Atten Percept Psychophys (2015) 77:272–281
DOI 10.3758/s13414-014-0749-5

The effects of character transposition within and across words in Chinese reading

Junjuan Gu · Xingshan Li

Published online: 20 August 2014

Abstract Given the lack of spaces between words in Chinese text, Chinese readers must parse these characters into words using their word knowledge. In this situation, are the characters belonging to a single word or to different words understood via different character-order encoding processes? In this study, we explored the effects of word boundaries in Chinese text on character-order encoding. We used four-character words (the one-word condition) and two two-character words (the two-word condition) as our targets. We embedded the target words into sentences and then manipulated the previews of the words using the boundary paradigm. The preview was identical to the target word (identity condition), had the two middle characters of the target word transposed (TC condition), or had two middle characters that were different from those in the target word (SC condition). Fixation durations on the target word in the TC condition were much longer than those in the identity condition for the two-word condition, but they were not significantly different for the one-word condition. Furthermore, for the one-word condition, gaze durations were longer in the SC than in the TC condition, whereas for the two-word condition, the difference between the TC and SC conditions was not significant. Word boundaries were found to affect the character-order encoding in Chinese reading, further suggesting the early occurrence of word segmentation.

Keywords Word boundaries · Character transposition · Parafoveal processing · Chinese reading

Letter position is important for word recognition in alphabetic writing systems (Gomez, Ratcliff, & Perea, 2008; Grainger & Heuven, 2003). The transposition of two letters within a word slows down the processing of that word or provides less priming to that particular word, suggesting the encoding of the letter order during the reading process (Johnson & Eisler, 2012; Rayner, White, Johnson, & Liversedge, 2006; White, Johnson, Liversedge, & Rayner, 2008). However, readers can still recognize the word even after two of the letters are transposed. As was shown in the famous "Cambridge" e-mail, readers could still read the message even after two of the letters in most of the words were transposed, and some readers even did not realize that letters had been transposed (see www.mrc-cbu.cam.ac.uk/personal/matt.davis/Cmabrigde/). Other studies have also shown that a transposed-letter nonword provides more priming to the target word than does substituting two letters at the same positions. In general, letter order is not strictly encoded during the reading process (Johnson & Dunne, 2012; Johnson & Eisler, 2012; Johnson, Perea, & Rayner, 2007; Rayner et al., 2006; White et al., 2008). Because there are spaces between words, letter position encoding in alphabetic writing systems is usually constrained within words, and no study has yet reported the possibility of letter position encoding between words. Chinese text has a unique structure that is formed by strings of equally spaced characters, with no spaces to separate the words. Therefore, Chinese readers must depend on lexical knowledge to parse these characters into words. However, it remains unclear whether character-order encoding can also be performed across words. By transposing two characters between words and within a word, we investigated whether character-order encoding was constrained by word boundaries and whether such encoding produced similar or dissimilar effects on the processing of that word.

A number of experiments have shown that a transposed-letter nonword (TL condition), created by transposing two adjacent letters of a word, is perceptually similar to the base word, much more so than a replacement-letter nonword (SL

J. Gu · X. Li (✉)
Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing, China
e-mail: lixs@psych.ac.cn

condition), in which two corresponding letters are replaced (i.e., the TL effect; Andrews, 1996; Bruner & O'Dowd, 1958; Chambers, 1979; Forster, Davis, Schoknecht, & Carter, 1987; Holmes & Ng, 1993; Kinoshita & Norris, 2009; Lee & Taft, 2009, 2011; O'Connor & Forster, 1981; Perea & Fraga, 2006; Perea & Lupker, 2003a, 2003b, 2004; Perea, Rosa, & Gómez, 2005; Schoonbaert & Grainger, 2004; Winskel, Perea, & Ratitamkul, 2012). Moreover, the letter order information is processed while the sentence is being read (Johnson & Eisler, 2012; Rayner et al., 2006; White et al., 2008). Thus, readers take more time when reading sentences with TL nonwords than when reading normal sentences, suggesting that the letter order is encoded under natural reading conditions.

Letter order encoding in parafoveal vision has also been studied through the eye-contingent display change technique known as the *boundary paradigm* (Rayner, 1975). In this paradigm, a preview appears at the target word position before the reader's eyes cross the invisible boundary on the left of the preview. When the eyes cross the boundary, the preview immediately changes to the target word. The preview may be a transposed nonword (TL condition) or a substituted nonword (SL condition), or it may also be identical to the target word (identity condition). Fixations on the target word are longer in the SL condition than in the TL condition, which in turn are longer than those in the identity condition (Johnson, 2007; Johnson & Dunne, 2012; Johnson et al., 2007; Perea, Nakatani, & van Leeuwen, 2011). These results indicate that the TL effect also occurs in parafoveal vision. The finding that the fixation on the target word is longer in the TL condition than in the identity condition shows that letter order information can be processed in parafoveal vision. Furthermore, the fact that the fixations are shorter in the TL than in the SL condition indicates that parafoveal letter order is not strictly encoded. Otherwise, there would be no difference between the TL and SL conditions, because two letters are in inaccurate positions for both types of nonwords. Therefore, the findings of letter order encoding in parafoveal vision are similar to those in foveal vision.

Some studies have investigated the effect of letter transpositions on the morpheme boundaries (e.g., *susnhine* for *sunshine*) of compound words in English and other alphabetic languages. Perea and Carreiras (2006) found that the TL effect was unaffected by morphological boundaries in Basque compounds. Additionally, some English studies have shown identical TL effects between the within- and across-morpheme letter transpositions in English-suffixed words in lexical decision tasks (Beyersmann, Coltheart, & Castles, 2012; Beyersmann, McCormick, & Rastle, 2013; Rueckl & Rimzhim, 2011). Transposition across morpheme boundaries did not modulate TL effects in two eye movement tasks and one masked-priming lexical decision task (Masserang & Pollatsek, 2012). Masserang and Pollatsek proposed that prefixes and suffixes are frequent word-beginning and -ending bigrams and trigrams and that English readers can easily correct minor errors in position coding at the early stages of processing English words.

Not all studies of this type, however, have generated similar results. In an English masked-priming naming task, transpositions across morpheme boundaries produced smaller TL effects than did those within monomorphemic words (Christianson, Johnson, & Rayner, 2005). The authors' findings indicate that across-morpheme transpositions are more disruptive to word naming than are within-morpheme transpositions. Letter transpositions within the stems of Spanish compound words (affixed words) can yield significant TL priming effects, and no TL effect is observed across the morpheme boundaries of either prefixed or suffixed words (Duñabeitia, Perea, & Carreiras, 2007). These findings affirm the sensitivity of letter position encoding to morphological boundaries in word identification. Across-morpheme transposition is more disruptive to word recognition than is within-morpheme transposition. Moreover, these results also suggest that morphological decomposition occurs at a very early stage in word processing, which may be before letter-position encoding.

Some researchers have reported that language systems with varying morphological processing systems, such as Spanish and English, may produce different patterns (Frost, 2009; Frost, Kugler, Deutsch, & Forster, 2005). Spanish has a significantly richer morphological variety and productivity than does English. Morphological parsing plays a central role in languages with morphologically rich structures (Frost, 2009). Therefore, languages such as Spanish tend to rely heavily on the precise positional encoding of the letters that constitute morpheme boundaries, which explains the disappearance of TL priming effects when the letters are transposed across morphemes. To investigate this issue, Sánchez-Gutiérrez and Rastle (2013) selected native Spanish and English speakers as participants and conducted two parallel masked-priming lexical decision experiments on the two languages using Spanish–English cognates. They found that the TL effects in both Spanish and English were not modulated by the position of the transposed letter in the prime stimulus. The TL effects of both experiments were also equal when the letters were transposed within the stem and across a morpheme boundary. Thus, they concluded that the two languages do not differ in terms of orthographic information coding and that the TL effect is not affected by the position of the transposed letters relative to the morpheme boundary (Sánchez-Gutiérrez & Rastle, 2013). To summarize, whether morpheme boundaries modulate TL effects may be related to differences in language properties or to differences across tasks, and this question appears to be controversial.

The effects of word boundaries on character-order encoding in Chinese reading have attracted much interest, given the unique properties of these texts. A Chinese text comprises equally spaced characters. One, two, three, four, or more characters can constitute a word, although successive characters are separated by equal-sized small spaces. These characters have highly complex structures. A Chinese character can comprise one or more radicals, with each radical consisting of one or more strokes, whereas a word may be composed of a single character. Except for punctuation marks, successive words in Chinese texts are not separated by spaces. However, the psychological realities of these words and their importance in Chinese reading have been reported in previous work (Bai, Yan, Liversedge, Zang, & Rayner, 2008; Cheng, 1981; Li, Gu, Liu, & Rayner, 2013; Li, Rayner, & Cave, 2009; Rayner, Li, Juhasz, & Yan, 2005; Rayner, Li, & Pollatsek, 2007; Yan, Tian, Bai, & Rayner, 2006). Therefore, Chinese readers must determine word boundaries by parsing strings of characters into words using their word knowledge (Li et al., 2013; Li et al., 2009). Therefore, character-order encoding in relation to word segmentation is critical in Chinese reading. If word segmentation occurs at an earlier period and if character encoding is constrained within a word, the transpositions of characters within a word versus between words might produce different effects on word processing. Previous studies have shown that word segmentation occurs at a very early stage, so that word boundaries affect character processing and attention deployment (Li & Ma, 2012; Li & Pollatsek, 2011; Li et al., 2009). Upon presentation of the stimuli, recognition accuracy decreases and probe detection reaction times increase rapidly from left to right at word boundary positions. If word segmentation happens at an early stage and word boundaries limit word processing, the transposition of characters between words should affect word recognition to a greater degree than the transposition of characters within a word. In comparison, transpositions of two characters between words and within a word should not produce different effects on word recognition if word segmentation occurs later than does character-order encoding.

We investigated whether word boundaries can affect character-order encoding in Chinese reading. We employed the boundary paradigm (Rayner, 1975) and compared fixation durations on the target, which was either one four-character word (one-word condition; e.g., 剑拔弩张) or two two-character words (two-word condition; e.g., 庄严肃穆). Three types of stimuli—namely, an identity condition (剑拔弩张), a transposed-character (TC) condition (e.g., 剑弩拔张), and a substituted-character (SC) condition (e.g., 剑昙耐张)—were used as preview stimuli. For the one-word condition, the swapping of the middle two characters was considered a within-word transposition. For the two-word condition, the swapping of the middle two characters was considered an across-word transposition. We measured the fixation durations on the target region.

To test whether the character order is strictly encoded in parafoveal vision, we then compared the fixation durations on the target region between the identity and TC conditions and between the TC and SC conditions, because the preview in the TC condition and the preview in the identity condition differed only in whether the character order was identical to or different from the target. Thus, if character order is not at all encoded in parafoveal vision, the previews in the TC and identical conditions should affect the processing of the target region in the same way, and thus we expected the fixation durations on the target region to be similar for these conditions. However, if character order is encoded in parafoveal vision, the identical preview should benefit the processing of the target region more than would the TC preview. Thus, in that case we would expect the fixation duration in the TC condition to be longer than that in the identity condition. Furthermore, if character-order information is not strictly encoded, the preview of the TC nonword should facilitate the processing of the target region more than would the preview in the SC condition. Thus, we would expect fixation durations to be shorter in the TC condition than in the SC condition. However, if character order is encoded strictly, then the TC preview should not facilitate the target region processing to a greater extent than the SC condition. Thus, we expected that the fixation durations on the target region would be not different between the SC and TC conditions.

As we stated above, if word segmentation occurs earlier than does character-order encoding, then we would expect that word boundaries would affect character-order encoding. The transposition of characters between words should affect word recognition to a greater degree than would the transposition of characters within a word. If this is the case, then the previews that involved an across-word transposition and the previews that involved a within-word transposition would be expected to produce varying results. The fixation durations in the identity and TC conditions should be similar for within-word transpositions (one-word condition), whereas the fixation durations in the identity condition should be shorter than those in the TC condition for across-word transpositions (two-word condition). The fixation duration in the TC condition should be shorter than that in the SC condition in the one-word condition, whereas in the two-word condition, there should be no difference between the TC and SC conditions. Therefore, an interaction between word type and preview condition would be expected. However, if character-order encoding is not affected by the word boundary, similar fixation duration patterns would be expected for across-word and within-word transpositions. In this case, the interaction between word type and preview condition would not be expected.

Method

Participants

A total of 30 native Chinese speakers (average age: 22.2 years), who were undergraduates or postgraduates from universities near the Institute of Psychology, Chinese Academy of Sciences, participated in the study. They were paid 25 Yuan (approximately US $4) to participate in the experiment. All of the participants had normal or corrected-to-normal vision, and all were unaware of the purpose of the experiment.

Apparatus

Eye movements were recorded using an SR EyeLink 2000 tracker, which had a resolution of approximately 30′ of arc. Participants read the target sentences (printed horizontally from left to right) on a 21-in. CRT monitor (SONY Multiscan G520) connected to a Dell computer. The eyetracking system sampled at 1000 Hz and provided eye movement data for further analysis using another PC. Participants rested their chins on a chinrest in order to minimize head movements during the experimental trials. Viewing was binocular, but eye movement data were collected only from the right eye. The refresh rate of the CRT monitor was 150 Hz, and the resolution was 1,024 × 768. Participants were seated 58 cm from the video monitor; at this distance, one character subtended 0.8° of visual angle.

Materials and design

Participants read 78 experimental sentences, with each frame containing two types of target words—namely, one of 78 four-character words (one-word condition) or 78 pairs of two-character words (two-word condition). All of the words were listed as words in a dictionary, and all of the nonword stimuli were not in this dictionary (Lexicon of Common Words in Contemporary Chinese Research Team, 2008). For the two-word condition, the two words had some semantic relationship,so that they could constitute phrases (e.g., 庄严肃穆, meaning "solemn and serene"). The target words were embedded in a single-line sentence no more than 31 characters long (ranging from 24 to 31 characters). The target words were in the middle of the sentence, such that the distance was at least seven characters away from both the beginning and the end of the sentence. Additionally, nine sentences were presented for participants to practice on before the formal experiment.

Each target word had three parafoveal preview conditions. In the identity condition, the preview was identical to the target word (e.g., 面黄肌瘦 as the preview of 面黄肌瘦, "emaciation with sallow complexion"). In the TC nonword condition, the preview was a transposition of the middle two characters (e.g., 面肌黄瘦 as the preview of 面黄肌瘦) of the target word. In the SC nonword condition, the preview differed from the target word at the middle two characters (e.g., 面朴密瘦 as the preview of 面黄肌瘦). The preview stimuli in the TC and SC conditions were nonwords. Furthermore, in both the TC and the SC conditions, no two-character word was produced in any of the four-character target positions. The first two characters were never the beginning of another word. The frequency and number of strokes of the second characters of the TC nonwords (frequency: M = 1,112.43 occurrences per million, SD = 1,635.96; number of strokes: M = 8.70, SD = 3.14) did not differ from those of the SC nonwords (frequency: M = 1,102.55 occurrences per million, SD = 1,589.50; number of strokes: M = 8.78, SD = 3.01; $ps > .1$). The frequency and number of strokes of the third characters of the TC nonwords (frequency: M = 898.76 occurrences per million, SD = 1,511.67; number of strokes: M = 9.04, SD = 3.00) did not differ from those of the SC nonwords (frequency: M = 890.09 occurrences per million, SD = 1,494.24; number of strokes: M = 9.08, SD = 2.90; $ps > .1$). We matched the character structure, which refers to the position relation of the radicals in a character. The two characters in the TC and SC conditions had similar structures (i.e., top to bottom and left to right). Given that the word frequencies were negatively correlated with the word length, we were unable to effectively control the word frequencies in the two word type conditions.[1] The frequencies of the four-character words ranged from 0.04 occurrences per million to 0.87 occurrences per million (M = 0.36, SD = 0.21). The frequency of the first word in the two-word condition was 36.31 (SD = 76.20) occurrences per million, and that of the second word was 19.79 (SD = 46.41) occurrences per million.

The experiment had a 2 (word type: one-word vs. two-word condition) × 3 (parafoveal preview condition: identity condition, TC condition, SC condition) design. We created six versions for each sentence frame. Each participant was asked to read only one version of each sentence frame. A sample sentence frame is shown in Table 1.

Procedure

When the participants arrived at the laboratory, they were asked to read a brief description of the experimental procedure and the apparatus. Then they were given verbal instructions about the task. The eyetracker was calibrated at the beginning of the experiment and then recalibrated as needed. For calibration and validation,

[1] Luke and Christianson (2013) investigated the effect of frequency on morphological processing across the time course of lexical access using the transposed-letter paradigm. They found that frequency did not affect early morphological processing. Furthermore, when whole-word frequency increased, transposition across the morpheme boundary became less disruptive. To exclude the possible influence of word frequency in our study, we used the lme4 package to build linear mixed models including word frequency as a control factor. For the one-word condition, the word frequency was counted from an online corpus (http://ccl.pku.edu.cn:8080/ccl_corpus/); for the two-word condition, word frequency was counted as the collocational frequency for the pair in the same corpus. In this model, word frequency did not affect the TC effects that we reported in the main text.

Table 1 Sample experimental sentence

Word Type	Display	Example
One word	Identity	这个有趣的问题立刻使几天来剑拔弩张的气氛变得轻松了。
	TC	这个有趣的问题立刻使几天来剑弩拔张的气氛变得轻松了。
	SC	这个有趣的问题立刻使几天来剑昙耐张的气氛变得轻松了。
Two words	Identity	这个有趣的问题立刻使几天来庄严肃穆的气氛变得轻松了。
	TC	这个有趣的问题立刻使几天来庄肃严穆的气氛变得轻松了。
	SC	这个有趣的问题立刻使几天来庄弯客穆的气氛变得轻松了。

TC, transposed-character; SC, substituted-character

participants were asked to look at a dot shown at each of three locations horizontally arranged at the center of the display in random order. The maximum error permitted for validation throughout the experiments was 0.5°. After validation, participants were asked to read nine practice sentences to familiarize themselves with the procedure.

Experimental sentences were presented randomly and one at a time in the center row of the monitor. Each trial began with a drift check procedure, during which the participant fixated on a circle located at the center of the monitor. After the drift check, a white square box (1° × 1°) appeared on the monitor at the location corresponding to the area where the first character of the sentence would appear. Once the eyetracker detected that the participant was looking at the box, the sentence was shown (see Fig. 1). When the participant's eyes crossed an invisible boundary located just to the left of the preview word, the preview stimulus was modified into the target word. Participants did not notice any change, because the modification occurred while their eyes were moving. The sentence remained on the screen until the participants had finished reading it. Participants were told to read silently and at a normal pace and to press a button on the response box when they had finished reading the sentence. A total of 52 filler items were intermixed with the 78 experimental items, and the experimental procedure was repeated until all sentences had been read. The stimuli did not change when the participants read the filler sentences. Each participant read the 130 sentences in a random order. For the experimental items, participants saw only one condition with each sentence frame and saw equal numbers of each type of target. Participants were required to answer comprehension questions after 30% of the sentences to ensure that they were reading the sentences carefully. These tested sentences were the same for each participant and included some filler sentences and some experimental sentences. Participants pressed a button on the response box in order to answer multiple-choice questions. The entire experimental procedure took approximately 30 min.

Normative data

The experimental sentences were evaluated by another group of participants, to ensure that each target word was appropriate within the context of its sentence. Ten participants were recruited to judge how well each target word matched the given sentence frame on a scale of 1 (*not natural at all*) to 7 (*very natural*). All of the target words were rated as being natural within their respective sentence frames (overall: min = 5.00, max = 6.80, M = 6.11, SD = 0.45; one-word items: min = 5.00, max = 6.80, M = 6.02, SD = 0.46; two-word items: min = 5.10, max = 6.80, M = 6.21, SD = 0.41). These participants did not participate in the eyetracking section of the experiment.

To ensure that the target words were equally predictable for all of the conditions, we structured the sentences in such a way that the target words were not predictable from their previous contexts. To accomplish this, ten participants were given the first part of the experimental sentence containing the sample word (up to the target word) and were then asked to provide the next word in the sentence (i.e., to predict the target word). The predictability of all items was close to zero, indicating that the target words were not predictable from their preceding contexts.

Results

The comprehension accuracy of the questions ranged from .92 to 1.00, with a mean of .97, suggesting that the participants understood the sentences well.

We measured the first fixation and gaze durations on the target region. The target regions were four characters long for both conditions. For the one-word condition, the target region included the four-character word. For the two-word condition, the target included the two two-character words. The *first fixation duration* referred to the amount of time spent on the initial fixation of the target, regardless of whether one or more than one fixation occurred. The *gaze duration* was the sum of the fixation durations on the target word before the reader left that target.

(1) 她们在公园里看到一个|面肌黄瘦的小男孩正坐在长椅上发呆。
*

(2) 她们在公园里看到一个|面黄肌瘦的小男孩正坐在长椅上发呆。
*

Fig. 1 Example sentence using the boundary paradigm. In line 1, a transposed-character (TC) nonword preview (面肌黄瘦, underlined here) was initially displayed in the target location. When the reader's eyes crossed the invisible boundary location (|) just to the left of the target word, the preview was replaced by the actual target word (面黄肌瘦, "emaciation," also underlined) shown in line 2. The asterisks represent the fixation locations. The English translation of this sentence is "They saw an emaciated boy staring blankly on a bench in the park"

Trials were eliminated from the data analysis if one or more blinks occurred when the eyes fixated on the pretarget character, target word, or posttarget character, or when tracker loss occurred during a trial (Johnson et al., 2007). Consistent with most eye movement research (Rayner, 1998), extremely short (<80-ms) isolated fixations and extremely long (>1,000-ms) fixations were excluded from the data set prior to analysis. In sum, 2.6% of the data were eliminated. The means of the first fixation and gaze durations for each of the three parafoveal preview conditions using the two types of words are shown in Table 2.

For each of the two eye movement measures, a 2 (word type: one-word or two-word condition) × 3 (parafoveal preview condition: identity condition, TC condition, SC condition) analysis of variance was conducted. The error variance was calculated over participants (F_1) and over items (F_2). In addition, we ran planned comparisons to compare the fixation durations between the TC and identity conditions and between the TC and SC conditions.

First fixation durations showed a significant main effect in the parafoveal preview condition, $F_1(2, 58) = 8.36$, $MSE = 674$, $p = .001$, $\eta_p^2 = .22$; $F_2(2, 154) = 7.93$, $MSE = 2{,}019$, $p = .001$, $\eta_p^2 = .09$. First fixation durations were longer in the TC condition ($M = 289.32$ ms, $SE = 7.94$ ms) than in the identity condition ($M = 279.40$ ms, $SE = 7.50$ ms), $F_1(1, 29)= 6.58$, $MSE = 896$, $p = .016$, $\eta_p^2 = .19$; $F_2(1, 77) = 3.64$, $MSE = 4{,}285$, $p = .06$, $\eta_p^2 = .05$. Such a difference suggested that character-order information was encoded from the parafovea during reading. We also found a trend for first fixation durations in the SC condition ($M = 298.77$ ms, $SE = 7.66$ ms) to be longer than those in the TC condition ($M = 289.32$ ms, $SE = 7.94$ ms), $F_1(1, 29) = 3.50$, $MSE = 1{,}533$, $p = .072$, $\eta_p^2 = .11$; $F_2(1, 77) = 3.43$, $MSE = 4{,}780$, $p = .068$, $\eta_p^2 = .04$. The effect of word type was not significant, Fs < 1. The interaction between word type and parafoveal preview condition was also not significant, $F_1(2, 58) = 2.50$, $MSE = 662$, $p = .091$, $\eta_p^2 = .08$; $F_2(1, 77) = 2.10$, $MSE = 2{,}046$, $p = .126$, $\eta_p^2 = .03$.

Gaze durations were affected by the parafoveal preview condition, $F_1(2, 58) = 8.84$, $MSE = 4{,}685$, $p < .001$, $\eta_p^2 = .23$; $F_2(2, 154) = 8.96$, $MSE = 12{,}972$, $p < .001$, $\eta_p^2 = .10$. Gaze durations in the TC condition ($M = 517.63$ ms, $SE = 25.97$ ms) were longer than those in the identity condition ($M = 483.72$ ms, $SE = 21.83$ ms), $F_1(1, 29) = 8.23$, $MSE = 8{,}388$, $p = .008$, $\eta_p^2 = .22$; $F_2(1, 77) = 7.60$, $MSE = 28{,}325$, $p = .007$, $\eta_p^2 = .09$. However, the difference between the SC ($M = 535.43$ ms, $SE = 22.54$ ms) and TC ($M = 517.63$ ms, $SE = 25.97$ ms) conditions was not significant, $F_1(1, 29) = 1.52$, $p = .228$; $F_2(1, 77) = 1.66$, $p = .202$. Gaze durations in the two-word condition ($M = 530.18$ ms, $SE = 25.64$ ms) were longer than those in the one-word condition ($M = 494.34$ ms, $SE = 20.25$ ms), $F_1(1, 29) = 9.77$, $MSE = 5{,}916$, $p = .004$, $\eta_p^2 = .25$; $F_2(1, 77) = 7.09$, $MSE = 27{,}269$, $p = .009$, $\eta_p^2 = .08$. The interaction between word type and the parafoveal preview condition was significant, $F_1(2, 58) = 4.12$, $MSE = 3{,}154$, $p = .021$, $\eta_p^2 = .12$; $F_2(2, 154) = 3.16$, $MSE = 10{,}978$, $p = .045$, $\eta_p^2 = .04$. For the one-word condition, gaze durations in the identity ($M = 478.53$ ms, $SE = 23.38$ ms) and TC ($M = 483.60$ ms, $SE = 21.28$ ms) conditions were not significantly different, Fs < 1. Gaze durations were longer in the SC condition ($M = 520.90$ ms, $SE = 21.46$ ms) than in the TC condition, $F_1(1, 29) = 5.78$, $MSE = 7{,}226$, $p = .023$, $\eta_p^2 = .17$; $F_2(1, 77) = 3.66$, $MSE = 25{,}227$, $p = .059$, $\eta_p^2 = .05$. For the two-word condition, gaze durations were longer in the TC condition ($M = 551.67$ ms, $SE = 31.98$ ms) than in the identity condition ($M = 488.90$ ms, $SE = 23.65$ ms), $F_1(1, 29) = 14.52$, $MSE = 8{,}141$, $p = .001$, $\eta_p^2 = .33$; $F_2(1, 77) = 13.77$, $MSE = 25{,}177$, $p < .001$, $\eta_p^2 = .15$. However, the difference between the TC and SC ($M = 549.97$ ms, $SE = 26.14$ ms) conditions was not significant, Fs < 1. These results suggest that character transpositions that occurred across word boundaries severely disrupted word identification relative to the within-word transpositions in the one-word condition.

Table 2 Mean fixation durations (and *SE*s) by word type (one-word, two-word) and preview condition (identity, transposed-character [TC], substituted-character [SC]) in our experiment

	First Fixation		Gaze Duration	
	One-Word	Two-Word	One-Word	Two-Word
Identity	286 (9)	273 (7)	479 (23)	489 (24)
TC	289 (7)	289 (10)	484 (21)	552 (32)
SC	295 (8)	303 (9)	521 (21)	550 (26)

All durations are in milliseconds.

Discussion

In the present study, we explored whether character-order encoding was affected by word boundaries in Chinese reading. Transposing two characters across word boundaries and character transpositions within a word resulted in different patterns of results. Gaze durations in the TC condition were much longer than those in the identity condition for the two-word condition, but they were not significantly different for the one-word condition. Furthermore, for the one-word condition, the gaze duration was longer in the SC condition than in the TC condition, whereas for the two-word condition, the difference between the TC and SC conditions was not significant. These results suggested that character-order encoding patterns were different between the one-word and two-word targets. TC nonwords were more similar to the base words in the one-word condition than was true for the two-word targets. Hence, character-order encoding was more disruptive for across-word than for within-word transpositions. In other words, the word boundaries affected character-order encoding in word identification during Chinese reading.

As we outlined in the introduction, native Chinese readers depend on word knowledge to determine word boundaries while reading Chinese text, because no spaces separate words (Bai et al., 2008; Cheng, 1981; Rayner, Li, Juhasz, & Yan, 2005; Rayner, Li, & Pollatsek, 2007; Yan, Tian, Bai, & Rayner, 2006). For Chinese words with two or more characters, encoding the identities and orders of these characters is indispensable in word recognition. The results of the present study showed that character-order encoding was sensitive to word boundaries, thus suggesting that word segmentation may occur early in word processing and that this may co-occur with, or occur earlier than, character-order encoding. In all other cases, character transpositions across word boundaries should not have had an effect on word recognition.

The results also showed that orthographic encoding interacted with word boundaries in the present study. Our results are consistent with the predictions of a formal model (see Fig. 2) proposed by Li et al. (2009).[2] According to that model, Chinese word segmentation and word recognition are unified processes. The model assumes that Chinese word recognition is an interactive process involving many nodes at multiple levels (i.e., a feature level, a character level, and a word level). Characters in the perception span are processed in parallel (with the constraint of visual acuity) at the character level. The activation of each unit containing a visible character feeds forward to the word recognition level, which activates the word unit, and when the activation of a word unit reaches a certain level, it feeds activation back to the characters belonging to the activated word. Hence, the characters belonging to the activated word will become increasingly activated more quickly than will other characters. In this way, the representations at the word level compete with each other until a single word unit wins the competition. At that time, the word is recognized and segmentation occurs. Characters are processed in parallel at the character level, but only a single word prevails at the word level. Moreover, the model assumes that the words on the left have the advantage during competition, so that the words on the left are processed sooner than are the words on the right. Because that model assumes a strict character-order encoding, distorting character order will result in failure to activate the related word at the word level. Thus, previewing a TC nonword should not benefit the processing

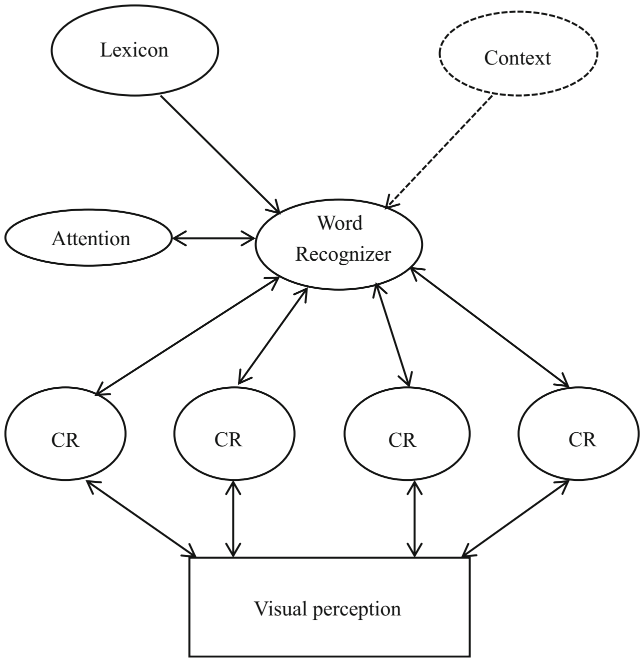

Fig. 2 Framework of a word segmentation and recognition model. "CR" refers to the character recognizer

of the target stimuli in any situation. This is not consistent with our observation that fixation durations were shorter when participants previewed a TC nonword than when they previewed an SC nonword in the one-word condition.

Li et al.'s (2009) word segmentation model could be improved to explain the results by introducing a flexible character-position encoding assumption. Some alphabetic letter-position encoding models do not assume a strict letter-order encoding. For example, the overlap model assumes that strings that are presented briefly have distributions over letter positions (Gomez et al., 2008). Each letter is assumed, at least initially, to be associated with more than one position, and the degree to which each letter is associated with different positions occurs as a function of a normal distribution centered about the letter's actual position. When adopting this type of flexible character-order encoding hypothesis, Li et al.'s (2009) improved word segmentation model could explain the different preview effects of across-word and within-word transpositions. In the one-word condition, the preview could still activate the base word even when the middle two characters were transposed. Thus, the reading time for the target word was much shorter in the TC condition than in the SC condition when the base word was active. However, because the character order was not accurate, the activation of the base word was not as high as to the preview of the identical preview condition. Thus, reading times were longer in the TC condition than in the identical condition. For the two-word condition, things were different. Because the visual acuity of character perception decreases from the center of the fovea, perception efficiency decreases from left to right in the parafoveal

[2] Taft et al., (1999) also proposed a model of Chinese word processing. They provided a theoretical framework that assumes that lexical memory is viewed as a hierarchy of levels. The lexical processing system includes orthographic, phonological, and semantic subsystems. However, only the model proposed by Li et al. (2009) focuses on word segmentation during Chinese reading.

region. Thus, the third character in that region would be perceived to a lesser degree than the second character, if it were perceived at all. Thus, the first two characters in the TC nonword preview would activate two different words, whereas the third character would be less likely to contribute to the activation of the base word. Thus, the TC nonword preview could not benefit the processing of the target word, resulting in reading times on the target word similar to those with the SC nonword preview.

The findings of our study are consistent with the results of some studies stating that transposing letters across morpheme boundaries and transposing letters within a morpheme in an alphabetic writing system have varying effects on word processing (Christianson et al., 2005; Duñabeitia et al., 2007). However, the present findings are inconsistent with other studies that have used lexical decision tasks to explore letter-position encoding problems for affixed words in English (Beyersmann et al., 2012; Beyersmann et al., 2013; Masserang & Pollatsek, 2012; Rueckl & Rimzhim, 2011). In those studies, equivalent transposed-letter effects were found when letters were transposed within the same morpheme and across morpheme boundaries. As has been suggested by some authors, two or three factors might have contributed to these differences. First, the dissociation may have been caused by the difference between bound derivational morphemes and free morphemes. Affixes are frequent word-beginning and -ending bigrams and trigrams, and native readers could overwrite minor letter-position errors at the early stages of word recognition. However, for compound words involving two free morphemes, the case might be different. Free morphemes can function independently as words and can appear with other lexemes. Free morphemes are also less frequent than affixes in compound words. Therefore, letter-position information about the morpheme boundaries of compound words might be more important for free morphemes than for affixes. Second, this dissociation could be explained from the perspective of processing stages. Morphological decomposition is necessary in the identification of multimorphemic words. Rastle, Davis, and New (2004) found that morphological decomposition occurs very early during word recognition. If morphological decomposition occurs at the very early stages of visual word recognition, before or coinciding with other low-level processes such as letter-position encoding, then TL effects should decrease or vanish when two letters across morpheme boundaries are transposed (Christianson et al., 2005; Duñabeitia et al., 2007; Perea & Carreiras, 2006). Conversely, if morphological decomposition occurs at the late stages of visual word recognition, early orthography effects could no longer interact with morphological effects.

In the present study, the argument that word boundaries affect character-order encoding in Chinese reading was mainly supported by the significant interaction between word type and preview condition for gaze durations. However, the interaction was not significant for first fixation durations. One might query why we found a significant interaction for gaze durations, but not for first fixation durations. It should be noted that the target regions were four characters long and usually required more than one fixation to process. Thus, the gaze duration in the target regions reflected the total amount of time needed to process the whole words, whereas first fixation durations only reflected part of the time. Hence, we believe that it is reasonable to focus on gaze durations and reach a conclusion based on the results of gaze durations when the target regions require more than one fixation to process, as was the case in our study.

In summary, word boundaries affected character-order encoding in Chinese reading, suggesting that character-order encoding is constrained within a word, even when no interword spaces mark the word boundaries. Our results generally suggest that words are processed as a unit even when there are no spaces between words. Furthermore, word segmentation must occur very early, so that it can affect character-order encoding in Chinese reading.

Author note This research was supported by the Knowledge Innovation Program of the Chinese Academic Sciences (Grant No. KSCX2-YW-BR-6) and by a grant from the Natural Science Foundation of China (No. 31070904). We thank Yanping Liu and Guojie Ma for their help on the data analyses. We sincerely thank Manuel Perea and Qingqing Qu for advice and helpful suggestions on an earlier draft of the article. We also thank Kiel Christianson and the anonymous reviewers for their expert advice and suggestions.

References

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language, 35,* 775–800. doi:10.1006/jmla.1996.0040

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1277–1287.

Beyersmann, E., Coltheart, M., & Castles, A. (2012). Parallel processing of whole words and morphemes in visual word recognition. *Quarterly Journal of Experimental Psychology, 65,* 1798–1819. doi:10.1080/17470218.2012.672437

Beyersmann, E., McCormick, S. F., & Rastle, K. (2013). Letter transpositions within morphemes and across morpheme boundaries. *Quarterly Journal of Experimental Psychology, 66,* 2389–2410. doi:10.1080/17470218.2013.782326

Bruner, J. S., & O'Dowd, D. (1958). A note on the informativeness of parts of words. *Language and Speech, 1,* 98–101.

Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior, 18,* 225–241. doi:10.1016/S0022-5371(79)90136-1

Cheng, C. M. (1981). Perception of Chinese-characters. *Acta Psychologica Taiwanica, 23,* 137–153.

Christianson, K., Johnson, R. L., & Rayner, K. (2005). Letter transpositions within and across morphemes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1327–1339. doi:10.1037/0278-7393.31.6.1327

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition, 105,* 691–703. doi:10.1016/j.cognition.2006.12.001

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *Quarterly Journal of Experimental Psychology, 39A,* 211–251. doi:10.1080/14640748708401785

Frost, R. (2009). Reading in Hebrew vs. reading in English: Is there a qualitative difference? In K. Pugh & P. McCradle (Eds.), *How children learn to read: Current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice* (pp. 235–254). New York: Psychology Press.

Frost, R., Kugler, T., Deutsch, A., & Forster, K. I. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1293–1326. doi:10.1037/0278-7393.31.6.1293

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115,* 577–600. doi:10.1037/a0012667

Grainger, J., & Van Heuven, W. J. B. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon: Some words to talk about words* (pp. 1–23). Hauppauge: Nova Science.

Holmes, V. M., & Ng, E. (1993). Word-specific knowledge, word-recognition strategies, and spelling ability. *Journal of Memory and Language, 32,* 230–257. doi:10.1006/jmla.1993.1013

Johnson, R. L. (2007). The flexibility of letter coding: Nonadjacent letter transposition effects in the parafovea. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 425–440). Oxford: Elsevier.

Johnson, R. L., & Dunne, M. D. (2012). Parafoveal processing of transposed-letter words and nonwords: Evidence against parafoveal lexical activation. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 191–212. doi:10.1037/a0025983

Johnson, R. L., & Eisler, M. E. (2012). The importance of the first and last letter in words during sentence reading. *Acta Psychologica, 141,* 336–351.

Johnson, R. L., Perea, M., & Rayner, K. (2007). Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 209–229. doi:10.1037/0096-1523.33.1.209

Kinoshita, S., & Norris, D. (2009). Transposed-letter priming of prelexical orthographic representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1–18. doi:10.1037/a0014277

Lee, C. H., & Taft, M. (2009). Are onsets and codas important in processing letter position? A comparison of TL effects in English and Korean. *Journal of Memory and Language, 60,* 530–542. doi:10.1016/j.jml.2009.01.002

Lee, C. H., & Taft, M. (2011). Subsyllabic structure reflected in letter confusability effects in Korean word recognition. *Psychonomic Bulletin & Review, 18,* 129–134. doi:10.3758/s13423-010-0028-y

Lexicon of Common Words in Contemporary Chinese Research Team. (2008). *Lexicon of common words in contemporary Chinese*. Beijing: Commercial Press.

Li, X., Gu, J., Liu, P., & Rayner, K. (2013). The advantage of word-based processing in Chinese reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 879–889. doi:10.1037/a0030337

Li, X., & Ma, G. (2012). Word boundaries affect visual attention in Chinese reading. *PLoS ONE, 7*(e48905), 1–7. doi:10.1371/journal.pone.0048905

Li, X., & Pollatsek, A. (2011). Word knowledge influences character perception. *Psychonomic Bulletin & Review, 18,* 833–839.

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology, 58,* 525–552. doi:10.1016/j.cogpsych.2009.02.003

Luke, S. G., & Christianson, K. (2013). The influence of frequency across the time course of morphological processing: Evidence from the transposed-letter effect. *Journal of Cognitive Psychology, 25,* 781–799. doi:10.1080/20445911.2013.832682

Masserang, K. M., & Pollatsek, A. (2012). Transposed letter effects in prefixed words: Implications for morphological decomposition. *Journal of Cognitive Psychology, 24,* 476–495.

O'Connor, R. E., & Forster, K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory & Cognition, 9,* 78–92. doi:10.3758/BF03196953

Perea, M., & Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic Bulletin & Review, 13,* 418–422. doi:10.3758/BF03193863

Perea, M., & Fraga, I. (2006). Transposed-letter and laterality effects in lexical decision. *Brain and Language, 97,* 102–109.

Perea, M., & Lupker, S. J. (2003a). Does *jugde* activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition, 31,* 829–841. doi:10.3758/BF03196438

Perea, M., & Lupker, S. J. (2003b). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 97–120). New York: Psychology Press.

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51,* 231–246. doi:10.1016/j.jml.2004.05.005

Perea, M., Nakatani, C., & van Leeuwen, C. (2011). Transposition effects in reading Japanese Kana: Are they orthographic in nature? *Memory & Cognition, 39,* 700–707. doi:10.3758/s13421-010-0052-1

Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics, 67,* 301–314. doi:10.3758/BF03206493

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review, 11,* 1090–1098. doi:10.3758/BF03196742

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65–81.

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

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

Rayner, K., Li, X., & Pollatsek, A. (2007). Extending the E-Z Reader model of eye movement control to Chinese readers. *Cognitive Science, 31,* 1021–1033.

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled lettres: There is a cost. *Psychological Science, 17,* 192–193. doi:10.1111/j.1467-9280.2006.01684.x

Rueckl, J. G., & Rimzhim, A. (2011). On the interaction of letter transpositions and morphemic boundaries. *Language & Cognitive Processes, 26,* 482–508. doi:10.1080/01690965.2010.500020

Sánchez-Gutiérrez, C., & Rastle, K. (2013). Letter transpositions within and across morphemic boundaries: Is there a cross-language difference? *Psychonomic Bulletin & Review, 20,* 988–996. doi:10.3758/s13423-013-0425-0

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes, 19,* 333–367. doi:10.1080/01690960344000198

Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory & Language, 40,* 498–519.

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1261–1276. doi:10.1037/0096-1523.34.5.1261

Winskel, H., Perea, M., & Ratitamkul, T. (2012). On the flexibility of letter position coding during lexical processing: The case of Thai. *Experimental Psychology, 59,* 68–73.

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