLers), early exposure to the target language might to some extent bring about advantages over adult L2 learners in terms of sensitivity to very subtle semantic features (e.g., Mai and Deng, 2017; Polinsky and Scontras, 2020). In this study, we examine whether and to what extent positive CLI takes place in establishing and maintaining a syntax-semantic condition in the L2 and heritage Mandarin Chinese using both offline and online reading tasks.

2. CLI in grammatical acquisition and sentence reading

2.1. Structural effects on CLI

When the target structure in the L2 does not have a close structural equivalent in the L1, or the L1 equivalent is substantially different from the L2 target **("L1-L2 different structures")**, positive CLI from the L1 on the L2 is not expected. Nevertheless, L1-L2 similarity is by no means a necessary condition for native-like sentence processing in the L2. Frenck-Mestre and Pynte (1997) showed that in processing structurally ambiguous English sentences (i.e., subject/object ambiguities, non-ambiguous in French), adult French-L1 English-L2 bilinguals performed in a similar manner to English monolinguals despite temporary difficulties. In Hoover and Dwivedi (1998), French clitics in causative and non-causative sentences were successfully acquired by highly proficient adult English-L1 learners of French. Similarly, Jackson and Dussias (2009) revealed that L2 learners of German, when reading *wh*-extractions in German, were sensitive to morphological case-marking not instantiated in their L1 English (other studies see Hasegawa et al., 2002; Yokoyama et al., 2006).

Nevertheless, when the L2 target structure has a close equivalent in the L1 **("L1-L2 similar structures")**, it is expected that structural similarity should lead to positive L1 transfer, facilitating the acquisition of the L2. However, the expected positive CLI is not always present (Bates and MacWhinney, 1989; Tokowicz and MacWhinney, 2005; Jing, 2008; Sabourin and Stowe, 2008, among others). For example, native speakers of Greek showed clear preferences in terms of high/low attachment in interpreting nouns in relative clauses, which is shared by a number of languages such as Spanish, German, and Russian, and are thus not expected to cause problems for Greek L2 learners whose L1 is any of these three languages. However, Papadopoulou and Clahsen (2003) found that L2 learners of Greek who were highly proficient in it, showed no preference for high/low attachment in the structure, suggesting the L1-L2 similarity might not play any facilitating role. Similar findings were reported with regard to the comprehension of non-local dependencies (Felser and Roberts, 2007; Omaki and Schulz, 2011).

Jeong et al. (2007) and Jeong et al. (2007) studied the processing of word order and related syntactic features (e.g., case-marking) by native speakers and L2 learners of SVO languages (Mandarin Chinese, English) and SOV languages (Korean, Japanese). Those learning an L2 with a word order similar to their L1 showed advantages in brain activities in processing the L2 compared with those learning an L2 with a word order different from their L1, but the advantage seems to be specific to the task adopted (i.e., a behavioral auditory sentence comprehension task), indicating task effects on the extent of positive CLI, which we will elaborate on in the next section.

2.2. Task effects on CLI

CLI, whether negative or positive, is extensively investigated in both **online** processing and **offline** comprehension and judgment tasks, and the findings are mixed. Jiang (2004) investigated subject-verb agreement marking in English by Chinese-L1 learners of L2 English. The L2 learners were quite accurate in an offline written test, yet insensitive to the same morphosyntactic manipulation in an online comprehension task. Elsewhere, Hopp (2006) showed that lower-proficient advanced German L2 speakers (whose L1s were English and Dutch) were able to utilize case-marking information as evidenced by comprehension accuracies, but were insensitive to the same information in online processing. However, in Turkish-L1 and German-L1 learners of L2 Dutch, CLI in pronoun resolution is observed in offline interpretation tasks yet absent in online reading tasks (Roberts et al., 2008). In Hopp and Grüter (2023), L1 effects were identified among German-L1 and Japanese-L1 L2 learners of English in offline comprehension of *wh*-questions rather than in online processing of the same questions (for more evidence on the online vs. offline asymmetry, see Kaan et al., 2015; Andersson et al., 2019; Mickan and Lemhöfer, 2020).

In bilingual processing, abundant evidence has supported **co-activation** of both languages in various types of bilinguals, even when the linguistic stimuli are presented monolingually (e.g., De Groot, 2011; Hopp and Grüter, 2023). Most studies have examined cross-linguistic co-activation at the **lexical** level, with a special focus on L1-L2 equivalents that are cognate words (Costa et al., 2000; Dijkstra, 2005). Processing these words in the L2 activates their L1 equivalents and facilitates lexical processing in the L2 (Dekydtspotter et al., 2006; Miller, 2014). However, less is known about the activation of **syntactic** structures in bilingual processing, except Hopp (2017) and Vaughan-Evans et al. (2020). Hopp (2017) touched upon syntactic co-activation in German-L1 learners of English through lexical co-activation in reading. Specifically, he investigated whether English relative clauses with different word orders that overlap with German-L1-embedded clauses activates the L1 grammars in online L2 sentence comprehension. His findings suggest the L1 word order was co-activated. Vaughan-Evans et al. (2020) examined Welsh-L1 English-L2 early bilinguals' online reading of English sentences with manipulated morphosyntactic rules (i.e., the Welsh soft mutation rule) and proposed that co-activation of the L1 syntax may occur through syntactic rules. As for the processing of complex syntactic features, such as syntax-semantic relationships investigated in the current study, much work is still needed.

2.3. Learner effects on CLI

It has been shown that bilingual co-activation may boost the processing of relevant syntactic units. However, this positive effect is not found across types of bilinguals. For instance, Canseco-Gonzalez

10.3389/fpsyg.2023.1145493

et al. (2010) compared three groups of Spanish-English bilinguals with various onset ages of bilingualism and proficiency in Spanish: the early bilinguals (with an onset age of both languages before 6 years) and the English-L2 group (onset age after 6 years) co-activated Spanish in word recognition tasks in English, whereas the Spanish-L2 group (onset age after 6 years) did not reveal any co-activation effect. The authors argued that the more proficient the early bilinguals are, the more successful and accustomed the learners are in terms of co-activation. This leads to an interesting prediction: early bilinguals of Languages A and B may be more efficient than late bilinguals (e.g., L2 learners of Language A) in co-activating Language B in reading in Language A. Moreover, early bilinguals may even outperform monolingually raised speakers of Language A, who are late bilinguals of the same languages. Our study follows up on this line of research and investigates learner effects in three types of proficient bilinguals in Chinese and English.

Heritage language speakers (HLers) are typically early bilinguals who acquire the HL as their L1 (or one of their L1s alongside the societal majority language) through interaction with parents who are native speakers of the language in naturalistic settings (Polinsky and Scontras, 2020). Because an HL is by definition a minority language in the larger community, adult HL speakers usually (but not always) develop stronger abilities in the societal majority language through mainstream schooling and possess varying proficiency in the HL (Montrul, 2016). Some HL speakers are able to develop literacy in the HL through bilingual education programs, community schools, or language courses. **L2 learners (L2ers)** are typically late bilinguals who acquire their L1 at early childhood and are raised in the L1 context. Their L2 was mostly learned through formal instruction after puberty. HLers, on the one hand, can display substantial differences from L2ers in many acquisition and processing aspects across linguistic domains (Montrul, 2016); on the other hand, they are also different from monolingually raised speakers of that language ("baseline") in judging and processing complex grammatical properties of the HL (e.g., Mai and Deng, 2017; Polinsky and Scontras, 2020). For HLers, it is likely that the experience of developing both the HL and the majority language concurrently in childhood gives rise to more accurate and accessible mapping between corresponding structures in the two languages and bring advantages over L2 and monolingual baselines in L1-L2 similar structures. In this sense, it might be possible that the CLI is modulated by learner effect, and we further explore this in the current study. In comparing English-dominant L2 and HL Spanish speakers, Regulez and Montrul (2023) found great learner variations regarding acceptability, production, and online comprehension in Spanish differential object marking, reporting the presence of task effects. This study tests L1-L2 similar structures involving complex syntax-semantics coordination, the sentence-final particle (SFP) *le* in Mandarin Chinese and the adverb *already* in English through both offline and online tasks.

3. The SFP *le* in Mandarin Chinese and the adverb *already* in English

3.1. The SFP *le*

Mandarin Chinese has a rich set of SFPs, among which *le* denotes a newly obtained state or "change of state" (Chao, 1968; Li and Thompson, 1981; Paul, 2015; Paul and Yan, 2022) and entails that the current state does not hold before speech time (Soh, 2009).[1] When the change is associated with a naturally occurring or commonly perceived ordered **scale** (e.g., numbers in ascending order, or stages of human development), the SFP *le* imposes, among other conditions, a semantic constraint on the direction of change encoded, as outlined in (1):

(1) The change must take place from an earlier/lower point to a later/higher point along the associated scale, rather than vice versa (Xiang, 1998; Xing, 2001).

For example, in stative [*be*-NP] structures as presented in (2) and (3), when the NP identifies a point which is straightforwardly interpretable as a high point on the ordered scale, entailing at least one lower point on the scale (termed as **"higher NP,"** e.g., where *dà-háizi* ('big child') is a point higher than *xiǎo-háizi* ('little child') on the scale of human development), [*be*-NP] is naturally compatible with the SFP *le*, as shown in (4). However, when the NP denotes a low point on the scale without the entailment of an even lower point (termed as **"lower NP,"** e.g., *xiǎo-háizi* ('little child')), [*be*-NP] is less compatible with the SFP *le*, as in (5). Notice that both types of NP are natural without the SFP *le*, as illustrated in (2) and (3).

(2) Higher NP in [*be*-NP]
Xiǎowáng shì dà háizi.
Xiaowang COP big child
'Xiaowang is a big child.'

(3) Lower NP in [*be*-NP]
Xiǎowáng shì xiǎo háizi.
Xiaowang COP small child
'Xiaowang is a small child.'

(4) Higher NP in [*be*-NP-*le*$_{\text{SFP}}$]
Xiǎowáng shì **dà háizi le**.
Xiaowang be big child LE
'Xiaowang is a big child already.'

(5) */?Lower NP in [*be*-NP-*le*$_{\text{SFP}}$]
*/?Xiǎowáng shì **xiǎo háizi le**.
Xiaowang be small child LE
*/?'Xiaowang is a small child already.'

Another manifestation of (1) can be found regarding the SFP *le* with **upward/downward-entailing duration phrases in [V-*le*$_{\text{ASP}}$-duration-*le*$_{\text{SFP}}$] structures**. Upward-entailing contexts identify, on a scale, the lowest possible value that is true for the relevant proposition and entail values higher than this point [e.g., *zhìshǎo wǔnián* ('at least 5 years'), or *chāoguò sāntiān* ('exceeding 3 days')]; they are natural with

1 Mandarin has a perfective aspect marker *le,* which is homomorphous to the SFP *le* (Huang et al., 2009; Yan, 2023; among others). Aspect *le* follows a verb and appears before the object of the verb, if any (e.g., *Wǒ chī le píngguǒ* 'I have eaten an apple'), whereas the SFP *le* occurs after the object at the clause-final position. Aspect *le* and the SFP *le* can co-occur and appear in [V-*le*$_{\text{ASP}}$-O-*le*$_{\text{SFP}}$]. In cases where the object is grammatically omitted, the distinction is less straightforward. Therefore, in our experimental stimuli to be introduced below, the SFP *le* always follows an overt object of the verb [e.g., *Wǒ chī píngguǒ le* 'I have eaten apples, (which is not the case before)'].

the SFP *le*, for the direction of the entailment aligns with the direction of change as stipulated by the SFP *le* above, as illustrated in (8). In sharp contrast, **downward-entailing contexts** identify the highest point that holds for the relevant proposition [e.g., *búdào wǔnián* ('less than 5 years'), or *zuìduō sāntiān* ('at most 3 days')] and are thus not compatible with the SFP *le*, as depicted in (9). Notice here that both upward and downward elements are possible in sentences without the SFP *le*, as demonstrated in (6) and (7).[2]

(6) Upward-entailing duration in [V-le_{ASP}-duration]
Wǒ zài zhèr [$_{VP}$ zhù le **zhìshǎo** wǔ nián].
1SG at here live PERF at-least five CL
'I lived here for at least 5 years.'

(7) Downward-entailing duration in [V-le_{ASP}-duration]
Wǒ zài zhèr [$_{VP}$ zhù le **búdào** wǔ nián].
1SG at here live PERF less-than five CL
'I lived here for less than 5 years.'

(8) Upward-entailing duration in [V-le_{ASP}-duration-le_{SFP}]
Wǒ zài zhèr [$_{VP}$ zhù le **zhìshǎo** wǔ nián] **le**.
1SG at here live PERF at-least five CL LE
'I have lived here for at least 5 years already (which was not the case before).'

(9) */?Downward-entailing in [V-le_{ASP}-duration-le_{SFP}]
*/?Wǒ zài zhèr [$_{VP}$ zhù le **búdào** wǔ nián] **le**.
1SG at here live PERF less-than five CL LE
*/?'I have lived here for less than 5 years already.'
[slightly modified examples from Soh (2009, p. 638–641)].

3.2. The adverb *Already* in English

Lacking SFPs, English has an adverb *already*, which is also subject to a "change of state" interpretation and is considered a semantic counterpart of the SFP *le* in Mandarin in many contexts (Traugott and Waterhouse, 1969; Soh, 2009). It has been argued that the English adverb *already* entails a "cognitive modeling of event sequence" including canonical courses of development (Michaelis, 1992, 1996). As shown in the translations in (4), (5), (8), and (9), the acceptability of *already* in English is subject to the same restrictions as the SFP *le* in contexts with scalar inferences: upward-entailing contexts (e.g., *at least*) and higher NPs (e.g., *big child*) are clearly more acceptable than downward-entailing contexts and lower NPs (e.g., *less than,* or *small child*).

In the absence of literature on cross-linguistic mapping between the SFP *le* and the English adverb *already* in Mandarin-English bilinguals, we searched the literature for evidence in Chinese-English bilinguals in general. Szeto et al. (2017) found that Cantonese-English bilinguals associated the English adverb *already* with the Cantonese SFP *laa3* and aspect marker *zo2*, which are well-established counterparts of the SFP *le* and aspect marker *le* in Mandarin.

4. This study

4.1. Research questions and predictions

As reviewed above, L1-L2 similarity facilitates both acquisition and processing due to positive CLI and bilingual co-activation, but whether positive CLI happens in different types of bilingual speakers and in complex grammatical processing involving semantic computation is an empirical question that has been raised only recently in studies. To fill this gap, this study looked into the acquisition and processing of higher/lower NPs in [*be*-NP-le_{SFP}] and upward/downward-entailing duration phrases in the [V-le_{ASP}-duration-le_{SFP}]. The following two research questions guided this study:

1. Given cross-linguistic similarities between the SFP *le* in Mandarin and the adverb *already* in English, will English-L1 learners of L2 Mandarin **(L2ers)** succeed in differentiating, in the SFP *le* sentences, the acceptability between higher and lower NPs such as in (4) and (5) and between upward- and downward-entailing duration phrases such as in (8) and (9) in both offline judgment and online reading?
2. Will adult **HLers**, who have acquired both Chinese and English at a young age, outperform proficiency-matched L2ers and pattern with monolingually raised Chinese L1 speakers in China (**baseline controls**), due to the early onset of Chinese-English bilingualism?

For English-L1 L2ers, consistent differentiation between the more acceptable and the less acceptable sentences hinges upon the mental representation of the semantic condition regulating the use of the SFP *le* in (1), which may be formed based on absorbing patterns in the input and profiting from positive transfer from *already*, if the mapping between the SFP *le* and *already* is established. We predict that L2ers will be able to make the differentiation, but they may be more successful in recognizing violations caused by lower NPs in [*be*-NP-le_{SFP}] than downward-entailing duration phrases in [V-le_{ASP}-duration-le_{SFP}], because the NPs in the former structure are less embedded, linearly closer to the SFP *le*, and semantically more concrete than the duration phrases in the latter. For HLers, we predict that they will outperform their L2 peers due to the early onset of bilingualism and likely early mapping between the SFP *le* in Mandarin and the adverb *already* in English and subsequently stronger facilitative effects. To test these predictions, we conducted an experiment consisting of an offline acceptability judgment task and an online self-paced reading task.

2 As suggested by a reviewer, we consulted the Center for Chinese Linguistics corpus at Peking University to evaluate the semantic restrictions in (1) in native naturalistic speech samples. We used the upward/downward-entailing elements (*zhìshǎo*, *búdào*) and higher/lower NPs (*dà-NP*, *xiǎo-NP*) as target words and identified 99 tokens *zhìshǎo*+SFP *le* (out of 2,487 tokens including the target words) and 114 tokens of *dà*~+ SFP *le* (out of 1,135 tokens including the target words). By contrast, no downward-entailing **búdào*+SFP *le* was found. We also found 49 cases of lower NPs *xiǎo*~+ SFP *le*. Nevertheless, the lower NPs were all within the scope of negators (e.g., *búshì xiǎo háizi*, 'NEG-be a small child'), which identifies a higher point on the scale and is thus consistent with the generalization in (1).

TABLE 1 Participant information (SD in brackets).

Groups	N	Age	Onset age	Months studying Chinese	Months in China	Cloze test
L2	18	23 (2.2)	18 (4.6)	60 (37.8)	38.1 (47.4)	31.3 (1.8)
HL	19	22 (2.9)	0	187 (102.3)	23.2 (28.9)	33.2 (2.4)
Baseline	18	23 (4.5)	0	N/A	N/A	38.8 (1.1)

4.2. Participants

The data of this study are from a larger project investigating the acquisition and processing of several grammatical structures by adult L2 and HL learners of Mandarin. The participants were recruited from Beijing, Shanghai, and Hangzhou in China, and Cambridge in the United Kingdom. A language background questionnaire was administered to collect information about each participant's language experience. The original sample consists of data from 150 participants. Given the research goals of this study and the complexity of the grammatical construction under investigation, we selected from the original sample participants who self-reported having intermediate or above-intermediate proficiency in written Chinese. Other criteria included: to qualify as L2ers, the participant must have reported English as their native language, and started to learn Chinese after puberty (average onset age 18 years, see Table 1); and to qualify as HLers, the participants must have reported early and substantial exposure to both Chinese and English in an English-speaking country and self-rated English as their dominant language at time of data collection. Ultimately, 18 L2ers (6 females) and 19 HLers (13 females) satisfied the above criteria and were included in this study.

The HLers had been exposed to Mandarin Chinese ($n = 15$) or Cantonese and Min Chinese[3] ($n = 4$) through everyday interactions with family members from birth, with at least one parent being a native speaker of Chinese. They were either born and raised in an English-dominant country (UK, US, etc., $n = 14$) or had immigrated with their family to an English-dominant country before the age of 5 (see Supplementary material for each HLer's language exposure). The baseline controls ($n = 18$) were undergraduate or graduate students at universities in Beijing (12 females), majoring in liberal arts ($n = 10$) or science ($n = 8$). They were all born and raised in northern China; none had visited an English-speaking country by the time of data collection.

An established cloze test (Zhao, 2006; Yuan, 2010; Mai and Yuan, 2016; etc.) was adopted to gauge the learners' reading proficiency in Chinese. The test included two short passages with 40 blanks and participants were required to fill in each blank with a Chinese character. No Pinyin (romanized form of Chinese) was provided. ANOVA and *post hoc* Tukey tests indicated that both the L2ers and HLers were significantly different from the baseline controls [$F(2, 52) = 78.24$, all $p < 0.000$], but they were not significantly different from each other ($p = 0.921$), suggesting comparable proficiency between the L2ers and the HLers. Information on the participants ($n = 55$) is summarized in Table 1.

4.3. Tasks

4.3.1. Acceptability judgment (AJ) task

In the AJ task, the participants judged the acceptability of four types of sentences on a four-point Likert scale (ranging from 'completely unacceptable,' to 'probably unacceptable,' to 'probably acceptable,' and to 'completely acceptable'):

Higher and lower NPs in sentences with and without the SFP *le*:

Type 1 ($n = 4$): [*be*-NP$_{HIGHER}$-*le*$_{SFP}$], as in (4),

Type 2 ($n = 4$): */?[*be*-NP$_{LOWER}$-*le*$_{SFP}$], as in (5),

Type 3 ($n = 4$): [*be*-NP$_{HIGHER}$], as in (2), and

Type 4 ($n = 4$): [*be*-NP$_{LOWER}$], as in (3).

Upward/downward-entailing duration phrases with and without the SFP *le*:

Type 5 ($n = 4$): [V-*le*$_{ASP}$-Duration$_{UP}$-*le*$_{SFP}$], as in (8),

Type 6 ($n = 4$): */?[V-*le*$_{ASP}$-Duration$_{DOWN}$-*le*$_{SFP}$], as in (9),

Type 7 ($n = 4$): [V-*le*$_{ASP}$-Duration$_{UP}$], as (6), and

Type 8 ($n = 4$): [V-*le*$_{ASP}$-Duration$_{DOWN}$], as in (7).

Four tokens were constructed for each type, rendering a total of 32 tokens. These critical items were mixed with 128 other items testing other grammatical structures such as *wh*-questions in the larger study. The sentences were randomized and presented one-by-one on a computer screen. The participants pressed one of five designated keys on the keyboard to indicate their rating on the scale (the four-point scale plus a separate "I do not know" option), and the program was designed such that once the testing has started, the participant could see only one sentence on the screen and they could not return to a pervious item or change their answers. Six additional training items were provided before the critical items.

4.3.2. Self-paced reading (SPR) task

In this "read and judge" task, the participants read a critical sentence with the SFP *le* and provided a binary truth-value judgment ("correct" or "incorrect") on a follow-up comprehension sentence based on their understanding of the previous SFP *le*-sentence. Unlike previous studies in which participants were required to read the sentences as quickly as possible, in this task we asked them to read the sentences at a natural and comfortable pace. This is because participants in pilot studies reported considerable anxiety and fatigue when they had to read the sentences as quickly as possible. The test items were presented segment by segment on a computer monitor. Thereafter, participants had to press a key to see the next segment of the SFP *le*-sentence. Upon reaching the end of the SFP *le*-sentence, the comprehension sentence was presented in full for the participants' judgment.

Four conditions, as shown in (10), (11), (12), and (13), were created to test the two structures, with six or eight trails in each

3 Our own informants, who were monolingual speakers of Min from Zhangzhou, Xiamen, and Taiwan, identified *liau* in Min Chinese as the equivalent of the Mandarin SFP *le*.

condition. The SFP *le*-sentences in the SPR task were longer than those in the AJ task, with an additional clause following the SFP *le*-clause to allow for spill-over effects and potentially delayed processing of the critical element (the SFP *le*). The sentences were segmented into seven windows for [*be*-NP-le_{SFP}] and nine for [V-le_{ASP}-duration-le_{SFP}] so that only one word (mostly disyllabic, bimorphemic with only a few systematic exceptional cases) was presented to the reader in a window. Note that in [*be*-NP-le_{SFP}] conditions, the monosyllabic SFP *le* [the critical window, W4 in (10) and (11)] was presented separately from the other words; in [V-le_{ASP}-duration-le_{SFP}] conditions, the upward/downward quantifiers were presented in a single time window without the temporal nouns, which were presented together with the SFP *le* in the subsequent window [the critical window, W5 in (12) and (13)]. Half of the comprehension sentences were directly related to the SFP *le*-clause and the other half to the additional clause. In addition, half of them were expected to elicit "correct" responses and the other half "incorrect" ones.

(10) [*be*-NP_{HIGHER}-le_{SFP}] ($n = 6$)

		pre-critical	critical	1st post-critical	2nd post-critical	
W1	W2	W3	W4	W5	W6	W7
Xiǎomíng	/shì	/dànánhái	/le	/,yīnggāi	/hǎohǎo	/xuéxí.
Xiaoming	COP	big boy	LE	should	well	study

'Xiaoming is a big boy already; he should study hard.'

Comprehension sentence: Xiǎomíng bùyīnggāi hǎohǎo xuéxí ('Xiaoming should not study hard')

Expected response: Incorrect

(11) */?[*be*-NP_{LOWER}-le_{SFP}] ($n = 6$)

		pre-critical	critical	1st post-critical	2nd post-critical	
W1	W2	W3	W4	W5	W6	W7
Xiǎomíng	/shì	/xiǎonánhái	/le	/,yīnggāi	/hǎohǎo	/xuéxí.
Xiaoming	COP	small boy	LE	should	well	study

Intended: 'Xiaoming is a small boy, and he should study hard.'

Comprehension sentence: Xiǎomíng búshì xiǎonánhái ('Xiaoming is not a small boy')

Expected response: Incorrect

The 28 SPR items were randomized and mixed with 44 items testing other structures such as aspect markers, negation markers in the larger study using the Latin square design (Dean and Voss, 1999). Two lists (Lists A and B) were generated manually to ensure that the SPF *le*-sentences that were minimally different were not included in the same list. Half of the participants completed List A and the other half List B. In both cases, the test items were divided into two blocks evenly, with an obligatory break between them to reduce the impact of fatigue. Six training items were provided to ensure that the participants understood the instructions.

4.4. Procedures

The participants completed both tasks individually in a quiet classroom in one meeting with the first author. Test materials were presented using E-Prime 2.0. Meanwhile, test instructions were given to L2ers and HLers in English, and to Chinese baseline controls in Chinese. An "I do not know" option was available across items throughout the test. The SPR task was administered before the AJ task so that the latter, which activates explicit and metalinguistic knowledge to a larger extent than SPR, did not impact on the SPR results.

5. Results

5.1. Data coding, trimming, and statistical tools

Judgments in the AJ task were transformed into numerical scores of 1, 2, 3, and 4 for "completely unacceptable," "probably unacceptable," "probably acceptable," and "completely acceptable" respectively. Reading times and responses to comprehension sentences in the SPR task were trimmed and transformed following several steps: (i) responses in the comprehension sentences were checked to screen out participants who did not reach 75% accuracy (following Jiang, 2004; Keating and Jegerski, 2015) – no participant was eliminated in this procedure; (ii) reading times were screened to identify outlying values falling outside an absolute cutoff (shorter than 100 ms or longer than 2000 ms for baseline controls, shorter than 200 ms or higher than 4,000 ms for learners) or a variable cutoff (i.e., three standard deviations above mean reading times of the relevant group), which were then replaced with the absolute cutoff value of the relevant group in statistical analysis (following Keating and Jegerski, 2015); and (iii) to reduce confounding inter-participant differences due to general reading strategies, raw reading times (after adjusting for outlying values as described above) were transformed into **residual reading times**, which reflect differences between the actual raw reading times and the predicted reading times based on the regression equation, with word length as the predictor and raw reading times as the response variable (following Keating and Jegerski, 2015; also known as "deviations from regressions"). After the transformation, positive values indicate longer reading times than predicted by the regression, whereas negative ones mean they are shorter than predicted.

Linear mixed-effects models (LMMs) were performed for both tasks in R (version 4.1.1, R Core Team, 2014; *lme4*, Bates et al., 2013). The ratings in the AJ task and reaction times in the SPR task were dependent variables in the respective models. Fixed factors included CONDITION (acceptable vs. less acceptable), GROUP (baseline, L2, and HL), and interactions between CONDITION and GROUP. Random factors included intercepts for subjects and items, as well as by-subject and by-item random slopes. Factors were removed if they did not significantly improve the model according to the maximum likelihood ratio tests. Due to convergence issues, the random slopes for items were removed. Values $t > |2|$ were considered statistically significant (following Gelman and Hill, 2007). The results are reported in the following sub-sections.

(12) [V-le_{ASP}-$\text{Duration}_{\text{UP}}$-$le_{\text{SFP}}$] ($n = 8$)

			pre-critical	critical	1st post-critical	2nd post-critical		
W1	W2	W3	W4	W5	W6	W7	W8	W9
Xiǎowáng	/zài zhèr	/zhù le	/zhìshǎo	/wǔnián le	/, tā	/fēicháng	/xǐhuān	/ zhèr.
Xiaowang	PREP here	live PERF	at least	5 years LE	3SG	very	like	here

'Xiaowang has lived here for at least 5 years, he likes it here very much.'

Comprehension sentence: Xiǎowáng xǐhuān zhèr ('Xiaowang likes it here')

Expected response: Correct

(13) */?[V-*le*$_{ASP}$-Duration$_{DOWN}$-*le*$_{SFP}$] ($n = 8$)

			pre-critical	critical	1st post-critical	2nd post-critical		
W1	W2	W3	W4	W5	W6	W7	W8	W9
*Xiǎowáng	/zài zhèr	/zhù le	/búdào	/wǔnián le	/, tā	/fēicháng	/xǐhuān	/ zhèr.
Xiaowang	PREP here	live PERF	less than	5 years LE	3SG	very	like	here

'Xiaowang has lived here for less than 5 years, he likes it here very much.'

Comprehension sentence: Xiǎowáng méi zài zhèr zhù ('Xiaowang does not live here')

Expected response: Incorrect

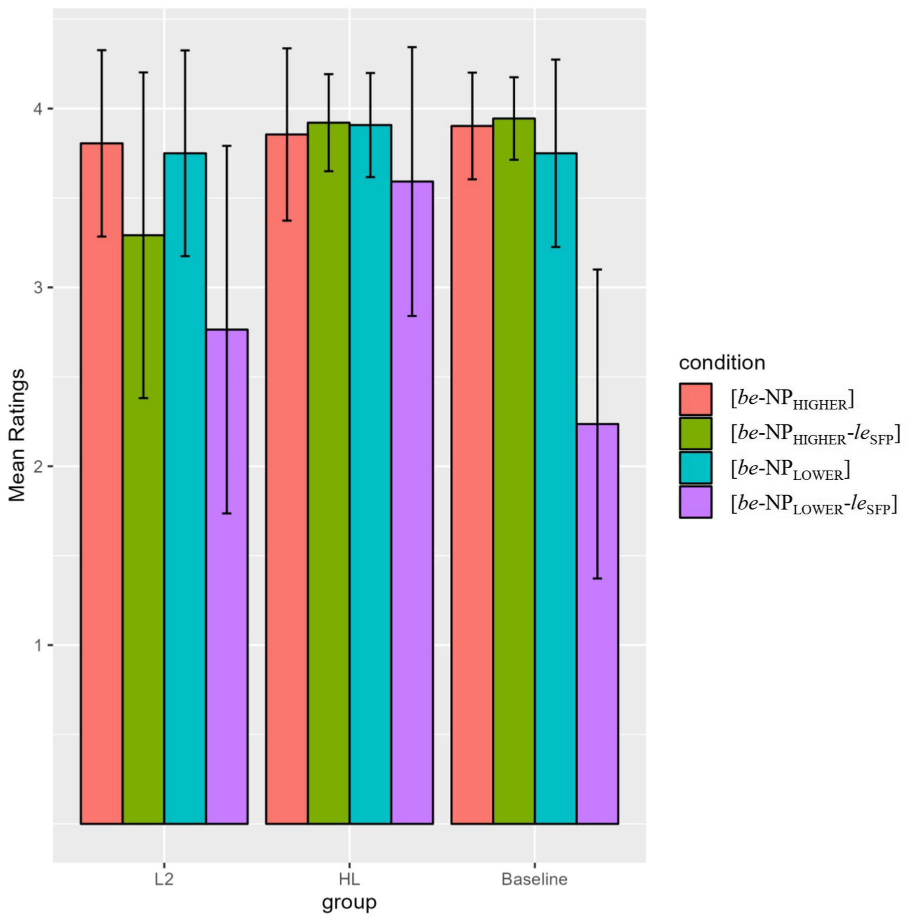


FIGURE 1
Mean ratings of the higher/lower NP conditions in the acceptability judgment (AJ) task.

TABLE 2 Mixed-effects models on [*be*-NP_{HIGHER}-le_{SFP}] and */?[be-NP_{LOWER}-le_{SFP}] in the acceptability judgment (AJ) task.

	Fixed effects		
Predictors	**Estimates**	***SE***	***t***
	L2 learners		
(Intercept)	2.76	0.12	22.91***
Condition [*be*-higher NP-le_{SFP}]	0.53	0.13	4.22***
Group Baseline	−0.53	0.16	−3.31**
Group HL	0.83	0.16	5.27***
Condition [*be*-higher NP-le_{SFP}]: Baseline	1.18	0.15	7.66***
Condition [*be*-higher NP-le_{SFP}]: HL	−0.20	0.15	−1.31
	HL learners		
(Intercept)	3.59	0.12	30.48***
Condition [*be*-higher NP-le_{SFP}]	0.33	0.12	2.69*
Group Baseline	−1.36	0.16	−8.63***
Group L2	−0.82	0.16	−5.27***
Condition [*be*-higher NP-le_{SFP}]: Baseline	1.38	0.15	9.07***
Condition [*be*-higher NP-le_{SFP}]: L2	0.20	0.15	1.31
	Baseline		
(Intercept)	2.24	0.12	18.53***
Condition [*be*-higher NP-le_{SFP}]	1.71	0.13	13.66***
Group HL	1.36	0.16	8.63***
Group L2	0.53	0.16	3.31**
Condition [*be*-higher NP-le_{SFP}]: HL	−1.38	0.15	−9.07***
Condition [*be*-higher NP-le_{SFP}]: L2	−1.18	0.15	−7.66***

Reference level for condition = */?[*be*-lower NP-le_{SFP}], Model formula for baseline group: ratings ~ condition + group + condition:group + (1 | subject) + (1 | item), Model formula for HL group: ratings ~ condition + group + condition:group + (1 | subject) + (1 | item), Model formula for L2 group: ratings ~ condition + group + condition:group + (1 | subject) + (1 | item); ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

5.2. AJ task

5.2.1. Higher/lower NP conditions

As shown in Figure 1; Table 2, all groups rated the [*be*-NP_{HIGHER}] and [*be*-NP_{LOWER}] sentences at ceiling (means above 3.5), indicating the [*be*-NP] structure is well represented in their Chinese grammars. All groups also rated the [*be*-NP_{HIGHER}-le_{SFP}] sentences higher than minimally different */?[*be*-NP_{LOWER}-le_{SFP}] sentences ($p < 0.001$ for baseline, *post hoc* power = 100%; and $p < 0.05$ for both HL and L2 groups, *post hoc* power for HL = 95.91%, for L2 = 100%).

5.2.2. Upward/downward-entailing duration conditions

As shown in Figure 2; Table 3, the baseline and HL groups were able to rate [V-le_{ASP}-$Duration_{UP}$-le_{SFP}] sentences higher than */?[V-le_{ASP}-$Duration_{DOWN}$-le_{SFP}] sentences ($p < 0.001$ for baseline, *post hoc* power = 100%; $p < 0.01$ for HLers, *post hoc* power = 96.14%). However, the L2 group showed no such sensitivity. In addition, all groups rated the [V-le_{ASP}-$Duration_{UP}$] and [V-le_{ASP}-$Duration_{DOWN}$] sentences high (means above 3), suggesting their acceptance of the basic upward/downward-entailing duration structures.

To sum up, the AJ task results indicate that like the baseline controls, both HL and L2 learner groups were sensitive to the incompatibility between lower NPs and the SFP *le*. However, only the HLers were able to show sensitivity to the incompatibility between downward-entailing duration phrases and the SFP *le*, patterning with the baseline controls.

5.3. SPR task

5.3.1. Higher/lower NP conditions

Figure 3 presents the reading times at each time window in each condition by each group in three panels. LMM analyses reveal that the three groups of participants showed a similar reading time pattern: no significant differences were found between the two conditions across groups from the pre-critical window (W3) to the first post-critical window (W5); and between-condition differences first showed up in the **second post-critical window** (W6, baseline controls: $p < 0.05$, *post hoc* power = 68.52%; HLers: $p < 0.1$, *post hoc* power = 52.63%; L2: $p < 0.1$, *post hoc* power = 42.59%). Table 4 presents the LMM results of the second post-critical window.

5.3.2. Upward/downward-entailing duration conditions

Figure 4 presents the residual reading times at each time window in each condition by each group in three panels. Unlike the NP conditions, the three groups displayed three different patterns. As

10.3389/fpsyg.2023.1145493

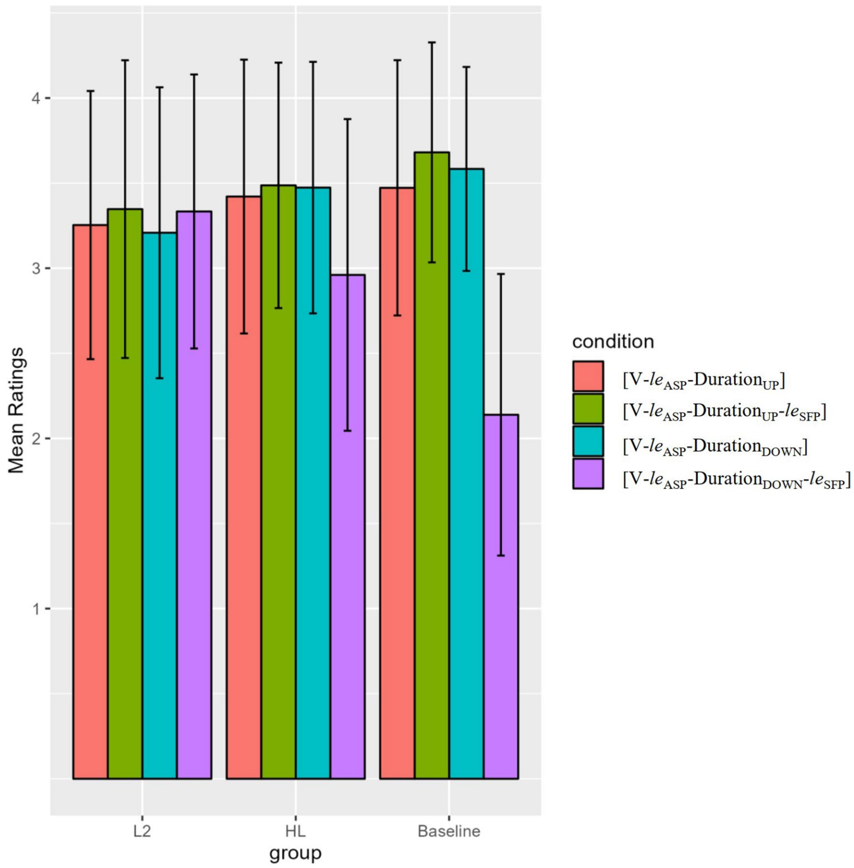


FIGURE 2
Mean ratings of the upward/downward-entailing duration phrase conditions in the acceptability judgment (AJ) task.

TABLE 3 Mixed-effects models on [V-le_{ASP}-Duration$_{UP}$-le_{SFP}] and */?[V-le_{ASP}-Duration$_{DOWN}$-le_{SFP}] in the acceptability judgment (AJ) task.

		Fixed effects	
Predictors	**Estimates**	***SE***	***t***
		HL learners	
(Intercept)	2.96	0.16	18.03***
Condition [V-le_{ASP}-Duration$_{UP}$-le_{SFP}]	0.53	0.20	2.57*
Group Baseline	−0.82	0.20	−4.18***
Group L2	0.37	0.20	1.90#
Condition [V-le_{ASP}-Duration$_{UP}$-le_{SFP}]: Baseline	1.02	0.23	4.23***
Condition [V-le_{ASP}-Duration$_{UP}$-le_{SFP}]: L2	−0.51	0.23	−2.23*
		Baseline	
(Intercept)	2.14	0.17	12.78***
Condition [V-le_{ASP}-Duration$_{UP}$-le_{SFP}]	1.54	0.21	7.40***
Group HL	0.82	0.20	4.18***
Group L2	1.19	0.20	6.00***
Condition [V-le_{ASP}-Duration$_{UP}$-le_{SFP}]: HL	−1.02	0.23	−4.42***
Condition [V-le_{ASP}-Duration$_{UP}$-le_{SFP}]: L2	−1.53	0.23	−6.57***

Reference level for condition = */?[V-le_{ASP}-Duration$_{DOWN}$-le_{SFP}], Model formula for baseline group: ratings ~ condition + group + condition:group + (1 + condition | subject) + (1 | item), Model formula for HL group: ratings ~ condition + group + condition:group + (1 + condition | subject) + (1 | item); ***$p < 0.001$, *$p < 0.05$, #$p < 0.1$.

expected, baseline controls showed statistical (though borderline) differences between the upward and downward conditions in the first post-critical window (W6, $p < 0.1$, *post hoc* power = 59.72%). What is surprising is that the HLers showed statistical differences (though also borderline) earlier than the baseline controls in the critical window (W5, $p < 0.1$, *post hoc* power = 42.11%), suggesting heightened

FIGURE 3
L2, HL, and baseline groups' residual reading times at each window between [*be*-NP$_{HIGHER}$-*le*$_{SFP}$] and */?[*be*-NP$_{LOWER}$-*le*$_{SFP}$].

sensitivity to violations caused by the downward-entailing contexts. The L2ers showed a third pattern and did not reveal any statistical differences in reading times in any window. Table 5 presents the LMM results.

To sum up, in reading the [*be*-NP$_{HIGHER}$-*le*$_{SFP}$] structures, both HL and L2 groups demonstrated sensitivity to violations of the semantic constraint in (1) caused by lower NPs in online sentence reading. Their sensitivity was evidenced in the same time window (second post-critical) as that of the baseline controls. By contrast, in reading the [V-*le*$_{ASP}$-duration-*le*$_{SFP}$] structures, both the HLers and the baseline controls, but not the L2ers, were sensitive to violations of the semantic constraint in (1) caused by downward-entailing duration phrases. Crucially, the HLers' sensitivity emerged very early in the critical time window during which the target SPF *le* was presented, earlier than that of the baseline controls at the first post-critical window after the target particle was presented.

TABLE 4 Mixed-effects models (dependent variable: residual reading time at the 2nd post-critical window [W6]).

	Fixed effects		
Predictors	**Estimates**	***SE***	***t***
	L2 learners		
(Intercept)	−105.21	73.94	−1.42
Condition [*be*-NP$_{HIGHER}$-*le*$_{SFP}$]	−82.69	46.16	−1.79#
	HL learners		
(Intercept)	−82.14	29.75	−2.76**
Condition [*be*-NP$_{HIGHER}$-*le*$_{SFP}$]	−80.68	42.07	−1.92#
	Baseline		
(Intercept)	−53.81	19.48	−2.76*
Condition [*be*-NP$_{HIGHER}$-*le*$_{SFP}$]	−46.50	18.65	−2.49*

Reference level for condition = */?[*be*-NP$_{LOWER}$-*le*$_{SFP}$], Model formula for baseline group: RT ~ condition + (1 | subject), Model formula for HL group: RT ~ condition, Model formula for L2 group: RT ~ condition + (1 | subject) + (1 | item); ** $p < 0.01$, * $p < 0.05$, # $p < 0.1$.

6. Discussion

6.1. Summary of findings

Our research questions investigated the acquisition and processing of Mandarin SFP *le*-sentences through comparisons along the following three dimensions: (i) **structure** effect – whether learners of Mandarin are equally sensitive to violations of the semantics of the SFP *le* caused by lower-NPs in [*be*-NP-*le*$_{SFP}$] as presented in (5) and downward-entailing elements in [V-*le*$_{ASP}$-duration-*le*$_{SFP}$] as demonstrated in (9); (ii) **task** effect - whether learners of Mandarin are equally successful in offline and online reading tasks; and (iii) **learner** effect – whether L2ers and HLers with matched proficiency in written Chinese are equally successful in recognizing such violations.

Our findings revealed structure, task, and the learner and their interactions all had an effect, albeit to different extents, as summarized in Table 6. In terms of **structure**, the L2ers were able to provide higher ratings for [*be*-NP-*le*$_{SFP}$] with higher NPs than those with lower NPs in the AJ task and displayed marginally longer processing times in reading [*be*-NP-*le*$_{SFP}$] with lower NPs than those with higher NPs; but they did not show signs of sensitivity to violations caused by the downward-entailing duration in [V-*le*$_{ASP}$-Duration-*le*$_{SFP}$]. The advantage of [*be*-NP-*le*$_{SFP}$] over [V-*le*$_{ASP}$-Duration-*le*$_{SFP}$] in eliciting sensitivity to violations is less clear in the other two groups, who recognized violations of both structures in both tasks. With respect to **task** effects, the offline AJ task elicited either stronger or equal (but never weaker) responses to violations compared with the online SPR task. This pattern holds across groups and structures. Finally, for different types of **learners**, the HLers outperformed the proficiency-matched L2ers in that the former recognized violations in both structures and in both online and offline tasks, whereas the latter only showed target-like sensitivity to the higher/lower NPs. Most strikingly, the HLers even outperformed the baseline controls, recognizing the violation earlier during the critical time window containing the critical SFP *le*, rather than in the post-critical window.

6.2. Conditions on positive cross-linguistic influence (CLI)

Given the similarities between the SFP *le* and *already*, it is expected that the English-dominant learners of Chinese (HLers and L2ers) would benefit from this cross-linguistic similarity and should not experience difficulty in acquiring the semantic constraint attached to the SFP *le*. Our findings revealed that this positive CLI is not always present, which is consistent with previous findings in the existing literature on the absence of positive CLI in acquiring L1-L2 similar structures (Bates and MacWhinney, 1989; Papadopoulou and Clahsen, 2003; Tokowicz and MacWhinney, 2005; Jing, 2008; Sabourin and Stowe, 2008). Our findings add new evidence to the conditional nature of CLI. The structure being tested obviously played a role, as shown by our L2 learners, who demonstrated little sensitivity to the violations caused by downward-entailing duration phrases, yet performed target-like with the lower NP conditions in both online and offline tasks. Note that violations caused by the two types of structures tested in this study are subject to the same semantic constraint in (1). The L2 learners' differential performance patterns clearly indicate that different structures under the same semantic condition in theoretical analyses could cause different problems for learners. Specifically, violations caused by the less embedded (linearly close to the target element) and semantically more transparent structures are more easily detected. This echoes the findings of Regulez and Montrul (2023) that L2 and HL acquisition are more sensitive to structural differences in the test stimuli.

So what prevents CLI from happening in this case? Despite similarities in terms of semantic meanings and conditions, the SFP *le* in Chinese and the adverb *already* in English lack apparent morphosyntactic similarity and correspondence. This adds to the difficulty for adult L2 learners to establish the necessary cross-linguistic mapping that would allow CLI to happen. In this light, positive CLI should not be taken for granted even when the L1 and L2 are similar. Furthermore, apparent morphosyntactic distance can prevent the desirable positive CLI from taking place. Pertinently, establishing the required mapping is not difficult for all Chinese-English bilinguals, which we return to in the next section.

6.3. Processing advantages of heritage bilinguals

The three groups of Chinese speakers in our study were all bilingual speakers of Chinese and English. On the one hand, the HLers and the baseline controls were both native and L1 speakers of Chinese; they differed in the onset age and context in which they acquired English and almost certainly in their proficiency in English, which was not tested in our study, representing a limitation. On the other, the HLers and the L2ers were both dominant in English; they also differed in the onset age and context in which they acquired Chinese. Although our cloze test, as a rapid and effective test for general proficiency in written Chinese, did not capture any difference between them, the HLers and L2ers did display considerable differences in the more challenging structure in our study (i.e., downward-entailing contexts).

We attribute the fast and accurate performance of the HLers to the advantage of co-activation of the SFP *le* and the adverb *already* in their bilingual representation. Conceivably, co-activation of Language A when processing Language B in bilinguals will either boost or compete with

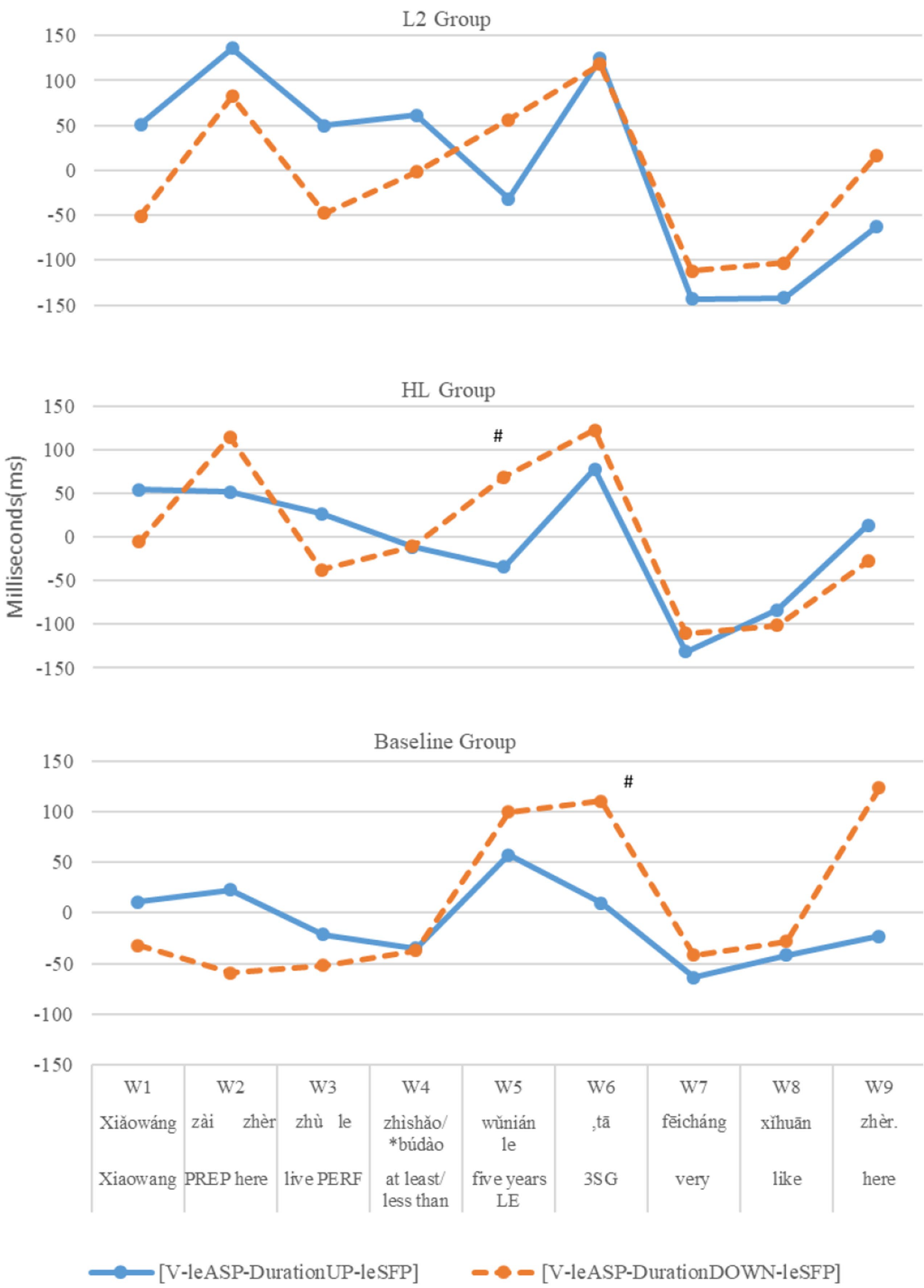


FIGURE 4
L2, HL, and baseline groups' residual reading times at each window between [V-leASP-DurationUP-leSFP] and */?[V-leASP-DurationDOWN-leSFP].

Language B. Since the L1-L2 structures in this study are subject to the same constraints, co-activating the English structure when processing the Chinese structure was more likely to boost and enhance, rather than damage or delay the processing. In our study, the HLers were early bilinguals who had been exposed to input in Chinese provided by native speakers of Chinese (from at least one of their parents) since birth as well as input in English provided by presumably native speakers of English in English-dominant contexts (UK, US, etc.). Among the three groups, the HLers stand out as the group that had the earliest and longest exposure to both languages and were probably the most balanced between the two languages. Crucially, English was added to the HLer's linguistic repertoire at a time when their Chinese was still rapidly developing and highly dynamic and fluid. It is possible that cross-linguistic mapping between the SFP *le* and the adverb *already* was established during this early stage and remained stable and accessible in the early bilinguals.

The L2ers were late bilinguals with an onset age for Chinese of 18 years on average. They did not seem to reliably and consistently co-activate the similar L1 structure while reading the L2. This is consistent with the findings of Canseco-Gonzalez et al. (2010) in terms of the effects of type of bilingualism in bilingual co-activation. This further explains why the L2 learners had problems in the more complex downward-entailing contexts even in offline tasks. The baseline controls were monolingually raised in Chinese and late bilinguals of Chinese and English. Given that *already* is a high-frequency adverb in English

TABLE 5 Mixed-effects models (dependent variable: residual reading time at the first post-critical window (W6) for baseline, and at critical window (W5) for HLers).

	Fixed effects		
Predictors	**Estimates**	***SE***	***t***
	HL learners		
(Intercept)	68.76	48.01	1.43
Condition [V-*le*$_{ASP}$-Duration$_{UP}$-*le*$_{SFP}$]	−102.97	61.97	−1.66#
	Baseline		
(Intercept)	110.67	36.98	2.99**
Condition [V-*le*$_{ASP}$-Duration$_{UP}$-*le*$_{SFP}$]	−100.92	52.29	−1.93#

Reference level for condition = */?[V-*le*$_{ASP}$-Duration$_{DOWN}$-*le*$_{SFP}$], Model formula for baseline group: RT ~ condition, Model formula for HL group: RT ~ condition + (1 | subject); ***p* < 0.01, #*p* < 0.1.

TABLE 6 Summary of findings: sensitivity to violations of the semantics of the SFP *le*.

	[*be*-NP-*le*$_{SFP}$]		[V-*le*$_{ASP}$-Duration$_{UP}$-*le*$_{SFP}$]	
	Offline AJ	Online SPR	Offline AJ	Online SPR
L2	√√	√	X	X
HL	√√	√	√√	√%
Baseline	√√	√√	√√	√

√√ and √ = significant results at and under 0.05 and 0.10 levels respectively; X = non-significant results; % = earlier sensitivity than baseline.

(regardless of variety of English), they must have been exposed to this usage in the English input they had received in China. Nevertheless, they did not seem to benefit from the knowledge of the adverb *already*. Our findings provide further evidence of the learner factor in modulating positive CLI, contributing to the ongoing discussion on bilingualism co-activation. As a reviewer suggested, the individual learner's processing strategies may also impact on their performance. We await further research to follow up on this issue.

7. Conclusion

This study has investigated potential positive CLI in judging and processing two grammatical structures with the SFP *le* by L2 and heritage Chinese bilinguals. Results of an offline judgment task and an online reading task show that positive CLI, though widely documented in the literature and desirable in this case, is not prevailing and is instead conditioned by the structure being tested (noun vs. verb phrase), the type of task administered (online vs. offline), and the type of the bilingual learner (early vs. late). Once in place, positive CLI and co-activation of L1 and L2 structures boost online processing proficiency, as evidenced in the superior performance of the heritage bilinguals in our study. Significantly, this study contributes to the discussion on CLI, particularly on the constraints affecting positive CLI, as well as early bilingual processing advantages.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee, Academic Committe, School of Chinese as a Second Language, Peking University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

SY designed the study and collected the experimental data. SY and ZM analysed the data and wrote up the manuscript. YZ provided funding, advised on the design and analysis, and revised the manuscript with SY and ZM. All authors approved the submitted version.

Funding

This work was supported by the National Social Sciences Project "Application of Chinese Proficiency Grading Standards and Strategies of Textbook Research and Development (21STA032)".

Acknowledgments

We are grateful to our mentors Boping Yuan and Waltraud Paul, the editor and reviewers for their kind comments and suggestions on this study, and Jiangling Zhou and Jingying Hu for their sharing and guidance on statistical R application. We also want to thank all participants who were generous with their time to make this study possible.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1145493/full#supplementary-material

References

Andersson, A., Sayehli, S., and Gullberg, M. (2019). Language background affects online word order processing in a second language but not offline. *Biling. Lang. Congn.* 22, 802–825. doi: 10.1017/S1366728918000573

Bates, D., Mächler, M., Bolker, B. M., and Walker, M. C. (2013). lme4: linear mixed-effects models using Eigen and S4. R package. Available at: https://cran.r-project.org/web/packages/lme4/index.html

Bates, E., and MacWhinney, B. (1989). "Functionalism and the competition model" in *The crosslinguistic study of sentence processing*. eds. B. MacWhinney and E. Bates (New York, US: Cambridge University Press), 3–73.

Bauke, L. S. (2020). "The role of verb-second word order for L1 German, Dutch and Norwegian L2 English learners: a grammar competition analysis," in *German and Dutch in contrast: Synchronic, diachronic and psycholinguistic perspectives* (*11th edition*). eds. VogelaerG. De, D. Koster and T. Leuschner (Berlin, Germany: De Gruyter), 241–270.

Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., and Wagner, K. (2010). Carpet or carcel: the effect of age of acquisition and language mode on bilingual lexical access. *Lang. Cogn. Processes.* 25, 669–705. doi: 10.1080/01690960903474912

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Costa, A., Caramazza, A., and Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1283–1296. doi: 10.1037/0278-7393.26.5.1283

De Groot, A. M. B. (2011). *Language and cognition in bilinguals and multilinguals: an introduction*. New York: Psychology Press.

Dean, A., and Voss, D. (1999). *Design and analysis of experiments* (pp. 649–657). New York: Springer-Verlag.

Dekydtspotter, L., Schwartz, B. D., and Sprouse, R. A. (2006). "The comparative fallacy in L2 processing research," in Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006): The Banff Conference). Eds. M. Grantham O'Brien, C. Shea and J. Archibald (Somerville, MA: Cascadilla Press), 33–40.

Dijkstra, T. (2005). "Bilingual visual word recognition and lexical access," in *Handbook of bilingualism: Psycholinguistic approaches*. Eds. J. F. Kroll and GrootA. M. B. De (Oxford: Oxford University Press), 179–201.

Felser, C., and Roberts, L. (2007). Processing wh-dependencies in a second language: a cross-modal priming study. *Second. Lang. Res.* 23, 9–36. doi: 10.1177/0267658307071600

Frenck-Mestre, C., and Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 50, 119–148. doi: 10.1080/027249897392251

Gelman, A., and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models (analytical methods for social research)*. Cambridge: Cambridge University Press.

Hasegawa, M., Carpenter, P. A., and Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage* 15, 647–660. doi: 10.1006/nimg.2001.1001

Hoover, M. L., and Dwivedi, V. D. (1998). Syntactic processing in skilled bilinguals. *Lang. Learn.* 48, 1–29. doi: 10.1111/1467-9922.00031

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second. Lang. Res.* 22, 369–397. doi: 10.1191/0267658306sr272oa

Hopp, H. (2017). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguist. Approach. Bilingualism* 7, 96–130. doi: 10.1075/lab.14027.hop

Hopp, H., and Grüter, T. (2023). The time-course of competition from the L1 grammar in L2 sentence processing: evidence from cross-linguistic structural priming. *Second. Lang. Res.* 39, 133–159. doi: 10.1177/02676583211009586

Huang, C., Li, Y., and Li, Y. (2009). *The syntax of Chinese (Cambridge syntax guides)*. Cambridge: Cambridge University Press.

Inagaki, S. (2001). Motion verbs with goal PPs in the L2 acquisition of English and Japanese. *Stud. Second. Lang. Acquis.* 23, 153–170. doi: 10.1017/S0272263101002029

Jackson, C. N., and Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Biling. Lang. Congn.* 12, 65–82. doi: 10.1017/S1366728908003908

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M., et al. (2007). Effect of syntactic similarity on cortical activation during second language processing: a comparison of English and Japanese among native Korean trilinguals. *Hum. Brain Mapp.* 28, 194–204. doi: 10.1002/hbm.20269

Jeong, H., Sugiura, M., Sassa, Y., Yokoyama, S., Horie, K., Sato, S., et al. (2007). Crosslinguistic influence on brain activation during second language processing: an fMRI study. *Biling. Lang. Congn.* 10, 175–187. doi: 10.1017/S1366728907002921

Jiang, N. (2004). Morphological insensitivity in second language processing. *Appl. Psycholinguist.* 25, 603–634. doi: 10.1017/S0142716404001298

Jing, M. (2008). Erroneous collocations caused by language transfer in Chinese EFL writing. *US-China Foreign Language* 6, 57–61. doi: 10.17265/1539-8080/2008.09.013

Kaan, E., Ballantyne, J. C., and Wijnen, F. (2015). Effects of reading speed on second-language sentence processing. *Appl. Psycholinguist.* 36, 799–830. doi: 10.1017/S0142716413000519

Keating, G. D., and Jegerski, J. (2015). Experimental designs in sentence processing research: a methodological review and user's guide. *Stud. Second. Lang. Acquis.* 37, 1–32. doi: 10.1017/S0272263114000187

Lago, S., Mosca, M., Stutter, , and Garcia, A. G. (2021). The role of crosslinguistic influence in multilingual processing: lexicon versus syntax. *Lang. Learn.* 71, 163–192. doi: 10.1111/lang.12412

Li, C. N., and Thompson, S. A. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Mai, Z., and Deng, X. (2017). Selective vulnerability and dominant language transfer in the acquisition of the Chinese cleft construction by heritage speakers. *Linguist. Approach. Bilingualism.* 9, 202–227. doi: 10.1075/lab.16040.mai

Mai, Z., and Yuan, B. (2016). Uneven reassembly of tense, telicity and discourse features in L2 acquisition of the Chinese shì … de cleft construction by adult English speakers. *Second. Lang. Res.* 32, 247–276. doi: 10.1177/0267658315623323

Michaelis, L. A. (1992). Aspect and the semantics-pragmatics interface: the case of already. *Lingua* 87, 321–339. doi: 10.1016/0024-3841(92)90015-B

Michaelis, L. A. (1996). On the use and meaning of "*already*". *Linguist. Philos.* 19, 477–502. doi: 10.1007/BF00632778

Mickan, A., and Lemhöfer, K. (2020). Tracking syntactic conflict between languages over the course of L2 acquisition: a cross-sectional event-related potential study. *J. Cogn. Neurosci.* 32, 822–846. doi: 10.1162/jocn_a_01528

Miller, K. A. (2014). Accessing and maintaining referents in L2 processing of wh-dependencies. *Linguist. Approach. Bilingualism* 4, 167–191. doi: 10.1075/lab.4.2.02mil

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge: Cambridge University Press.

Odlin, T. (1989). *Language transfer: cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, T. (2001). *Cross-linguistic influence in language learning*. Shanghai: Shanghai Foreign Language Education Press.

Omaki, A., and Schulz, B. (2011). Filler-gap dependencies and island constraints in second language sentence processing. *Stud. Second. Lang. Acquis.* 33, 563–588. doi: 10.1017/S0272263111000313

Papadopoulou, D., and Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: a study of relative clause attachment in Greek. *Stud. Second. Lang. Acquis.* 25, 501–528. doi: 10.1017/S0272263103000214

Paul, W. (2015). *New perspectives on Chinese syntax*. Berlin: De Gruyter Mouton.

Paul, W., and Yan, S. (2022). "Sentence-final particles in mandarin Chinese: syntax, semantics and acquisition" in *Discourse particles: syntactic, semantic, pragmatic and historical aspects, Linguistik Aktuell 276 (linguistics today)*. eds. X. Artiagoitia, A. Elordieta and S. Monforte (Amsterdam: John Benjamins Publishing Company), 179–206.

Polinsky, M., and Scontras, G. (2020). Understanding heritage languages. *Biling. Lang. Congn.* 23, 4–20. doi: 10.1017/S1366728919000245

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rankin, T. (2014). Variational learning in L2: the transfer of L1 syntax and parsing strategies in the interpretation of wh-questions by L1 German learners of L2 English. *Linguist. Approach. Bilingualism* 4, 432–461. doi: 10.1075/lab.4.4.02ran

Regulez, A. B., and Montrul, S. (2023). Production, acceptability, and online comprehension of Spanish differential object marking by heritage speakers and L2 learners. *Front. Psychol.* 14:1106613. doi: 10.3389/fpsyg.2023.1106613

Roberts, L., Gullberg, M., and Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Stud. Second. Lang. Acquis.* 30, 333–357. doi: 10.1017/S0272263108080480

Sabourin, L., and Stowe, L. A. (2008). Second language processing: when are first and second languages processed similarly? *Second. Lang. Res.* 24, 397–430. doi: 10.1177/0267658308090186

Soh, H. L. (2009). Speaker presupwindow and mandarin Chinese sentence-final-*le*: a unified analysis of the 'change of state' and the 'contrary to expectation' reading. *Nat. Lang. Linguist. Theory* 27, 623–657. doi: 10.1007/s11049-009-9074-4

Szeto, P. Y., Matthews, S., and Yip, V. (2017). Multiple correspondence and typological convergence in contact-induced grammaticalization. *J. Lang. Contact* 10, 485–518. doi: 10.1163/19552629-01002014

Tokowicz, N., and MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammars: an event-related potential investigation. *Stud. Second. Lang. Acquis.* 27, 173–204. doi: 10.1017/S0272263105050102

Traugott, E. C., and Waterhouse, J. (1969). 'Already' and 'yet': a suppletive set of aspect-markers? *J. Linguist.* 5, 193–320.

Vaughan-Evans, A., Paul, S., Fitzsimmons, G., and Manon, W. J. (2020). Syntactic co-activation in natural reading. *Vis. Cogn.* 28, 541–556. doi: 10.1080/13506285.2020.1841866

Xiang, K. (1998). On the prominence of entities and the marker *da*. *Hanyu Xuexi* 1, 12–16.

Xing, F. (2001). *A study of Chinese complex sentences*. Beijing: Shangwu yinshuguan.

Yan, S. (2023). *The right periphery in L2 Chinese: How sentence-final particles are represented in English-Chinese interlanguages*. Oxon/New York: Routledge.

Yip, V., Mai, Z., and Matthews, S. (2018). "CHILDES for bilingualism" in *Bilingual cognition and language: The state of the science across its subfields*. eds. D. Miller, F. Bayram, J. Rothman and L. Serratrice (Amsterdam/Philadelphia: John Benjamins Publishing Company), 183–201.

Yokoyama, S., Okamoto, H., Miyamoto, T., Yoshimoto, K., Kim, J., Iwata, K., et al. (2006). Cortical activation in the processing of passive sentences in L1 and L2: an fMRI study. *Neuroimage* 30, 570–579. doi: 10.1016/j.neuroimage.2005.09.066

Yuan, B. (2004). Negation in French–Chinese, German–Chinese and English–Chinese interlanguages. *Trans. Philol. Soc.* 102, 169–197. doi: 10.1111/j.0079-1636.2004.00134.x

Yuan, B. (2010). Domain-wide or variable-dependent vulnerability of the semantics-syntax interface in L2 acquisition? Evidence from wh-words used as existential polarity words in L2 Chinese grammars. *Second. Lang. Res.* 26, 219–260. doi: 10.1177/0267658309349421

Yuan, B., and Zhao, Y. (2005). Resumptive pronouns in English-Chinese and Arabic-Chinese interlanguages. *Int. Rev. Appl. Linguist. Lang. Teach.* 43, 219–237. doi: 10.1515/iral.2005.43.3.219

Zhao, Y. (2006). *Causativity in Chinese and its representations in English, Japanese and Korean speakers' L2 Chinese grammars*. Beijing: Peking University Press.