BMC Psychology

RESEARCH

Open Access

The role of format familiarity and semantic transparency in Chinese reading: evidence from eye movements

Mingjing Chen[1*], Li-chih Wang[1], Sisi Liu[1] and Duo Liu[1*]

Abstract

Unlike alphabetic language, Chinese is an ideographic language that does not contain spaces between words. Chinese readers must develop unique segmentation strategies for word recognition and reading comprehension. This study explored the role of format familiarity and semantic transparency in Chinese reading, reflecting the segmentation strategy and word processing characteristics in Chinese reading. Forty undergraduates read Chinese in familiar and unfamiliar formats, segmenting target words into semantically transparent and semantically opaque words. We used Eye Link 1000 to measure readers' eye movement index, which can reflect processing characteristics of word recognition in Chinese reading. The following findings were made: (1) Familiarity with the text format affects Chinese reading performance. The fixation time in the familiar direction is short, the skipping rate is high, and the processing efficiency is higher when the fixation point is close to the word center; (2) Semantic transparency affects the segmentation strategy and word processing in Chinese reading. Chinese readers have shorter fixation times, higher reading efficiency, and a fixation point closer to the word center when reading semantically transparent words. It supported the combined access model. (3) There is significant interaction in the early eye movement indicators, representing word processing characteristics in the early stage of Chinese reading. Specifically, the semantic-transparency effect appeared under a familiar rather than an unfamiliar format. The format familiarity effect was found in the early processing indexes of transparent words rather than opaque words. In the familiar format, since the meaning of the morpheme and the whole word of transparent words is consistent, readers tend to segment and process them as whole words. Due to the lack of reading experience, the reading difficulty increases in the unfamiliar format. To reduce the difficulty and promote comprehension, readers change their segmentation strategy and tend to segment transparent words by character. The word segmentation process slowed, and the format-familiarity effect did not show in the early indexes under unfamiliar format. More importantly, the separability of the lexical processing stages showed in the interaction of different indexes, which means that word segmentation and lexical recognition in Chinese reading may not be completely synchronized, supporting the Chinese E-Z reader model.

Keywords Eye movement, Chinese reading, Format familiarity, Semantic transparency, Chinese E-Z reader model, Integrated model, Combined access model

Background

Relationship between format familiarity and segmentation strategy

Chinese is an ancient ideographic language with a history of over a thousand years [16]. It is distinct from alphabetic languages and possesses unique characteristics and

*Correspondence:
Mingjing Chen
Lsr_psy@126.com
Duo Liu
s1144055@eduhk.hk
[1] Department of Special Education and Counselling, The Education University of Hong Kong, Hong Kong, SAR, China

global influence [3]. Generally, the smallest semantic unit in reading is a word; in some contexts, such as ambiguous words or fuzzy words, the semantic units may also be characters [31]. Chinese texts lack inter-word spaces, necessitating a different approach to segmenting them [11]. Consequently, it is necessary to develop a unique segmentation strategy to divide continuous Chinese texts into words or characters [11].

Because the smallest semantic units differ, two distinct text segmentation strategies are generally employed in Chinese reading. The former approach is that the basic semantic unit of Chinese reading is the word [3, 24, 26, 27, 35, 39]. When the text is difficult to read or ambiguous, the basic unit of semantic understanding is not the word but the character, which means that Chinese readers must segment continuous texts into single characters one by one. For instance, there are some texts containing low-frequency words [36] or opaque words [53], where a high level of lexical processing difficulty for readers makes it possible to adopt a character-segmented strategy rather than a word-segmented strategy [18, 22, 30, 50].

The two segmented strategies also result in two distinct processing strategies at the level of lexical processing: whole-word processing and morpheme processing [5]. These two processing strategies in Chinese reflect the mental representation and processing mechanism of Chinese compound words, which have been a widespread concern in psycholinguistic research [7]. The latest Chinese E-Z reader model uses six models to simulate and explore the mental representation and processing mechanism of Chinese compound words in Chinese reading. The eye movement characteristics of the readers were analyzed, and the results showed that the best matching model was the familiarity segmentation model: readers used relative familiarity to segment Chinese texts for word recognition, which was the most efficient reading strategy [33]. Combined with previous studies, a trade-off exists between format familiarity and word segmentation, and unfamiliarity offsets the facilitation of inter-word spaces in Chinese reading [11]. Accordingly, this study considered the effect of familiarity in exploring word processing and segmentation strategies.

The relationship between segmentation and lexical representation behind semantic transparency

To investigate this issue, the semantic-transparency effect of two-character compound words (transparent and opaque) is generally used to examine the lexical processing mechanism, including segmentation and lexical recognition in Chinese reading [13, 49, 52, 53, 55, 56, 62]. Semantic transparency refers to the semantic relatedness between the word and its morphemes, which means to what extent the semantics of the compound word can be inferred from the semantics of each morpheme [17]. According to it, Chinese two-character compound words can be divided into two categories: transparent and opaque. The former refers to compound words with a high correlation between morphemes and whole words (e.g., 道路: the first morpheme "道" and the second morpheme "路" of the word "道路" both have the same Chinese meaning as "road"). The latter refers to compound words with a low correlation between morphemes and whole words (e.g., 马虎: the first morpheme, "马"(horse), and the second morpheme, "虎" (tiger) both represent an animal, but the meaning of the whole word "马虎" represents carelessness.). The consistency between morphemes and whole word meaning means that transparent words can be quickly recognized and processed, while opaque words are more difficult to process and recognize [38]. The reading difference between the two is called the semantic transparency effect, and eye trackers have generally been used to capture this effect during natural reading through eye movement indexes [19, 55, 60–62].

Behind the semantic-transparency effect is the lexical representation of transparent or opaque words in two characters, and there is no consistent conclusion yet. Three representations of lexical processing are whole word processing, morpheme processing, and combined processing [29]. The combined access model (CA) suggests both morpheme and whole-word representations are in the mental lexicon. The recognition of compound words results from the interaction between morpheme and whole-word activation. Whether it is a transparent or opaque word, whole word processing and morpheme processing are involved [29, 41, 52]. The meaning computation view analyzes from another perspective, suggesting that the morpheme semantics of transparent compound words are consistent with the whole word meaning obtained by integration and calculation, thereby promoting the accessibility of the whole word semantics. In contrast, the morpheme semantics of opaque compound words compete with the meaning obtained by integration and calculation, thereby hindering the accessibility of the whole word semantics. This view also means that both transparent and opaque words involve two processes and further refines the relationship between morpheme processing and whole word processing: competition or cooperation [14, 20]. Neither of the above two views prioritizes whole word processing over morpheme processing. However, some researchers assume that morphemes are secondary in recognizing opaque words. Transparent words are processed as morphemes, while Chinese readers process opaque words as whole words. This view of prioritizing whole-word recognition and morpheme recognition is also supported by some findings [53]. In

summary, the theoretical explanation of the lexical representation of the semantic transparency effect of Chinese compound words is still somewhat controversial. The first objective of this study is to provide empirical evidence for the above theories through the semantic transparency effect.

Relationship between format familiarity and semantic transparency

The representations debate between semantically transparent and opaque words may come from the reading experience of Chinese native adults, which masks differences between lexical or segmentation strategies [11]. When readers are native Chinese adults, the segmentation and lexical processing are automated. Even if there are differences in segmentation or recognition strategies between transparent and opaque words, a rich reading experience as a higher-level cognitive factor can compensate for the differences between transparent and opaque words from top to bottom [44]. After adjustment of reading experience, those subtle differences between transparent and opaque words can not be perceived in regular reading performance. Therefore, to study the lexical processing mechanism in Chinese reading more effectively, a recent study, unlike traditional studies, generated novel findings by varying familiarity with the Chinese format by changing the reading direction [11]. By manipulating the two variables (inter-word space and reading direction), the researchers found that the inter-word space as a word segmentation cue cannot facilitate regular reading performance in unfamiliar directions. Readers lack reading experience in unfamiliar directions; daily reading experience can not offset the facilitation of word segmentation on this condition [10]. Furthermore, the researchers subsequently manipulated word frequency and format familiarity and found that readers displayed more flexible processing strategies under unfamiliar reading directions [9]. Without the adjustment of reading experience, reading differences between high-frequency and low-frequency words that could not be discovered in daily life were often manifested under unfamiliar reading directions. Thus, readers adopt more flexible strategies of lexical processing in unfamiliar directions, which can be extended in the processing mechanism of semantic transparency in Chinese reading. Following the previous definition, this study refers to readers' familiarity with the direction they read as 'format familiarity' since modern Chinese texts are read horizontally, from left to right [3]. At the same time, readers read from right to left in many languages, such as Hebrew [10]. Thus, Chinese readers are unfamiliar with the right-to-left format of Chinese texts as they are more accustomed to the left-to-right format. The second question is about the role of format familiarity in the lexical processing of Chinese transparent and opaque words.

The research questions proposed the following hypothesis: Readers read in a familiar direction for shorter reading times, higher skip rates, and longer reading distances. Transparent words were read in shorter reading times, higher skip rates, and longer reading distances. If there is an interaction between format familiarity and semantic transparency, the transparency effect observed in the familiar format (left to right) and the unfamiliar format (right to left) should be inconsistent, perhaps appearing in the familiar format but not in the unfamiliar format, or being smaller in the unfamiliar format. Conversely, there is no significant interaction between the two variables. Furthermore, the format-familiarity effect may exist in transparent words but not in opaque words from the early processing indicators related to the lexical representation. In the familiar format, the morpheme meaning of transparent words is consistent with the meaning of the whole word; readers tend to segment as whole words. Due to the lack of reading experience, the reading difficulty of unfamiliar formats increases. To reduce the difficulty and promote understanding, readers change their segmentation strategies and tend to segment transparent words by character. Therefore, the processing difference between transparent words in familiar and unfamiliar formats is significant, and the format familiarity effect appears on transparent words. However, opaque words' morphemes and whole meanings have semantic conflicts, and readers need to compete to acquire the word's meaning. Readers tend to segment opaque words by character in the familiar format, so the segmentation strategy does not change in the unfamiliar format, and the reading difference between the two formats is insignificant, so the format-familiarity effect does not appear on opaque words in the early indexes (e.g., First fixation Duration and Single fixation Duration). In the later stages of lexical processing, the segmentation strategy change has also been completed. There should be no difference in the interaction, whether it is a transparent or opaque word. Furthermore, the difference in interaction between early and late indexes reflects the separability of word segmentation and recognition, supporting the latest E-Z reader model of Chinese reading [33].

Methods

The aim is to investigate the role of format familiarity and semantic transparency in Chinese reading.

Participants

Forty undergraduate students (mean age 20.50 ± 1.63 years), 30 female and 10 male, participated in the experiment. All were right-handed, native Chinese speakers with normal

or corrected-to-normal vision. Before the experiment, a signed informed consent form was obtained from each participant. The ethics committee of the author's university approved the experiment as complying with the Declaration of Helsinki.

Design

The experiment had a 2 (format familiarity: reading from left to right, reading from right to left)×2 (semantic transparency: transparent word, opaque word) two factors within-subjects design.

Materials

Semantic transparency refers to the degree to which the semantics of a compound word can be inferred from the semantics of its constituent morphemes. Its operational definition is the degree of semantic relevance between the whole word and its morphemes [29, 53, 55, 62]. The experimental materials come from the Modern Chinese Dictionary (2005). In the first step, the researchers selected 620 two-character words, of which 316 were transparent and 304 were opaque two-character words. Two psychology graduate students completed the evaluation of 620 two-character words, and the rating agreement reached 97%. Then, 100 college students who were native Chinese speakers did not participate in the eye-tracking experiment. They were asked to evaluate the semantic transparency on a 7-point scale, that is, to evaluate the degree of association between the morpheme (the first character) of a two-character word and the meaning of the whole word, as well as the degree of association between the final morpheme (the second character) and the meaning of the whole word. Among them, 1 represents the lowest degree of meaning association, that is, the lowest transparency, and 7 represents the highest degree of meaning association, that is, the highest transparency [27, 54]. For example, the first character "道" and the second character "路" of the word "道路" both have the same Chinese meaning as "road," and the meaning of the whole word is also "road," so whether it is the first character or the last character, the meaning is consistent with the whole word. Therefore, the participants gave "道" a score of 7 and "路" a score of 7. The semantic transparency score of this word is 7, so it is a transparent word. The first character, "马"(horse) of the word "马虎" and the second character, "虎" (tiger), both represent an animal, but the meaning of the whole word "马虎" represents carelessness. Hence, the first and last characters have different meanings from the whole word. Therefore, the participants gave "马" a score of 1 and "虎" a score of 1, so the semantic transparency score of this word is 1, which is an opaque word.

The scoring questionnaire is included in the Appendix, and the semantic relatedness between the first and second characters was averaged to form the semantic transparency score of the target word [27, 55]. According to the semantic transparency scores, the highest 25% were selected as high transparency words (79 words), M = 6.45, SD = 0.94; the lowest 25% were selected as low transparency words (76 words), M = 2.02, SD = 1.53. Then, the word frequency, first character frequency, last character frequency, first character stroke number, and last character stroke number of the selected 155 words were checked based on the public Chinese word database [6]. Finally, forty groups of two-character words were selected as target words. There was no significant difference between semantically transparent words and opaque words in the frequency of target words, the frequency of first characters, the frequency of last characters, the number of strokes of the first character, and the number of strokes of the last character ($ts < 0.137$, $p > 0.1$). The unit of word frequency and character frequency was the occurrences per million words (OPM). The results are shown in Table 1. The semantic transparency of high-transparency words (M = 6.44, SD = 0.21) and low-transparency words (M = 2.02, SD = 0.26) was significantly different (t = 80.22, $p < 0.001$).

The 80 target words were placed in the corresponding contexts, and 80 declarative sentences with the target words that were not at the beginning or end were created. A total of 80 sentences were created, comprising 40 sentences containing semantically transparent target words and 40 sentences containing semantically opaque words. The sentences were presented in two different text formats: unfamiliar and familiar. The unfamiliar format consisted of 20 sentences with semantically opaque

Table 1 The statistical characteristics of the experimental sentences and target words

Experimental conditions	Word Frequency	First- character strokes	Second-character strokes	First- character frequency	Second-character frequency	Naturalness
Transparent Word	85.57 (137.90)	9.4 (1.84)	9.10 (3.08)	111.57 (233.02)	152.78 (402.42)	6.30 (0.23)
Opaque Word	79.50 (146.71)	8.80 (2.50)	8.40 (3.23)	124.62 (213.18)	185.13 (496.87)	6.30 (0.28)

The word frequency unit is 1/millions, and the standard deviation is in brackets

target words and 20 sentences with semantically transparent target words. The familiar format comprised 20 sentences with semantically opaque target words and 20 sentences with semantically transparent target words. The sentences were 12 to 18 Chinese characters in length. Thirty-one undergraduates who were not involved in the experiment were selected to evaluate the fluency of the experimental sentences using a 7-point scale. The scores of five subjects who did not meet the criteria were excluded. The fluency of the sentences was M = 6.30 (SD = 0.26). There was no significant difference in the fluency of the sentences under the four conditions, as indicated by the $Fs < 0.46$. The fluency of the sentences met the experimental requirements. Table 1 provides further details on the experimental materials. Table 2 presents examples of the experimental materials on four conditions.

Apparatus

The experiment employed the EyeLink 1000 (SR Research, Canada) to record right-eye movements. The sampling rate was 1000 Hz. The stimuli were presented on a 19-inch Dell monitor with a resolution of 1024 × 768. The participants were required to maintain a distance of 70 cm from the screen. Each character was 25 pixels wide by 25 pixels tall, with a visual angle of 0.80°. The size of the Song font was 20.8 in the stimuli presentation.

Procedure

Each participant was tested individually. Upon the participant's arrival at the laboratory, they were provided a brief introduction to the laboratory environment and then completed a series of primary information forms.

Table 2 Example Chinese stimuli from the four experimental conditions

Format Familiarity	Semantic Transparency	Sentence
Familiar Format (left to right)	Transparent	盘山公路上危险的地方都立着警示牌。
	Opaque	我们推选书记为新一届人民大会代表。
Unfamiliar Format (right to left)	Transparent	。牌示警着立都方地的险危上路公山盘
	Opaque	。表代会大民人届一新为记书选推们我

The target words are highlighted in blue and bold, but in the experiment, the target words are the same as other words. Under the semantic transparency-familiar format, the transparent target words were presented from left to right; under the semantic transparency-unfamiliar format, the transparent target words were presented from right to left. In the semantic opaque-familiar format, the opaque target words were presented from left to right, and in the semantic opaque-unfamiliar format, the opaque target words were presented from right to left. The italic words are target words, which appeared in normal form in the experiment

Subsequently, the participants were seated at a distance of 70 cm from the eye tracker, with their chins placed on the chin rests, which were employed to ensure that the heads remained resting with no movement. A three-point calibration of the eye was then conducted. A calibration is deemed successful when the average value is less than 0.2. Once the eye calibration was completed, the participant began reading the experimental sentences on the screen.

Participants were then instructed to read sentences in various conditions. They were required to comprehend the meaning of the sentences as rapidly as possible and then to press the space bar to read the subsequent sentence. A comprehension question was posed for specific sentences, which the participants had to answer as accurately as possible. The initial 12 sentences were practice sentences. The experiment comprised four groups of materials, each containing 80 formal sentences. Twenty sentences were presented per condition, with the sentences in each group being randomly presented. Each participant was required to complete only one of the four groups. During the experimental phase, A reading comprehension question was inserted after 18 of the experimental sentences, followed by simple "yes" or "no" judgment questions. The participants needed to read carefully before answering correctly. The rate of correct responses to these questions was 93%, indicating that the sentences were predominantly read and understood. Eye calibration was re-performed during the experiment when necessary, and the entire experimental process took 20–30 min.

Data preparation and analysis

In line with criteria from previous studies, if the fixation point is either excessively long or short, or if the fixation time is less than 80 ms or greater than 800 ms, the data were excluded [3, 9, 11, 21, 22, 36, 51, 54]. In addition, the data were filtered according to the following criteria, which are consistent with previous studies [2, 3, 9, 11, 15, 23, 36, 63]: (1) The participant initiates the key press too early or incorrectly during the experiment, resulting in a sentence interruption. (2) The tracking data are lost due to accidental factors (such as the participant's head movement) during the experiment. (3) The number of fixations is less than four times. (4) The data whose mean is outside three standard deviations. Invalid data constituted 3.2% of the total data set.

Eye movement indicators are usually divided into two categories—time indicators and position indicators. Time indicators are divided into two categories: one can reflect the characteristics of early vocabulary processing, and the other can reflect the characteristics of late processing. According to the calculation method of eye

movement indicators, FFD, SFD, and GD are regarded as eye movement indicators that can reflect readers' early lexical processing. (1) First fixation duration (FFD), the duration of the first fixation on a word, irrespective of the number of fixations. It is one of the most commonly used eye movement indicators, effectively reflecting the early-stage characteristics of lexical processing [46, 47]. However, it is not a perfect indicator because it confuses the situation of single fixation and multiple fixations in the interest area, and the psychological processing in these two situations is different. Therefore, we need to supplement (2) single fixation duration (SFD), the fixation duration when only one fixation was made on the word during first-pass reading, and it is considered to be a good indicator of the semantic activation stage in word recognition [8]; (3) gaze duration (GD), the sum of all fixations on a word before moving to another word, GD is also an indicator reflecting the early stage of vocabulary access [46]. It is sensitive to pre-lexical and lexical features and is also one of the most widely used indicators. It is proposed that there is a difference between the FFD and the GD, so it can be inferred that the reader has difficulty in the first processing of words [9]. Based on the Chinese E-Z reader model, FFD and SFD represent the early stage of lexical recognition [33, 45]. They can sensitively reflect that readers achieved word recognition in the first or only one fixation without re or multiple fixations. The characteristic well reflects the early processing stage of word recognition [65]. When the regression appears, single or first processing characteristics will be submerged. The later process indicators usually include the following indicators: (4) regression-path duration (RPD), the sum of all gaze times looking back to the current word; (5) total time (TT), the sum of all fixations on the target word, including regressions, RPD is an indicator of the later processing of words. At the same time, the TT is sensitive to slower and longer cognitive processing [46]. Regardless of the time index, the reader needs more time when the reading comprehension difficulty is higher, and the processing efficiency will be reflected in the time indexes [9, 12, 36, 54]. When readers spend more time understanding the target words, it also means that the reader's reading efficiency is lower [10]. The RPD and TT are eye movement indicators that reflect the late and slower cognitive processes of word recognition; they are unaffected by the number of fixations [8, 12, 33, 57, 64, 65].

In addition, time indicators indicate the reader's reading efficiency, processing difficulty, and cognitive characteristics through the fixation duration. In contrast, position indicators usually reflect the reader's processing characteristics during reading based on the distance and landing point of the eye saccade [2, 37, 58, 63]. For instance, the center of the word is the best fixation position for eye movements, and the reading efficiency is highest at the best fixation position. The farther the fixation position is from the word center, the lower the fixation efficiency [9, 11, 36]. The center of the word is the best fixation position for eye movements, and the reading efficiency is highest at the best fixation position. The farther the fixation position is from the word center, the lower the fixation efficiency [2, 23, 37, 58, 66]. Position indicators include: (6) skipping probability (SP), the probability of skipping the target region in the first reading, specifically, the ratio of the frequency of the interest area being skipped to the sum of the frequency of the interest area being skipped and fixed in the first reading [9–11]. SP indicates that the skipped words have been processed in the parafoveal visual area, which may occur before or after the skipping [43]. The lower SP means that the text is more difficult to comprehend, and the reader's reading efficiency is inefficient. (7) refixation rate (RR), the probability of the target region being gazed at multiple times in the first reading. It is sensitive to cognitive variables and is considered to be of great significance in eye movement research. When the refixation rate is high, the reader cannot fully understand the first fixation and needs to make the next fixation, which also reflects the reader's low reading efficiency [42]; and (8) initial landing position (ILP), the distance to the beginning of the target word for the first time [3, 9, 11, 36, 54, 58, 63]. The farther from the beginning of the word, the greater the reader's eye saccade distance, indicating that the subject obtains relatively more information in the fixation before the eye saccade, and the reading speed is faster. As the difficulty of the article increases, the reader's eye saccade distance becomes shorter, and the ILP is shorter [40, 42]. The time index units were milliseconds, and the unit of ILP is the character [9–11, 28].

Data analysis includes the local analysis with target words as the area of interest [3, 9, 11, 36, 46]. A total of eight eye-movement measures were computed for the target words. All collected data were analyzed using the linear mixed model (LMM), implemented in the R language (R Development Core Team, 2016), and the lme4 data processing packages [4]. The data were classified as significant if the t-value exceeded 1.96 at the 5% level [9–11]. The participants and items were specified as crossed random effects during the linear mixed model. The Markov-Chain Monte Carlo algorithm was employed to derive the model parameters of the post-hoc distribution, which served as an estimate of significance [9–11]. This estimate simultaneously reflected the variation from participants and items [1]. The analysis indicators were log-transformed during the model operation, and the Logistic LME transformation was performed on skipping

data and re-fixation probability. The LMM was designed to assess the impact of semantic transparency and format familiarity on the interaction between the two factors [9–11]. When the interaction was significant, transparent words were compared with opaque words in the familiar format (Comparison 1), and transparent words were compared with opaque words in the unfamiliar format (Comparison 2) [9–11].

Results

The results presented in Table 3 show the means and standard deviations for the eye movement measures of the target words, which were analyzed using an Linear Mixed Model (LMM) approach with format familiarity, semantic transparency, and their interaction as the fixed factors. Table 4 shows the t-values of the eight indexes of the target words [9–11]. To show the reading process of readers more visually, Figs. 1, 2, 3 and 4 simulated the eye-tracking process under different conditions.

The results demonstrated that the format familiarity effect and semantic transparency effect in the fixation duration (FFD was shorter for the transparent words than for the opaque words: $b=0.038$, $SE=0.012$, $t=3.294$, $p=0.001$. Furthermore, the FFD was also shorter from left to right than from right to left: $b=0.078$, $SE=0.014$, $t=5.388$, $p<0.001$. The SFD was shorter for the transparent words than the opaque words: $b=0.035$, $SE=0.015$, $t=2.424$, $p=0.016$. Additionally, the SFD was shorter from left to right than from right to left: $b=0.05$, $SE=0.017$, $t=3.139$, $p=0.002$. The GD was shorter for the transparent words than the opaque words: $b=0.119$, $SE=0.018$, $t=6.540$, $p<0.001$. GD was shorter from left to right than right to left: $b=0.253$, $SE=0.023$, $t=11.206$, $p<0.001$. The RPD was shorter for the transparent words than the opaque words: $b=0.180$, $SE=0.025$, $t=7.134$,

Table 3 Eye-movement indexes for the target words

Indexes	Familiar format (reading from left to right)		Unfamiliar format (reading from right to left)	
	Transparent words	Opaque words	Transparent words	Opaque words
FFD (ms)	237(70)	261(85)	271(94)	275(97)
SFD (ms)	239(70)	262(85)	267(91)	274(99)
GD (ms)	269(106)	322(143)	382(181)	422(205)
RPD (ms)	327(196)	406(260)	527(360)	626(400)
TT (ms)	364(180)	479 (254)	611(330)	737(382)
SP	0.24(0.39)	0.1(90.33)	0.18(0.30)	0.12(0.27)
RR	0.13(0.26)	0.44(0.39)	0.25(0.53)	0.35(0.45)
ILP	0.92(0.54)	0.88(0.52)	0.81(0.50)	0.74(0.45)

Standard deviations are provided in parentheses

FFD First fixation duration, *SFD* Single fixation duration, *GD* Gaze duration, *RPD* Regression-path duration, *TT* Total time, *SP* Skipping probability, *RR* Refixation rate, *ILP* Initial landing position

Table 4 T-values of total indicators of target words

Indexes	Semantic Transparency	Format Familiarity	Interaction	Compare 1	Compare 2
FFD (ms)	3.294^{**}	5.388^{***}	-2.442^{*}	4.055^{***}	0.694
SFD (ms)	2.424^{*}	3.139^{**}	-2.155^{*}	3.879^{***}	0.584
GD (ms)	6.540^{***}	11.206^{***}	−1.472		
RPD (ms)	7.134^{***}	12.655^{***}	−0.598		
TT (ms)	8.500^{***}	14.729^{***}	−0.454		
SP	-3.191^{**}	-6.230^{***}	0.584		
RR	5.856^{***}	14.000^{***}	−0.554		
ILP	-2.43^{*}	-4.692^{***}	−0.275		

Interaction = the interaction between the semantic transparency and the format familiarity. Comparison 1: the semantic transparent condition would be compared with the opaque condition in the familiar format (reading from left to right). Compare 2: the transparent condition and the opaque condition would be compared in the unfamiliar format (reading from right to left)

FFD First fixation duration, *SFD* Single fixation duration, *GD* Gaze duration, *RPD* Regression-path duration, *TT* Total time, *SP* Skipping probability, *RR* Refixation rate, *ILP* Initial landing position

$^{***}p<0.001$, $^{**}p<0.01$, $^{*}p<0.05$

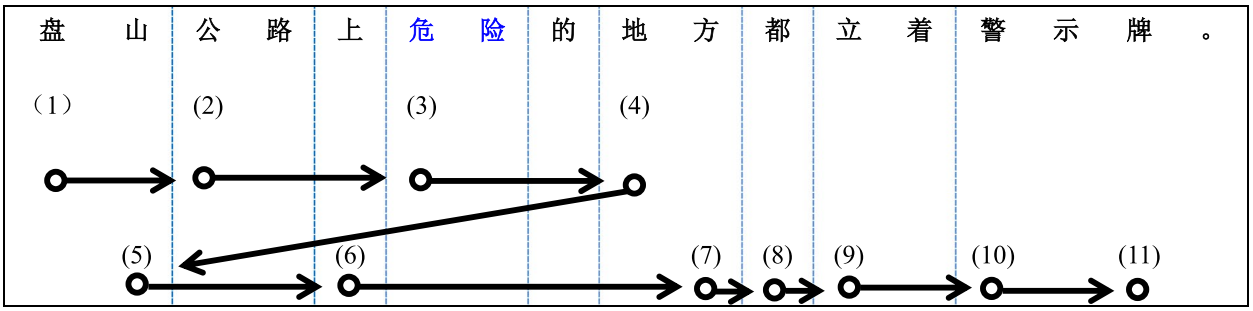

Fig. 1 The simulation trajectory diagram under familiar and transparent condition. Note: This trajectory diagram uses words as interest areas, and interest areas are separated by blue lines. The dots represent fixation points, the arrows represent eye-saccade directions, and the numbers represent the fixation order. The entire trajectory is designed to simulate the eye movement trajectory of readers during reading. For example, if readers have already fixed their gaze on the fourth fixation point (4) but do not fully understand the word meaning before the fourth fixation point (4), the reader's eye saccade will return to the fifth fixation point (5), and then continue reading. If the readers understand the meaning, they will not return to the fifth fixation point (5). The same applies below

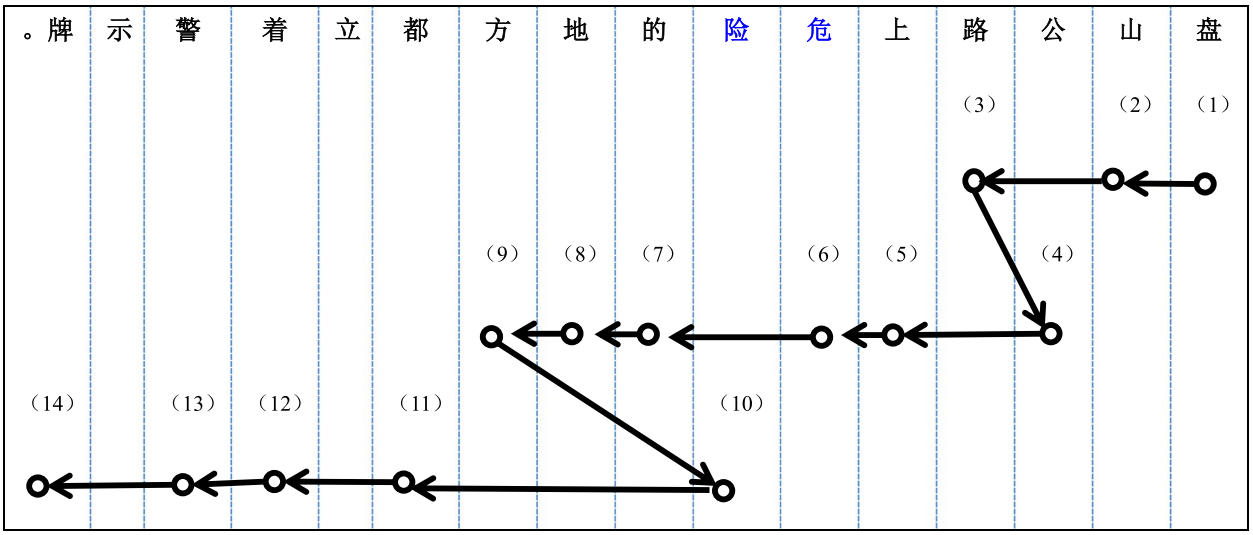

Fig. 2 The simulation trajectory diagram under unfamiliar and transparent condition. Note: This trajectory diagram uses characters as interest areas, and interest areas are separated by blue lines. The dots represent fixation points, the arrows represent eye-saccade directions, and the numbers represent the fixation order. The entire trajectory is designed to simulate the eye movement trajectory of readers during right-to-left reading. For example, if readers have already fixed their gaze on the third fixation point (路) but do not fully understand the word meaning before the third fixation point (路), the reader's eye saccade will return to the fourth fixation point (公), and then continue reading until the ninth fixation point (方). If the readers understand the meaning, they will continue reading until the fourteenth fixation point (牌). The same applies below

$p < 0.001$. The RPD was also shorter from left to right than from right to left: $b = 0.393$, $SE = 0.031$, $t = 12.655$, $p < 0.001$. TT was shorter for the transparent words than the opaque words: $b = 0.23$, $SE = 0.027$, $t = 8.500$, $p < 0.001$. Furthermore, TT was also shorter from left to right than from right to left: $b = 0.471$, $SE = 0.032$, $t = 14.729$, $p < 0.001$). The format familiarity effect found in this study is consistent with the results of previous studies [9–11].

Readers with familiar formats have better reading performance, shorter reading time, and higher reading efficiency [9–11]. It also means the reading experience behind format familiarity will significantly impact Chinese readers. Even for very proficient Chinese native speakers, the reader's reading performance will drop rapidly once the text direction is changed. It also supports the latest E-Z reader model in Chinese reading [33]. The reading experience, as a high-level cognitive factor, always affects the reading process of Chinese readers from top to bottom [44, 46, 47], including the initial visual familiarity verification stage, the word segmentation stage, and the lexical recognition stage. The latest eye movement model of Chinese reading proposes that familiarity affects the segmentation stage of Chinese reading,

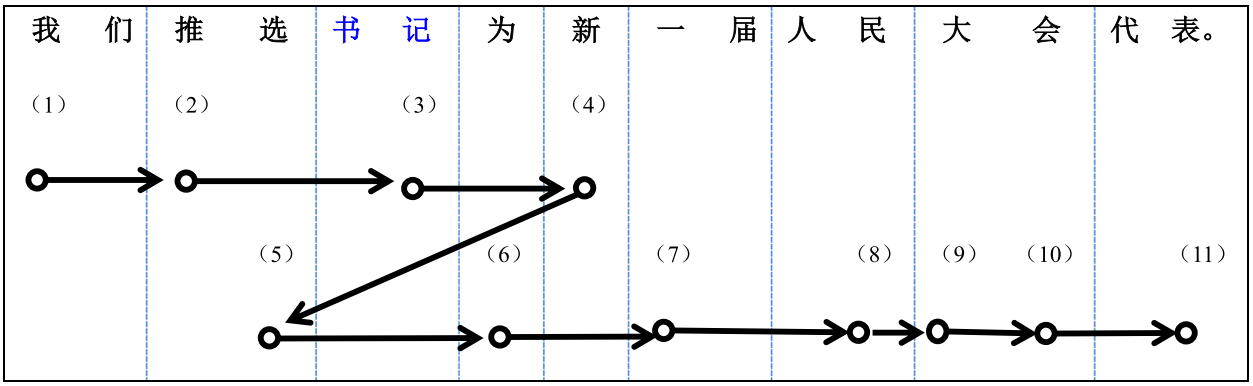

Fig. 3 The simulation trajectory diagram under familiar and opaque condition. Note: This trajectory diagram uses words as interest areas, and interest areas are separated by blue lines. The dots represent fixation points, the arrows represent eye-saccade directions, and the numbers represent the fixation order. The entire trajectory is designed to simulate the eye movement trajectory of readers during reading. For example, if readers have already fixed their gaze on the fourth fixation point (新) but do not fully understand the word's meaning before the fourth fixation point (新), the reader's eye saccade will return to the fifth fixation point (推选), and then continue reading until the eleventh fixation point(代表). The same applies below

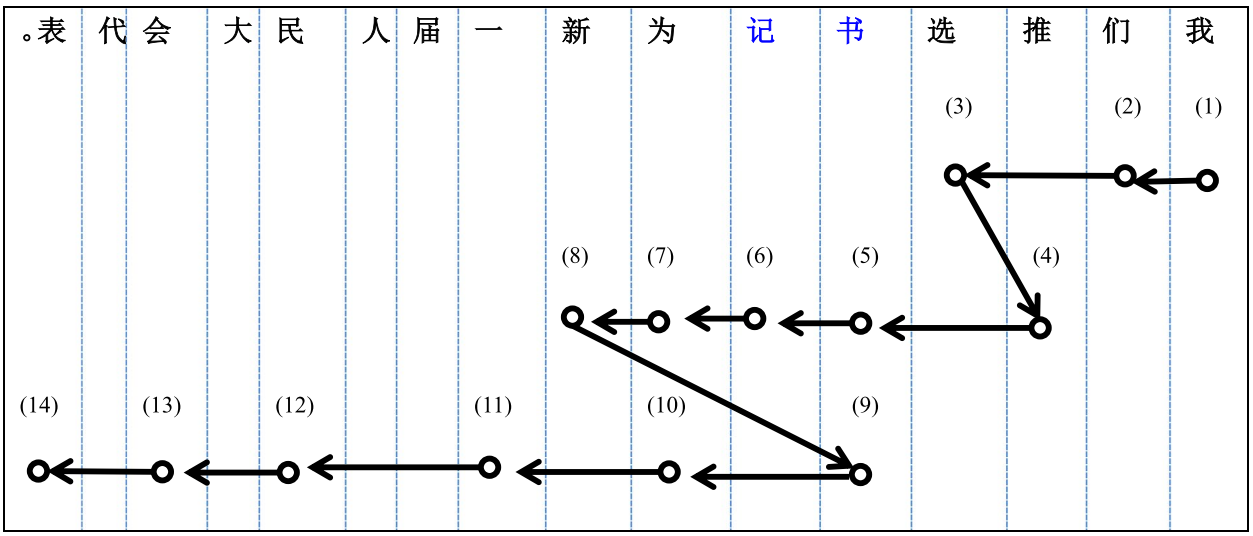

Fig. 4 The simulation trajectory diagram under familiar and opaque condition. Note: This trajectory diagram uses characters as interest areas, and interest areas are separated by blue lines. The dots represent fixation points, the arrows represent eye-saccade directions, and the numbers represent the fixation order. The entire trajectory is designed to simulate the eye movement trajectory of readers during right-to-left reading. For example, if readers have already fixed their gaze on the third fixation point (选) but do not fully understand the word meaning before the third fixation point (选), the reader's eye saccade will return to the fourth fixation point (推), and then continue reading until the eighth fixation point (新). While readers do not fully understand the word's meaning before the eighth fixation point (新), the reader's eye saccade will return to the ninth fixation point (书), and then continue reading until the fourteenth fixation point (表). The same applies below

which is consistent with the results of this study. Readers perform better in familiar formats and can more quickly segment sentences into the smallest semantic units and perform word recognition and comprehension [33]. At the same time, the results are consistent with previous studies. Whether it is Chinese sentence reading [11] or word processing [10], the format familiarity effect can be found: sentences reading are faster, lexical processing is faster, and reading efficiency is higher under familiar formats. In addition, the semantic transparency effect is consistent with previous studies [55, 59, 62]. Compared with opaque words, readers can understand transparent words faster and have higher reading efficiency. It is because the meaning of the morphemes of transparent words is consistent with the meaning of the whole word, while the meaning of the morphemes of opaque words is

inconsistent with the meaning of the whole word. When readers process the opaque words, semantic conflicts will occur, and more cognitive resources are needed to carry out the semantic access process, which requires longer reading time. The semantic transparency effect was found even in unfamiliar formats, but the format unfamiliarity delayed the appearance stage.

In comparison with the familiar format, the initial landing position was situated closer to the beginning of the word when reading efficiency was lower from right to left ($b=-0.206$, $SE=0.031$, $t=-4.69$, $p<0.001$), which was the unfamiliar format for Chinese readers [9–11]. This interference may be attributed to the cost of reading from right to left [23, 36]. A significant semantic transparency effect was observed on ILP ($b=-0.098$, $SE=0.031$, $t=-2.45$, $p=0.013$). The first fixation position for the transparent words was longer than that for the opaque words [8–10]. Similarly, for the skipping rate, the transparent target words were skipped more often than the opaque target words ($b=-0.035$, $SE=0.011$, $t=-3.191$, $p=0.002$). More skipping rate was also observed from left to right than from right to left ($b=$, $SE=0.011$, $t=-6.23$, $p<0.001$). The refixation probability was higher in the opaque condition ($b=0.096$, $SE=0.016$, $t=5.856$, $p<0.001$). RR was also higher from left to right than from right to left ($b=0.22$, $SE=0.016$, $t=14.000$, $p<0.001$). The main effect of format familiarity was significant ($b=-0.185$, $SE=0.039$, $t=-4.692$, $p<0.001$). Compared with the right-to-left and opaque words, there were shorter fixation duration, higher SP, lower RP, and shorter ALP for left-to-right and transparent words, consistent with assumptions and previous findings [9, 11, 19, 55, 62].

The most important findings were the interaction between semantic transparency and format familiarity. It is noteworthy that the interactions of the different indexes were inconsistent. The FFD and SFD are typically regarded as preliminary indicators; they are the reading duration of a single fixation or first fixation of the target word, and they can reflect the characteristics of the initial processing stage of word recognition in Chinese reading [11, 12, 33, 36, 47, 57]. The interaction between format familiarity and semantic transparency was significant in the early indexes. The interaction was statistically significant in the FFD ($b=-0.07$, $SE=0.029$, $t=-2.442$, $p=0.015$). From left to right, the FFD on the transparent words was shorter than that on the opaque words ($b=0.084$, $SE=0.021$, $t=4.055$, $p<0.001$). However, the results demonstrated no significant interaction between the transparent and opaque words for the right-to-left condition ($b=0.014$, $SE=0.020$, $t=0.694$, $p=0.488$). Moreover, a significant interaction was found with the SFD ($b=-0.073$, $SE=0.034$, $t=-2.155$, $p=0.032$). The SFD was deemed an effective indicator of the semantic stage in word recognition, exhibiting a pronounced influence of semantic transparency. There were significant differences between transparent and opaque words from left to right ($b=0.088$, $SE=0.023$, $t=3.879$, $p=0.000$). However, there were no significant differences between transparent and opaque words from right to left ($b=0.015$, $SE=0.025$, $t=0.584$, $p=0.559$). This finding is insightful and consistent with previous studies, which showed that format familiarity changed the word segmentation strategies [11]. Readers tend to conduct whole-word segmentation to obtain higher reading efficiency. However, readers lack reading experience in unfamiliar formats, and the difficulty of lexical processing increases. The segmentation strategies changed from word segmentation under a familiar format to character segmentation under an unfamiliar format, which slowed down the segmentation steps and reduced reading efficiency. Thus, there were significant differences between familiar and unfamiliar formats in eye movement indicators of transparent words. For opaque words, whether in familiar or unfamiliar formats, readers tend to use a character-segmentation strategy for the semantic conflict between morphemes and whole words, so there is little difference in reading performance on eye movement indicators. It is why the semantic-transparency effect appears in familiar rather than unfamiliar formats in early eye movement indicators (FFD and SFD), which mainly showed the change in segmentation strategy. In addition, the effect of format familiarity on transparent and opaque words also supports the above findings. From the interaction in the early stages of lexical processing, format familiarity significantly impacts word processing more than semantic transparency. In the FFD and SFD, the effect of format familiarity in the transparent words is significant (FFD: $b=0.117$, $SE=0.030$, $t=3.923$, $p<0.001$; SFD: $b=0.081$, $SE=0.023$, $t=3.503$, $p<0.001$). In contrast, it is not significant on opaque words (FFD: $b=0.037$, $SE=0.030$, $t=1.226$, $p=0.222$; SFD: $b=0.014$, $SE=0.022$, $t=0.607$, $p=0.545$).

In the later stage, transparent and opaque words complete the segmentation and semantic conflict stages, enter the semantic integration stage, and access lexical recognition. Thus, there is no significant difference between the segmentation strategy of transparent and opaque words. It is also an essential finding of this study, showing that readers are more flexible in choosing segmentation strategies when processing transparent and opaque words, supporting the combined access model [56, 63]. As assumed, except for those above two early time indexes, the interaction was not statistically significant in the remaining time indexes, typically representing the late and full-time stages of lexical processing. There

were no significant interactions between gazing duration ($b=$, $SE=0.045$, $t=-1.472$, $p=0.142$), regression duration ($b=-0.037$, $SE=0.062$, $t=-0.598$, $p=0.550$), or total time ($b=-0.030$, $SE=0.0670$, $t=-0.454$, $p=0.650$). The semantic-transparency effect always exists regardless of the familiar or unfamiliar format. In addition to the time indicators, for the position indicators, the skipping probability ($b=0.013$, $SE=0.022$, $t=0.584$, $p=0.559$), the refixation probability ($b=-0.017$, $SE=0.03$, $t=-0.554$, $p=0.580$), and the first fixation position ($b=-0.022$, $SE=0.078$, $t=-0.275$, $p=0.783$) were not significant in the interactions between semantic transparency and format familiarity. These results were consistent with the research assumption and supported the Chinese E-Z reader model.

Discussion

This study manipulated format familiarity and semantic transparency to investigate the processing mechanism of word segmentation and recognition, which found that the two processes were not completely indistinguishable. It supported the Chinese E-Z reader (CEZR) model [33].

The facilitation of format familiarity in the Chinese lexical processing

There are format-familiarity effects on Chinese reading and lexical processing. This study found that readers in familiar formats had better reading performance, higher reading efficiency, and shorter saccade distances. Reading under the unconventional format may result in a leftward bias, which can be attributed to the asymmetry of the visual field [48]. There is a left-side processing bias for Chinese reading, in which observers rely more heavily on information conveyed by the left side of stimuli than the right side, related to habitual format familiarity [10]. Furthermore, given that the directional eye movement of readers in daily reading and writing activities is not aligned with that in unfamiliar formats, readers will engage in a compensatory effect, whereby they sacrifice a degree of reading performance to compensate for the format unfamiliarity [9–11].

The latest Chinese E-Z reader (CEZR) model found that Chinese readers dynamically adjust their eye saccade based on relative familiarity to change the segmentation strategy, thereby achieving efficient word recognition [33]. The researchers used six models for simulation and found that the familiarity segmentation model performed best. However, CEZR does not explain the word recognition mechanism in Chinese reading in detail. The results of this study provide empirical evidence for this model. Format familiarity affects readers' word recognition in Chinese reading. There are two interacting processes here: the bottom-up process controlled by the oculomotor nerve and the top-down regulation of the reading experience. The reading direction controls the oculomotor nerve because the direction of eye movement from left to right is more familiar to readers. When an unfamiliar reading direction is presented, the eye movement process is more complicated, and readers need to use more cognitive resources to control it. Furthermore, reading experience, as a high-level cognitive factor, can affect visual processing, character processing, word recognition, and segmentation from top to bottom based on the Chinese E-Z reader model [33, 46]. We conclude that reading experience behind format familiarity affects the speed and efficiency of Chinese word recognition. Under familiar formats, readers read faster and more efficiently, and the eye saccade is farther. In an unfamiliar format, readers need more reading time, reading efficiency is reduced, and the reading distance becomes shorter.

The lexical representation behind semantic transparency

The second finding is the effect of semantic transparency in Chinese reading and its impact on word recognition. Readers have shorter fixation times, higher reading efficiency, and a fixation point closer to the word center when reading semantically transparent words. It is consistent with previous studies, which show that Chinese readers are more efficient at processing transparent words [13, 49, 52, 53, 55, 56, 62]. More importantly, the results support morpheme processing, whole word processing, or combined access model. The possibilities for the two formers are as follows. Suppose the research results supported that transparent and opaque words are processed entirely through morpheme processing. In that case, the difference between transparent and opaque words should not be significant because two-character words are all processed by morphemes, and the semantic conflict does not affect the morpheme processing, so the result can refute this view. The next question is whether the representation mechanism involves whole words or a combined processing of whole words and morphemes. The interaction between semantic transparency and format familiarity illustrates the combined representation. The early significant interaction indicated that format familiarity has a minor impact on opaque words. In the FFD, the difference between transparent words in familiar and unfamiliar formats is 34ms, but for opaque words, the difference is only 14ms. It is consistent with our research hypothesis because the segmentation strategy of transparent words changes from whole-word segmentation to character segmentation in familiar and unfamiliar formats. Simultaneously, it is also consistent with the mixed representation model. When the segmentation strategy of transparent words changes, readers tend to recognize morphemes, and access to whole words

is faster than access after competition between whole words and morphemes [29].

The combined access (CA) model has morpheme and whole-word representations [29]. Chinese word recognition is the interactive activation of morpheme representation and whole-word representation. The psychological dictionary's whole word representation and morpheme representation are at the same level. There is a filtering mechanism between them for cognitive resource allocation and selection. The morphemes and whole words of transparent words have the same meaning. Morpheme representation is activated first and diffused and integrated into whole word representation. The former can promote the latter. However, in recognizing opaque words, morpheme representation cannot diffuse and integrate into whole word representation. There is a conflict of meaning between morphemes and whole words. Morphemes will inhibit whole-word activation. Individuals must return to the filter, readjust cognitive resources, reduce whole-word inhibition, and increase morpheme inhibition to achieve whole-word activation and access [41]. Combined with the results, the segmentation strategy change has little effect on the later integration. However, it significantly impacts the early indexes (e.g., FFD), which represent the cognitive resource selection allocation and mainly affect the attention allocation stage before the filter. Combined with the meaning computation viewpoint, the difference in processing transparent and opaque compound words may lie in the integration stage of morpheme meaning and whole-word semantics [14, 20, 53]. Furthermore, the CEZR model showed that readers dynamically adjust eye movement trajectories according to relative familiarity, thereby efficiently recognizing words [33]. Format familiarity has a greater impact than semantic transparency in early word recognition processing.

In sum, the result supported the combined access model, which indicates that both morpheme representation and whole-word representation exist in the mental lexicon, and the recognition is the interactive activation between morpheme and whole word in Chinese [29, 41].

Mechanism between segmentation behind format familiarity and recognition behind semantic transparency

The most important result of this study is the interaction between format familiarity and semantic transparency, which provides empirical evidence for the Chinese E-Z reader model (CEZR). The results show that there is a significant interaction in the early indicators. In FFD and SFD, the semantic transparency effect is significant in familiar but not unfamiliar formats. It is consistent with the previous studies [10, 11]. Previous studies showed that format familiarity has a trade-off effect on Chinese word segmentation [11]. It means that format familiarity, an important factor affecting the word segmentation process, occupies a greater dominant position in early processing. As an important indicator of word recognition, the semantic transparency effect did not appear in the early indexes (FFD and SFD) under unfamiliar formats. There is a possibility that word segmentation and recognition are dynamically interactively adjusted. In the early lexical processing, the influencing factor of word segmentation may be dominant. Readers spend more cognitive resources adjusting word segmentation strategies, which may involve competition between parafoveal and foveal processing. Why is the transparency effect significant in familiar formats but disappears in unfamiliar formats? There may be explanations from three perspectives.

First, under an unfamiliar format, the lack of reading experience consumes prior readers' cognitive resources, making it impossible for them to show good reading performance for lexical processing, whether transparent or opaque words, making the main effect of semantic transparency insignificant in early processing. Therefore, format familiarity plays a more significant role in the early stages of lexical processing. Following the first perspective, the second perspective is that format familiarity leads to changes in word segmentation strategies. The extension of the word segmentation process makes readers enter vocabulary recognition more slowly. It is similar to the results of previous studies: word frequency and inter-word space were manipulated to examine Chinese reading and lexical processing. The results showed that inter-word space delayed the onset of the word frequency effect by 21ms [36]. This view is consistent with the latest CEZR and provides empirical research evidence [33]. It simulates the segmentation strategies and lexical processing in Chinese reading according to the unique characteristics of Chinese that are different from English by six models. According to CEZR, Chinese readers dynamically adapt their eye movement distance and segmentation strategies based on relative familiarity, thereby achieving efficient word recognition [33]. Therefore, readers had a longer eye saccade distance and adopted word segmentation strategies under a familiar format. However, the eye saccade distance was shorter in the unfamiliar format, and only character segmentation could be adopted. Under an unfamiliar format, readers may not be able to enter the word recognition in the early stage quickly, and there will be no semantic transparency effect as word recognition representation on early indicators (FFD and SFD). At the same time, this view has also been supported by physiological evidence. It is found that the independent brain area for Chinese segmentation is the left middle temporal gyrus [66]. In summary, the results of this study support the research

hypothesis. When the reader's fovea is fixated on a word n, the reader's parafovea may partially or entirely complete the segmentation of the next word n + 1 [34]. Therefore, segmentation may begin earlier than recognition, which is not indistinguishable [46]. Furthermore, this view supports the existence of an independent word segmentation stage in Chinese reading. It does not regard word segmentation as a concomitant word recognition in Chinese reading.

More importantly, this study is different from the integration model's view. The integrated model proposes that Chinese word segmentation and recognition are indistinguishable. Word segmentation is automatically completed when a reader recognizes a word successfully [25]. If true, the format familiarity and semantic transparency simultaneously affect the same processing stage, and there would be no priority effect on processing time. In the results, there would be significant interactions in all-time indexes. Then, the semantic transparency effect should exist in both familiar and unfamiliar formats because it is an indicator of word recognition, and there should also be a semantic transparency effect on early indicators. However, this was not the case. It is more likely that word segmentation starts earlier than word recognition. When word recognition starts, the parafovea may have already processed the segmentation of the next word while the reader is processing the previous word [32, 34]. Therefore, even if the two end at the same time, they do not start simultaneously, and it cannot be said that the two are entirely indistinguishable. Accordingly, the inconsistency of this interaction in different indicators (the interaction is significant in early indicators but not significant in late or sorting indicators) indicates that in Chinese reading, word segmentation and recognition are not entirely inseparable. This study provides empirical research evidence for the Chinese E-Z reader model.

Implications for reading strategies and second language education

This study provides suggestions for Chinese readers, including native speakers, second-language learners, and beginners. Even proficient native speakers are affected by unfamiliarity with the format. Thus, more attention should be given to the impact of unfamiliarity on readers' reading. When readers encounter ambiguous or unfamiliar texts, their cognitive difficulty should be considered, and more assistance, such as word segmentation clues or relevant training, should be provided to improve their reading ability in daily life and help them cope with various reading situations.

Regarding second-language learners, Chinese is different from phonetic writing. There is no word segmentation strategy to help it segment. It is challenging for readers to read Chinese sentences and requires a lot of cognitive resources. As the results show, even with skilled native Chinese speakers, their reading performance will also deteriorate when the format is unfamiliar. When second language learners want to read Chinese sentences, they first need to master basic Chinese words and form corresponding representations, so accumulating words is a basic skill required for Chinese reading. Then, readers need to master basic syntax. The continuous text can be segmented into the smallest semantic units to achieve word, sentence, and text comprehension only by understanding the structure and segmentation strategies. This step requires readers to master flexible segmentation strategies, as shown in the results of this study. When readers encounter ambiguous sentences or words, they must judge based on the context and ultimately select the most appropriate reading strategy. This step requires readers to have a basic reading volume and be familiar with Chinese, which can be achieved through reading training. Research shows that readers perform better when in a familiar context.

In daily teaching, teachers should adopt reading scenarios that readers are more familiar with, such as adding contextual information, providing bilingual comparisons, and increasing reading training to improve readers' familiarity. In addition, the semantic transparency effect also suggests that readers' reading performance will be worse in conflict situations. Two strategies can be used in the teaching process to help improve readers' abilities. One is to add conflicting semantic vocabulary in daily training. When second language learners encounter opaque words, they are encouraged to do more relevant training and be taught the correct discrimination skills. The other is to reduce semantic conflict when reading text is unfamiliar. Because the unfamiliarity of the text will increase the cognitive difficulty of readers, it will consume a lot of cognitive resources. At this time, if the text also contains many words with semantic conflict, readers may give up the reading task because of the high difficulty and cognitive overload. Finally, long-term training can improve reading ability. In sum, the neurophysiological evidence for word recognition in Chinese reading is still insufficient, and more empirical evidence can be expanded in the future. Subsequent studies can select second-language readers or children with insufficient Chinese reading experience to examine how they process different types of words in texts, such as real and fake. In different situations, readers' relative familiarity is different, and the eye-saccade distance may be adjusted accordingly; the segmentation strategy and recognition may change [33], which may provide new findings for the Chinese E-Z reader model.

Furthermore, this study can also be extended to other language speakers, such as Hebrew or Uyghur, whose reading direction is opposite to that of Chinese. When native speakers of Hebrew or Uyghur learn Chinese, the directional unfamiliarity and the difficulty of the segmentation make their learning more difficult. Therefore, we can help them learn Chinese by improving semantic transparency, reducing text unfamiliarity, and increasing reading training.

Conclusion

In summary, this study supports the following conclusions:

(1) Familiarity with the format affects the lexical processing in Chinese reading. Under the familiar format, adult readers exhibited a shorter fixation time, a higher skipping rate, and a reading position closer to the word's center. (2) The Semantic transparency of two-character words influences word recognition in Chinese reading. The fixation time for transparent words is shorter, the skipping rate is higher, and the reading position is closer to the center of the word. (3) Two-character word processing in Chinese reading is more inclined towards the combined access model (CA) [29]. Whether it is transparent or ambiguous word processing, lexical access results from the interaction between whole word processing and morpheme processing, and which one holds the dominant position depends on the reader's segmentation strategy. Under different formats, the reader would change the segmentation strategy of words to promote lexical processing. (4) The interaction between format familiarity and semantic transparency in the early stages is significant. There is a significant semantic transparency effect from left to right, and the format unfamiliarity delayed the appearance of the semantic transparency effect under an unfamiliar format. The delay may be due to a change in segmentation strategy. The inconsistency of early and late interactions indicates a partial separation of Chinese word segmentation and recognition. In conclusion, this study supports the Chinese E-Z reader model [33].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40359-025-02397-6.

Supplementary Material 1.

Acknowledgements
The author would like to thank all participants.

Authors' contributions
Mingjing Chen designed the study. Mingjing Chen collected the data. Mingjing Chen analyzed and interpreted the data. Mingjing Chen drafted the manuscript. Mingjing Chen, Li-chih Wang Sisi Liu, and Duo Liu revised the paper. Mingjing Chen, Li-chih Wang Sisi Liu, and Duo Liu proofread the paper. Mingjing Chen, Li-chih Wang Sisi Liu, and Duo Liu agreed to be accountable and verifed the submitted version.

Funding
This work was partially supported by the Multi-disciplinary Research Capacity Building Scheme Grant of The Education University of Hong Kong (Reference No. 1–32-04A29).

Data availability
The data will be made available upon the request from the author.

Declarations

Ethics approval and consent to participate
The ethical approval committee of Education University of Hong Kong approved this study and confirmed that the study has no side effects on the participants of the study. All experiments were performed in accordance with relevant guidelines and regulations with the Declaration of Helsinki. All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from the individual participated in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 2 October 2024 Accepted: 17 January 2025
Published online: 06 March 2025

References

1. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang. 2008;59(4):390–412.
2. Bai X, Liang F, Blythe HI, et al. Interword spacing effects on the acquisition of new vocabulary for readers of Chinese as a second language. J Res Reading. 2013;36:S4–17.
3. Bai X, Yan G, Liversedge SP, et al. Reading spaced and unspaced Chinese text: evidence from eye movements. J Exp Psychol Hum Percept Perform. 2008;34(5):1277.
4. Bates D. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823, 2014.
5. Beyersmann E, Coltheart M, Castles A. Parallel processing of whole words and morphemes in visual word recognition. Quarterly journal of experimental psychology. 2012;65(9):1798–819.
6. Cai Q, Brysbaert M. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. PLoS One. 2010;5(6):e10729.
7. Cai W, Zhang X, Wang X, et al. Time course of the integration of the morpho-semantics and the meaning of two-character Chinese compound words. Acta Psychol Sin. 2023;55(8):1207.
8. Chang M, Hao L, Zhao S, et al. Flexible parafoveal encoding of character order supports word predictability effects in Chinese reading: evidence from eye movements. Atten Percept Psychophys. 2020;82:2793–801.
9. Chen M J, Lu J M. The role of format familiarity and word frequency in Chinese reading. J Eye Mov Res. 2023;16(4): https://doi.org/10.16910/jemr.16.4.5.
10. Chen M, Wang Y, Zhao B, et al. The role of text familiarity in Chinese word segmentation and Chinese vocabulary recognition. Acta Psychol Sin. 2022;54(10):1151.
11. Chen M, Wang Y, Zhao B, et al. The trade-off between format familiarity and word-segmentation facilitation in Chinese reading. Front Psychol. 2021;12:602931.
12. Cui L, Wang J, Zhang Y, et al. Compound word frequency modifies the effect of character frequency in reading Chinese. Quarterly Journal of Experimental Psychology. 2021;74(4):610–33.
13. Dang M, Zhang R, Wang X, et al. The interaction between phonological and semantic processing in reading Chinese characters. Front Psychol. 2019;9:2748.
14. El-Bialy R, Gagné CL, Spalding TL. Processing of English compounds is sensitive to the constituents' semantic transparency. The Mental Lexicon. 2013;8(1):75–95.

15. Fan X, Reilly R. Reading development at the text level: an investigation of surprisal and embedding based text similarity effects on eye movements in Chinese early readers. J Mov Res. 2020;13(6). https://doi.org/10.16910/jemr.13.6.2.
16. Gu J, Li X, Liversedge SP. Character order processing in Chinese reading. J Exp Psychol Hum Percept Perform. 2015;41(1):127.
17. Han YJ, Huang S, Lee CY, et al. The modulation of semantic transparency on the recognition memory for two-character Chinese words. Mem Cognit. 2014;42:1315–24.
18. Huang L, Li X. Early, but not overwhelming: the effect of prior context on segmenting overlapping ambiguous strings when reading Chinese. Quarterly Journal of Experimental Psychology. 2020;73(9):1382–95.
19. Hyönä J, Cui L, Heikkilä TT, et al. Reading compound words in Finnish and Chinese: an eye-tracking study. J Mem Lang. 2024;134:104474.
20. Ji H, Gagné CL, Spalding TL. Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. J Mem Lang. 2011;65(4):406–30.
21. Kumar A, Netzel R, Burch M, et al. Visual multi-metric grouping of eye-tracking data. J Eye Mov Res. 2018;10(5). https://doi.org/10.16910/jemr.10.5.10.
22. Li T, McBride-Chang C. How character reading can be different from word reading in Chinese and why it matters for Chinese reading development. In: Reading development and difficulties in monolingual and bilingual Chinese children. 2014. p. 49–65.
23. Li XS, Liu PP, Ma GJ. Advances in cognitive mechanisms of word segmentation during Chinese reading. Advances in Psychological Science. 2011;19(4):459.
24. Li X, Gu J, Liu P, et al. The advantage of word-based processing in Chinese reading: evidence from eye movements. J Exp Psychol Learn Mem Cogn. 2013;39(3):879.
25. Li X, Pollatsek A. An integrated model of word processing and eye-movement control during Chinese reading. Psychol Rev. 2020;127(6):1139.
26. Li X, Rayner K, Cave KR. On the segmentation of Chinese words during reading. Cogn Psychol. 2009;58(4):525–52.
27. Li X, Zang C, Liversedge SP, et al. 16 The role of words in Chinese reading. Oxford Handb Read. 2015;232.
28. Liang F, Blythe HI, Bai X, et al. The role of character positional frequency on Chinese word learning during natural reading. PLoS One. 2017;12(11):e0187656.
29. Libben G, Gagné CL, Dressler WU. The representation and processing of compounds words. Word Knowl Word Usage. 2020;336.
30. Liu PD, Chung KKH, McBride-Chang C, et al. Holistic versus analytic processing: evidence for a different approach to processing of Chinese at the word and character levels in Chinese children. J Exp Child Psychol. 2010;107(4):466–78.
31. Liu P, Lu Q. The effects of spaces on word segmentation in Chinese reading: evidence from eye movements. J Res Reading. 2018;41(2):329–49.
32. Liu Y, Reichle ED, Li X. The effect of word frequency and parafoveal preview on saccade length during the reading of Chinese. J Exp Psychol Hum Percept Perform. 2016;42(7):1008.
33. Liu Y, Yu L, Reichle ED. Towards a model of eye-movement control in Chinese reading. Psychon Bull Rev. 2024:1–35.
34. Lv Y, Zhang L, Chen W, et al. The influence of foveal load on parafoveal processing of N+ 2 during Chinese reading. Vis Cogn. 2023;31(2):97–106.
35. Ma G, Li X, Rayner K. Word segmentation of overlapping ambiguous strings during Chinese reading. J Exp Psychol Hum Percept Perform. 2014;40(3):1046.
36. Ma G. Does interword spacing influence lexical processing in Chinese reading? Vis Cogn. 2017;25(7–8):815–24.
37. Ma MY, Chuang HC. How form and structure of Chinese characters affect eye movement control. J Eye Mov Res. 2015;8(3). https://doi.org/10.16910/jemr.8.3.3.
38. Mok LW. Word-superiority effect as a function of semantic transparency of Chinese bimorphemic compound words. Lang Cognit Process. 2009;24(7–8):1039–81.
39. Myers J. Words as basic lexical units in Chinese. Natl Chung Cheng Univ ms; 2017.
40. Phillips MH, Edelman JA. The dependence of visual scanning performance on saccade, fixation, and perceptual metrics. Vision Res. 2008;48(7):926–36.
41. Pollatsek A, Hyönä J, Bertram R. The role of morphological constituents in reading Finnish compound words. J Exp Psychol Hum Percept Perform. 2000;26(2):820.
42. Rayner K, Fischer MH, Pollatsek A. Unspaced text interferes with both word identification and eye movement control. Vision Res. 1998;38(8):1129–44.
43. Rayner K, Juhasz B, Ashby J, et al. Inhibition of saccade return in reading. Vision Res. 2003;43(9):1027–34.
44. Rayner K, Li X, Pollatsek A. Extending the E-Z reader model of eye movement control to Chinese readers. Cogn Sci. 2007;31(6):1021–33.
45. Rayner K, Schotter ER, Masson MEJ, et al. So much to read, so little time: how do we read, and can speed reading help? Psychological Science in the Public Interest. 2016;17(1):4–34.
46. Rayner K. Eye movements in reading: models and data. Journal of eye movement research. 2009;2(5):1.
47. Rayner K. The 35th Sir Frederick Bartlett lecture: eye movements and attention in reading, scene perception, and visual search. Quarterly journal of experimental psychology. 2009;62(8):1457–506.
48. Rinaldi L, Di Luca S, Henik A, et al. Reading direction shifts visuospatial attention: an interactive account of attentional biases. Acta Physiol (Oxf). 2014;151:98–105.
49. Shen M, Niu Z, Gao L, et al. Examining the extraction of parafoveal semantic information in Tibetan. PLoS One. 2023;18(4):e0281608.
50. Shen W, Li X, Pollatsek A. The processing of Chinese compound words with ambiguous morphemes in sentence context. Quarterly Journal of Experimental Psychology. 2018;71(1):131–9.
51. Strandberg A, Nilsson M, Östberg P, et al. Eye movements during reading and their relationship to reading assessment outcomes in Swedish elementary school children. J Eye Mov Res. 2022;15(4). https://doi.org/10.16910/jemr.15.4.3.
52. Tang M, Chan SD. 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. 2022;51(1):33–53.
53. Tsang YK, Chen HC. Activation of morphemic meanings in processing opaque words. Psychon Bull Rev. 2014;21:1281–6.
54. Wang J, Li L, Li S, et al. Effects of aging and text-stimulus quality on the word-frequency effect during Chinese reading. Psychol Aging. 2018;33(4):693.
55. Wang J, Yang J, Biemann C, et al. Mechanism of semantic processing of lexicalized and novel compound words: an eye movement study. J Exp Psychol Learn Mem Cogn. 2023;49(11):1812.
56. Wang S, Huang CR, Yao Y, et al. The effect of morphological structure on semantic transparency ratings. Language and Linguistics. 2019;20(2):225–55.
57. Xiong J, Yu L, Veldre A, et al. A multitask comparison of word-and character-frequency effects in Chinese reading. J Exp Psychol Learn Mem Cogn. 2023;49(5):793.
58. Yan M, Kliegl R, Richter EM, et al. Flexible saccade-target selection in Chinese reading. Quarterly Journal of Experimental Psychology. 2010;63(4):705–25.
59. Yan M, Zhou W, Shu H, et al. Lexical and sublexical semantic preview benefits in Chinese reading. J Exp Psychol Learn Mem Cogn. 2012;38(4):1069.
60. Yang H, Chen J, Spinelli G, et al. The impact of text orientation on form priming effects in four-character Chinese words. J Exp Psychol Learn Mem Cogn. 2019;45(8):1511.
61. Yang J, Wang S, Tong X, et al. Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. Read Writ. 2012;25:1031–52.
62. Yi W, DeKeyser R. Incidental learning of semantically transparent and opaque Chinese compounds from reading: an eye-tracking approach. System. 2022;107:102825.
63. Zang C, Liang F, Bai X, et al. Interword spacing and landing position effects during Chinese reading in children and adults. J Exp Psychol Hum Percept Perform. 2013;39(3):720.
64. Zang C, Fu Y, Bai X, et al. Investigating word length effects in Chinese reading. J Exp Psychol Hum Percept Perform. 2018;44(12):1831.
65. Zhang G, Yao P, Ma G, et al. The database of eye-movement measures on words in Chinese reading. Scientific Data. 2022;9(1):411.
66. Zhou W, Wang S, Yan M. Fixation-related fMRI analysis reveals the neural basis of natural reading of unspaced and spaced Chinese sentences. Cereb Cortex. 2023;33(19):10401–10.

Publisher's Note