Cognitive Mechanisms of Chinese Character Position Processing and Word Boundary Effects

Authors: Gu Junjuan, Shi Jinfu, Gu Junjuan

Date: 2020-10-18T00:00:00+00:00

Abstract

In Chinese reading, to perform lexical recognition, it is necessary to process the relative order of characters within words, that is, character position processing. Character position processing is a crucial component of lexical recognition and reading comprehension. If character position is not processed, transposable words will be difficult to distinguish. Transposable words are composed of the same characters but with each character occupying a different position. Previous studies have found that within-word character position processing is relatively flexible; moreover, word boundary information influences character position processing, that is, there exists a word boundary effect in character position processing. What are the causes and influencing factors of this effect? This project employs eye-tracking technology to systematically investigate the following three questions in natural reading: (1) How the first/last characters of a word modulate the character position processing mechanism; (2) How nested words and ambiguous word boundaries affect character position processing; (3) How contextual plausibility and predictability influence the word boundary effect in character position processing. This project will provide data support and a scientific foundation for establishing the first Chinese character position processing model, promote the further refinement of existing Chinese lexical recognition and word segmentation models, and provide scientific guidance for efficient vocabulary teaching and learning.

Full Text

The Cognitive Mechanism of Chinese Character Position Processing and Word Boundary Effects

GU Junjuan; SHI Jinfu
(School of Psychology and Mental Health, North China University of Science and Technology, Tangshan 063210, China)

Abstract: In visual word recognition of Chinese words, it is necessary to process the relative order of Chinese characters within words, i.e., Chinese character position processing. This processing is crucial for word recognition and reading comprehension. If character positions are not processed, transposable words would be difficult to distinguish. Transposable words are composed of the same characters but with different positional arrangements. Previous studies have found that within-word character position processing is relatively flexible, and that word boundary information affects character position processing—demonstrating a word boundary effect. What factors influence this effect? This project systematically investigates three questions using eye-tracking technology during natural reading: (1) How do the initial and final characters of a word modulate character position processing? (2) How do the word boundaries of embedded words and ambiguous words affect character position processing? (3) How do contextual plausibility and predictability influence the word boundary effect in character position processing? This project will provide empirical support and a scientific foundation for establishing the first Chinese character position processing model, advance the refinement of existing Chinese word segmentation and recognition models, and offer scientific guidance for efficient vocabulary teaching and learning.

Keywords: Chinese; word segmentation; word boundaries; Chinese character position; eye movements
Classification Code: B842

1 Research Background

In Chinese reading, recognizing words requires processing the sequential order of characters within words, a process known as Chinese character position processing (Hua et al., 2017). Some two-character words in Chinese are transposable—reversing the positions of the two characters creates different words, such as "笔画" (stroke) and "画笔" (paintbrush). If character positions were not processed, the two members of a transposable word pair would be difficult to distinguish (Peng et al., 1999). Therefore, word recognition involves not only processing the identity of constituent characters but also their positions within the word—both are indispensable. Moreover, research has found that some children with reading disabilities confuse character positions within phrases (Meng & Shu, 1999). Tian et al. (2006) conducted a case study showing that children with developmental dyslexia exhibited difficulty in lexical decision tasks for non-words with reversed character positions (e.g., "律规" for "规律"), performing significantly worse than control children. This indicates that some children with reading disabilities have problems processing character positions within words, which may prevent their word recognition ability from reaching normal levels. These developmental studies also demonstrate the important role of character position processing in word recognition and even reading comprehension.

Most alphabetic writing systems use spatial cues to mark word boundaries. For example, spaces between words in English separate continuous letter strings into distinct words. Research shows that spaces not only facilitate word recognition for English readers but also effectively guide their eye movements (Juhasz et al., 2005; Rayner et al., 1998). Unlike alphabetic scripts, Chinese written or printed texts lack boundary markers such as spaces between words, except for punctuation marks that indicate semantic units and pauses (Li et al., 2013; Li et al., 2009; Zhou et al., 2017). In alphabetic scripts like English, letter position coding is constrained by inter-word spaces, whereas Chinese character position coding is not limited by word boundaries and can be encoded across lexical boundaries.

A word is the smallest unit in language that can be used independently with both sound and meaning (Huang & Liao, 2007). The process by which readers segment words from sentences or texts during Chinese reading is referred to as "word segmentation" (Li et al., 2011). Word segmentation—i.e., determining word boundaries—is a crucial component of word recognition in text reading. According to the Chinese word segmentation and recognition model proposed by Li et al. (2009), word recognition and segmentation are unified processes. In Chinese reading, besides processing the characters themselves, character position processing is also an essential part of lexical processing. Moreover, when reading texts, character position processing includes not only within-word character position processing but also inter-word position information processing during word segmentation. So how does boundary information between words affect character position processing during word segmentation? A thorough investigation of the relationship between word boundaries and character position processing will help us understand how bottom-up processing at the character level influences word segmentation and text comprehension, and how lexical and contextual information top-down influences character position processing during dynamic word segmentation. Therefore, understanding the relationship between character position processing and word boundary effects is a key link in revealing the cognitive mechanisms of Chinese reading.

2.1 Effects of Morpheme Boundaries on Letter Position Processing in Alphabetic Scripts

In alphabetic scripts, letter position processing has gradually become a focal point in language cognition research. Using the masked priming lexical decision task, researchers first present a brief prime stimulus followed by a target word. Three priming conditions are used: (1) identity condition, where the prime is identical to the target (e.g., speak-speak); (2) transposition condition, where the prime is a non-word formed by transposing two letters in the word (e.g., spaek-speak); and (3) substitution condition, where the prime is a non-word formed by substituting the transposed letters with other letters (e.g., spuok-speak). Participants must judge as quickly as possible whether the target is a real word. Results show that reaction times are shortest in the identity condition, and significantly longer in the substitution condition than in the transposition condition (Forster et al., 1987; Johnson & Dunne, 2012; Masserang & Pollatsek, 2012; Perea & Lupker, 2003; Winskel et al., 2012). Additionally, some masked priming studies using only the two non-word conditions also found significantly longer reaction times in the substitution than transposition condition (Acha & Perea, 2008; Duñabeitia et al., 2007; Duñabeitia et al., 2014; Perea & Acha, 2009; Perea et al., 2008; Perea & Carreiras, 2008; Schoonbaert & Grainger, 2004; Witzel et al., 2011; Ziegler et al., 2014). The finding that transposed-letter non-words facilitate activation of the original word more than substituted-letter non-words is known as the transposed-letter effect.

Transposed-letter effects have also been observed in sentence reading (Blythe et al., 2014; Johnson, 2007; Johnson et al., 2007; Johnson & Dunne, 2012; Masserang & Pollatsek, 2012; Pagán et al., 2016; Tiffin-Richards & Schroeder, 2015; Winskel & Perea, 2013). Researchers typically use the boundary paradigm: an invisible boundary is set to the left of the target word, and as readers' eyes cross this boundary during left-to-right sentence reading, the preview stimulus changes to the target word (Rayner, 1975). Similar to masked priming lexical decision tasks, parafoveal preview conditions include identity, transposition, and substitution conditions. Results show that fixation durations on target words are shortest in the identity condition, and significantly longer in the substitution than transposition condition. Thus, in sentence reading, transposed-letter non-words facilitate word recognition more than substituted non-words, demonstrating a transposed-letter effect. Longer reaction times (fixation durations) in the transposition than identity condition indicate that letter positions are processed and play an important role in lexical processing. The emergence of transposed-letter effects suggests that letter position coding is relatively flexible. If letter position coding were strict, reaction times should not differ between transposed-letter and substituted-letter non-word conditions, as both contain two incorrect letters while the other letters remain the same.

Some researchers in alphabetic scripts have examined how morpheme boundaries in compound words affect letter position processing. In masked priming lexical decision tasks, priming conditions include transposition and substitution conditions, with position manipulations occurring either within or across morpheme boundaries. For English suffixed words, reaction times are significantly longer in the substitution than transposition condition both within and across morpheme boundaries, demonstrating transposed-letter effects in both conditions (Beyersmann et al., 2012; Beyersmann et al., 2013; Rueckl & Rimzhim, 2011). Masserang and Pollatsek (2012) also found in two eye-tracking experiments and one masked priming experiment that for English prefixed words, reaction times (fixation durations) were significantly longer in the substitution than transposition condition both within and across morpheme boundaries. This suggests that morpheme boundaries do not affect transposed-letter effects, and that letter position processing is not influenced by morpheme boundaries. Masserang and Pollatsek argued that because prefixes and suffixes are high-frequency letter clusters, skilled readers can easily correct letter position coding errors during early lexical recognition.

However, other studies have yielded inconsistent results. Christianson et al. (2005) used a masked priming naming task to investigate the effect of morpheme boundaries on letter position processing (using compound words composed of free morphemes, such as "cowboy"). They found no significant difference in naming times between within-morpheme transposition and identity conditions, but significantly shorter naming times in within-morpheme transposition than substitution conditions. In contrast, naming times were significantly longer in cross-morpheme boundary transposition than identity conditions, with no difference between cross-morpheme transposition and substitution conditions. Thus, cross-morpheme boundary transposition interfered with letter position processing and inhibited word recognition.

Duñabeitia et al. (2007) used Spanish prefixed words, suffixed words, and monomorphemic words as materials, with priming conditions including transposition and substitution. The researchers manipulated two letters at affix boundaries and two letters within monomorphemic words. Results showed that for monomorphemic words, reaction times were significantly longer in the substitution than transposition condition, producing a strong transposed-letter effect. However, for prefixed and suffixed words, reaction times did not differ between substitution and transposition conditions, indicating that transposition across morpheme boundaries eliminated the transposed-letter effect. Several subsequent studies using English compound words yielded consistent results (Luke & Christianson, 2013; Taft & Nillsen, 2013). These studies confirmed that letter transposition across morpheme boundaries interferes more strongly with word recognition than within-morpheme transposition, and that letter position processing is affected by morpheme boundaries.

Some researchers have proposed that English and Spanish have different morphological processing systems, and that letter position coding patterns may differ accordingly (Frost, 2009; Frost et al., 2005). Spanish has richer morphological diversity than English, and morphological parsing plays a more important role (Frost, 2009). Therefore, the precision of letter position coding at morpheme boundaries is crucial for word recognition in Spanish compound words, which explains why transposed-letter effects disappear when letters are transposed across morpheme boundaries. To further clarify this issue, Sánchez-Gutiérrez and Rastle (2013) conducted two masked priming lexical decision tasks with native Spanish and native English speakers, using Spanish-English cognates as target words. They found no difference in transposed-letter effects between within-stem and cross-morpheme boundary transpositions, suggesting that letter transposition effects are not modulated by the position of transposed letters within words. Therefore, they argued that orthographic information coding does not differ between Spanish and English, and that letter position processing is not affected by morpheme boundaries.

In summary, although morpheme boundaries may affect letter position processing, this effect likely depends on the type of morphemes in compound words and requires further investigation.

2.2 Effects of Word Boundaries on Chinese Character Position Processing

Research on the cognitive mechanisms of Chinese character position processing is still in its early exploratory stages. Researchers have used both masked priming paradigms and eye-tracking technology to examine character position processing. Target words were non-transposable two-character words, with three conditions: (a) identity condition (e.g., 扑簌—扑簌); (b) transposition condition (e.g., 簌扑—扑簌); and (c) substitution condition (e.g., 藐幼—扑簌). The first characters in transposition and substitution conditions were matched for character frequency and stroke count, as were the second characters. To control for consistency in character structure, corresponding characters in transposition and substitution conditions shared the same structure (e.g., top-bottom, left-right). Additionally, corresponding characters differed in phonology and had no semantic relationship. Results showed that reaction times and fixation durations on target words were shortest in the identity condition, with significantly shorter times in the transposition than substitution condition. This study initially confirmed that character positions are processed and that transposed-character effects exist in Chinese reading (Gu et al., 2015). However, this conclusion was limited to within-word character position processing in two-character words.

Given that Chinese texts lack boundary markers between words, Gu and Li (2015) examined how word boundaries affect character position processing in sentence reading. Experimental materials consisted of four-character idioms (one-word condition) and two two-character words (two-word condition), manipulating the second and third characters of target words to create three preview conditions: (a) identity condition (e.g., 捷足先登, 资金雄厚); (b) transposition condition (e.g., 捷先足登, 资雄金厚); and (c) substitution condition (e.g., 捷安兵登, 资湾界厚). Results showed that in the one-word condition, fixation durations on targets were significantly shorter in the transposition than substitution condition, with no difference between identity and transposition conditions. In the two-word condition, fixation durations did not differ between transposition and substitution conditions, but were significantly shorter in the identity than transposition condition. This indicates that within-word character transposition does not affect character position processing, whereas cross-word-boundary character transposition interferes with character position processing and inhibits word recognition. The study confirmed the existence of a word boundary effect in character position processing, showing that word boundary information influences character position processing.

In summary, character position information is processed during reading, and word boundary information affects this processing. However, the mechanism underlying the word boundary effect in character position processing requires further investigation.

2.3.1 Effects of Different Character Positions Within Words on Character Position Processing

Previous research has found that transposing two characters at word boundaries interferes with word recognition, whereas transposing the middle two characters of four-character words does not affect lexical processing (Gu & Li, 2015; Gu et al., 2020). This suggests that within-word character position coding is relatively flexible, while character position coding at word boundaries is more strict. Thus, characters at different positions within words may have varying importance in lexical processing. Within-word character positions may be less critical for lexical access, while the initial or final characters at word boundaries may be particularly important.

In alphabetic scripts, numerous studies have found that initial and final letters are more important for word recognition than internal letters (Johnson et al., 2007; Johnson & Eisler, 2012; Lee & Taft, 2009; Perea & Lupker, 2003; Rayner et al., 2006; Schoonbaert & Grainger, 2004; White et al., 2008). Johnson and Eisler (2012) investigated the cognitive mechanisms underlying the importance of initial and final letters in text reading. They found that with normal inter-word spacing, final-letter transposition caused significantly greater interference than internal-letter transposition. When spaces were inserted before and after each letter or when "#" filled inter-word spaces, the interference from internal-letter and final-letter transposition did not differ significantly. Across all three spacing conditions, initial-letter transposition caused significantly greater interference than internal-letter transposition. These results suggest that the intrinsic importance of initial letters may be related to higher-level cognitive functions involving lexical information processing and access, while the importance of final letters is likely related to lower-level perceptual factors—final letters receive less lateral inhibition from preceding and following letters compared to internal letters.

Unlike alphabetic scripts, Chinese lacks spaces between words. The perceptual span in Chinese reading is one character to the left and 2-3 characters to the right of fixation. Readers can process the two characters to the right of fixation relatively precisely, with information acquisition limited to within three characters to the right of fixation (Inhoff & Liu, 1998). Given that Chinese reading' s perceptual span is much smaller than that of alphabetic scripts, do initial and final characters of multi-character Chinese words have equal importance in lexical access? This requires further investigation.

2.3.2 Effects of Different Types of Word Boundary Information on Character Position Processing

Previous research has found that character transposition at the boundaries of two consecutive two-character words affects character position processing and interferes with word recognition (Gu & Li, 2015; Gu et al., 2020). Ma et al. (2017) examined word segmentation and recognition processes in Chinese reading using two conditions: cross-character-word condition (e.g., "材料质量," containing a cross-character word "材质") and control condition (e.g., "材料数量," containing only two two-character words). Results showed no difference in fixation durations for the first two characters between conditions, but significantly longer fixation durations for the last two characters in the cross-character-word condition. This demonstrates that character processing is not strictly left-to-right serial processing and is not constrained by word boundaries.

In Chinese texts, besides boundaries between consecutive words, there are also embedded word boundaries and ambiguous word boundaries. For example, embedded words like "介绍信" and "贵金属" contain two-character words "介绍" and "金属," respectively. If we simultaneously manipulate the last two characters of both types of embedded words, would character transposition across word boundaries (介信绍) versus within embedded two-character words (贵属金) lead to different processing patterns? This requires further investigation. Additionally, the boundary information of overlapping ambiguous words is unclear. For instance, "外交谈判" contains three words: "外交," "交谈," and "谈判," making word boundaries ambiguous. According to the left-dominance hypothesis and unidirectional syntactic analysis hypothesis, lexical recognition proceeds strictly left-to-right, with absolute processing advantage for left-side words. For ambiguous words ABCD, two-character words AB and CD are activated sequentially, while BC will not be activated (Ma et al., 2014; Inhoff & Wu, 2005). If the middle two characters of an ambiguous word boundary are transposed, will a word boundary effect emerge? This needs experimental verification.

2.3.3 Effects of Textual Contextual Information on Character Position Processing

According to the lexical recognition and word segmentation model constructed by Li et al. (2009), Chinese word recognition is an interactive activation process involving multiple processing levels: feature level, character level, and word level. Input information feeds forward from lower to higher levels, while information also feeds back from higher to lower levels, influencing lower-level processing. At the character level, characters within the perceptual span are processed in parallel. Activation at the character level feeds forward to the word level to activate word units. When a word unit' s activation reaches a certain threshold, this activation feeds back to the character level of its constituent characters. Characters belonging to that word become activated faster than other characters. Word-level representations compete until only one word wins. However, this model posits that character coding is constrained by slots, limiting its ability to explain findings in character position processing.

In the revised integrated model of lexical processing and eye-movement control, there are also no hypotheses regarding character position processing mechanisms. Researchers have proposed that the model needs partial revision to explain character position processing. For example, they hypothesize that connections from visual feature level to character level may not be strictly limited to one slot, and a visual feature-level unit may connect to character units in adjacent slots (Li & Pollatsek, 2020). However, the different processing mechanisms for within-word and word-boundary character positions during lexical recognition and word segmentation are primary considerations for establishing a character position processing model, and also provide directions for revising lexical recognition and word segmentation models.

3 Research Questions

In summary, although previous research has preliminarily identified word boundary effects in character position processing, a comprehensive and in-depth investigation of the cognitive mechanisms underlying character position processing and word boundary effects in Chinese reading is still needed. The following scientific questions remain to be addressed:

First, the word boundary effect in character position processing mainly manifests as adjacent character transposition at word boundaries interfering with word recognition. This suggests that initial and final characters are more important than within-word characters. The coding mechanisms of initial and final characters and how they influence lexical processing need further exploration. Are initial and final characters equally important, or are initial characters more important than final characters? How do they modulate character position processing? A systematic investigation of this question will help reveal the cognitive mechanisms of the word boundary effect in character position processing at the character level in Chinese reading.

Second, besides boundary information between consecutive words, Chinese texts contain other types of word boundaries, such as those in embedded words and overlapping ambiguous words. The mechanisms by which different types of word boundary information affect character position processing during word segmentation and recognition need systematic investigation. How do embedded word and ambiguous word boundaries influence character position processing? Research on this question will help us comprehensively understand the influence of various word boundary information on character position processing.

Third, previous research has demonstrated that contextual information plays an important regulatory role in lexical processing during Chinese reading. In previous studies, the semantic context of targets had low predictability, and the plausibility of targets in sentence contexts was not manipulated (Gu et al., 2015; Gu & Li, 2015). At the contextual level, how do plausibility and predictability affect the word boundary effect in character position processing? Investigating this question will explore how high-level contextual information influences character position processing during word segmentation from the perspective of macro-level text comprehension, thereby demonstrating that character position processing and word segmentation are dynamic, context-dependent cognitive processes.

4.1 Research Plan

This project will use eye-tracking technology to systematically investigate the cognitive mechanisms of character position processing and word boundary effects during Chinese reading. The research approach will first examine how initial and final characters of three- and four-character words modulate character position processing at the character level, then investigate how embedded word and ambiguous word boundaries affect character position processing at the word level, and finally explore how contextual plausibility and predictability influence the word boundary effect in character position processing at the contextual level. To address the three major questions proposed, the following research designs are planned:

4.1.1 Effects of Initial and Final Characters on Character Position Processing

Previous studies have found that character position information is processed in Chinese reading and that a word boundary effect exists in character position processing (Gu et al., 2015; Gu & Li, 2015; Gu et al., 2020). However, no study has systematically examined the roles of initial characters, final characters, and within-word characters as word boundary information in lexical recognition. Study 1 will investigate how initial and final characters modulate character position processing in natural reading scenarios through two experiments.

Experiment 1 will examine the effects of initial and final characters on character position processing in three-character words (e.g., 座右铭). Target presentation conditions include four types: (a) identity condition; (b) front-transposition condition, where the first two characters are transposed to form a non-word; (c) rear-transposition condition, where the last two characters are transposed to form a non-word; and (d) initial-final transposition condition, where the first and last characters are transposed to form a non-word. The focus is to compare whether initial and final characters differ in their roles in character position processing. We predict that if results support the left-dominance hypothesis (Ma et al., 2014), the position of the initial character is more important than that of the final character. When the initial character participates in transposition, it will cause the strongest interference with lexical recognition and reading. Fixation durations are expected to be significantly longer in both front-transposition and initial-final transposition conditions than in the rear-transposition condition. If results support Li et al.' s (2009) word segmentation model, which posits parallel processing of characters within the perceptual span, then initial and final characters should be equally important for lexical recognition. Fixation durations are expected to not differ significantly among front-transposition, rear-transposition, and initial-final transposition conditions.

Experiment 2 will further examine the effects of initial and final characters on character position processing in four-character words (e.g., 望而生畏). Target presentation conditions include four types: (a) identity condition; (b) front-transposition condition, where the first two characters are transposed to form a non-word; (c) middle-transposition condition, where the middle two characters are transposed to form a non-word; and (d) rear-transposition condition, where the last two characters are transposed to form a non-word. Again, if results support the left-dominance hypothesis (Ma et al., 2014), the position of the initial character should be more important than that of the final character, with initial-character transposition causing the strongest interference with lexical recognition and reading. Fixation durations are expected to be significantly longer in the front-transposition condition than in both rear-transposition and middle-transposition conditions. If results support Li et al.' s (2009) word segmentation model positing parallel processing of characters within the perceptual span, then initial and final characters should be equally important for lexical recognition. Fixation durations are expected to not differ significantly among front-transposition, rear-transposition, and middle-transposition conditions.

4.1.2 Effects of Embedded Word and Ambiguous Word Boundaries on Character Position Processing

Study 2 will investigate how embedded word and ambiguous word boundaries affect character position processing during sentence reading through two experiments, aiming to explore the word boundary effect in character position processing from different word-level perspectives.

Experiment 3 will use eye-tracking to examine the effect of embedded word boundaries on character position processing. We will use front-embedded three-character words (AB-C) and rear-embedded three-character words (A-BC), simultaneously manipulating the last two characters of both types to create inter-word transposition/substitution and intra-word transposition/substitution conditions. The focus is to examine whether character transposition across word boundaries versus within embedded two-character words leads to different processing patterns for these two word types. Each target word has three preview conditions: (a) identity condition; (b) transposition condition, where the last two characters are transposed to form a non-word; and (c) substitution condition, where the last two characters are substituted to form a non-word. This experiment uses the boundary paradigm. According to Taft's (2004) hierarchical lexical processing model, activation of Chinese multi-character words requires activation of embedded words within them—that is, whole-word and constituent word processing are not parallel. This model assumes that activation of an embedded three-character word requires joint activation of the embedded two-character word and the remaining single character. Activation of the embedded two-character word is a critical step in whole-word activation (Zhou et al., 2017). Therefore, during processing of embedded three-character words, boundary information between the embedded two-character word and the other character must be processed. We predict that if results support the hierarchical lexical processing model, then embedded word boundaries will affect character position processing, showing a word boundary effect with significant interaction between word type and preview condition. For A-BC type words, fixation durations will be significantly shorter in the transposition than substitution condition; for AB-C type words, fixation durations will not differ between transposition and substitution conditions. If embedded three-character words are processed as whole words, then embedded word boundaries will not affect character position processing, with no significant interaction between word type and preview condition, and both word types should show transposed-character effects.

Experiment 4 will use eye-tracking combined with the boundary paradigm to examine the role of ambiguous word boundaries on character position processing. Targets include three types: ambiguous words AB-CD, non-ambiguous words AB-CD, and ABCD-type four-character words. Ambiguous words AB-CD contain ambiguous word boundaries, such as "外交谈判." Non-ambiguous words AB-CD contain only two two-character words, such as "外交策略." ABCD-type four-character words contain no two- or three-character words, such as "危言耸听." Additionally, the three two-character words contained in ambiguous words are matched for word frequency and stroke count, and corresponding two-character words in ambiguous and non-ambiguous conditions are matched for frequency and stroke count. Based on the negative correlation between word frequency and word length (Deng & Feng, 2013), this experiment cannot effectively match word frequency across ABCD-type four-character words and the other two word types. Gu and Li (2015) used non-ambiguous AB-CD words and ABCD-type four-character words to examine the word boundary effect in character position processing, controlling for word frequency as a factor in linear mixed models during data analysis, and found that word frequency did not affect results. Experiment 4 will use the same statistical approach to rule out frequency effects across the three target types. Each target has three preview conditions: (a) identity condition; (b) transposition condition, where the middle two characters are transposed to form a non-word; and (c) substitution condition, where the middle two characters are substituted to form a non-word. According to the left-dominance hypothesis and unidirectional syntactic analysis hypothesis, lexical recognition proceeds strictly left-to-right, with absolute processing advantage for left-side words. Moreover, once a character forms a word with the character to its left, it cannot be assigned to other words. For ambiguous words AB-CD, after two-character word AB is activated, two-character word CD is activated, while BC will not be activated (Ma et al., 2014; Inhoff & Wu, 2005). Therefore, if ambiguous word boundary middle characters are transposed or substituted, a word boundary effect should emerge. According to the independent processing hypothesis and multiple activation hypothesis, all words within the perceptual span are activated. Thus, a character can be activated with the character to its left or right, and processing of AB, BC, and CD occurs in parallel and independently (Ma et al., 2014; Inhoff & Wu, 2005). Inhoff and Wu (2005) found that fixation durations on ambiguous words (e.g., 专科学生) were significantly longer than on two types of non-ambiguous words (e.g., 专科毕业 and 外地学生), suggesting that all three two-character words in ambiguous words may be activated, supporting the multiple activation hypothesis. If ambiguous word boundary middle characters are transposed or substituted, the word boundary effect may be weakened or eliminated. Therefore, we predict that if ambiguous word boundaries affect character position processing, there will be a significant interaction between word type and preview condition. For ambiguous AB-CD and non-ambiguous AB-CD targets, fixation durations will not differ between transposition and substitution conditions; for ABCD four-character targets, fixation durations will be significantly shorter in the transposition than substitution condition. If ambiguous word boundaries do not affect character position processing, there will still be a significant interaction between word type and preview condition. For ambiguous AB-CD and ABCD four-character targets, fixation durations will be significantly shorter in the transposition than substitution condition; for non-ambiguous AB-CD targets, fixation durations will not differ between transposition and substitution conditions.

4.1.3 Effects of Contextual Plausibility and Predictability on the Word Boundary Effect in Character Position Processing

Study 3 will investigate how contextual plausibility and predictability affect the word boundary effect in character position processing through two experiments, aiming to explore how high-level contextual information modulates character position processing during word segmentation in Chinese reading.

Experiment 5 will use eye-tracking combined with the boundary paradigm to examine how contextual plausibility affects the word boundary effect in character position processing. We will manipulate two types of four-character words: one where transposition of the middle two characters forms two two-character words with unchanged meaning (e.g., 眉清目秀), which fits the sentence context; and another where transposition forms two two-character words with changed meaning (e.g., 安贫乐道), which does not fit the context. This experiment primarily tests whether contextual plausibility modulates the word boundary effect in transposed-character processing of four-character words. Each target word has three preview conditions: (a) identity condition; (b) transposition condition, where the middle two characters are transposed to form two two-character words; and (c) substitution condition, where the middle two characters are substituted to form a non-word. Yang (2013) used transposable two-character words to examine how contextual plausibility affects lexical processing in Chinese reading. Target words were of two types: transposable words whose two members had the same meaning (e.g., 适合 and 合适) and those with different meanings (e.g., 画笔 and 笔画). Results showed that when the preview word and target had different meanings and the preview did not fit the sentence context, fixation durations were significantly longer in the transposition than identity condition; when they had the same meaning and both fit the context, fixation durations did not differ between identity and transposition conditions. Yang further found that when preview and target had different meanings but both fit the context, fixation durations did not differ between identity and transposition conditions. These results indicate that sentence context plausibility affects transposable word recognition. We predict that in transposable four-character words, if contextual plausibility affects the word boundary effect in character position processing, there will be a significant interaction between word type and preview condition. For four-character words that remain plausible after transposition, fixation durations will not differ between identity and transposition conditions, but will be significantly longer in the substitution than transposition condition. For four-character words that become implausible after transposition, fixation durations will be significantly longer in the transposition than identity condition, with no difference between transposition and substitution conditions. If contextual plausibility does not affect the word boundary effect, there will be no significant interaction between word type and preview condition.

Experiment 6 will examine how contextual predictability affects the word boundary effect in character position processing. Target words are two consecutive two-character words, divided into two types based on predictability: high-predictability and low-predictability. By comparing differences in transposition effects when middle characters are transposed under these two predictability conditions, this experiment tests whether predictability modulates the word boundary effect in character position processing. Each target has three preview conditions: (a) identity condition; (b) transposition condition, where the middle two characters are transposed to form a non-word; and (c) substitution condition, where the middle two characters are substituted to form a non-word. According to the integrated model of lexical processing and eye-movement control proposed by Li and Pollatsek (2020), contextual predictability affects processing of characters within targets. High-predictability targets are more likely to be skipped because both the target and its constituent characters are processed in parallel. Moreover, in parafoveal vision, characters in high-predictability targets receive more processing than those in low-predictability targets. We therefore speculate that high-predictability contexts may weaken the word boundary effect in character position processing, with strong transposed-character effects for high-predictability targets and weak or absent effects for low-predictability targets. We predict that if contextual predictability affects the word boundary effect, there will be a significant interaction between predictability and preview condition. For high-predictability conditions, fixation durations will not differ between identity and transposition conditions, but will be significantly shorter in the transposition than substitution condition. For low-predictability conditions, fixation durations will be significantly longer in the transposition than identity condition, with no difference between transposition and substitution conditions. If contextual predictability does not affect the word boundary effect, there will be no significant interaction between predictability and preview condition.

5 Theoretical Construction

Character position information processing is an important component of word recognition. In alphabetic scripts, research on letter position coding is more systematic and in-depth. Through extensive empirical research on the cognitive mechanisms of letter position coding, researchers have proposed a series of visual word recognition models, such as the spatial coding model and the overlap model. These models use flexible mechanisms to encode letter identity and position, thus explaining transposed-letter effects. However, some models cannot yet explain findings such as the initial/final letter position effect. Currently, research on Chinese character position processing remains in its early exploratory stages. This project will systematically investigate the word boundary effect in character position processing during Chinese reading from three dimensions—character, word, and context—specifically examining the detailed cognitive mechanisms of character position processing during word segmentation. This has important implications for establishing a Chinese character position processing model.

The multiple activation model of Chinese lexical processing proposed by Taft and Zhu (1997) assumes that the lexical processing system includes three subsystems: orthographic, phonological, and semantic. The system comprises four levels: feature, radical, character, and multi-character word. When a word is presented visually, it enters the lexical processing system from the orthographic subsystem based on the lowest-level features (e.g., strokes). Activation then spreads to the radical level associated with activated features, then to the associated character level, and finally to the associated multi-character word level. Moreover, activation can spread from character and word levels to closely related phonological and semantic subsystem units, thereby activating the phonology and semantics of characters and words. Thus, this lexical processing system primarily activates whole-word representations through character-level representations. However, at the character level, this model does not define the activation mechanism for character positions.

The word segmentation and lexical recognition model proposed by Li et al. (2009) assumes that Chinese word recognition is an interactive activation process involving multiple processing levels, such as feature, character, and word levels. Input information feeds forward from lower to higher levels, while information also feeds back from higher to lower levels, influencing lower-level processing. At the character level, characters within the perceptual span are processed in parallel. Activation at the character level feeds forward to the word level to activate word units. When a word unit' s activation reaches a certain threshold, this activation feeds back to the character level of its constituent characters. Characters belonging to that word become activated faster than other characters. Word-level representations compete until only one word wins. However, this model 主张 s that character coding is constrained by slots, limiting its ability to explain findings in character position processing.

In the new version of the integrated model of lexical processing and eye-movement control, there are also no hypotheses regarding character position processing mechanisms. Researchers have proposed that the model needs partial revision to explain character position processing. For example, they hypothesize that connections from visual feature level to character level may not be strictly limited to one slot, and a visual feature-level unit may connect to character units in adjacent slots (Li & Pollatsek, 2020). However, the different processing mechanisms for within-word and word-boundary character positions during lexical recognition and word segmentation are primary considerations for establishing a character position processing model, and also provide directions for revising lexical recognition and word segmentation models.

References

Deng, Y., & Feng, Z. (2013). A quantitative linguistic study of the relationship between word length and word frequency. *Foreign Languages*, 36, 29–39. doi: CNKI:SUN:WYXY.0.2013-03-002

Gu, J., Gao, Z., & Qu, Q. (2020). The word boundary effect in Chinese character position processing. *Studies of Psychology and Behavior*, 18, 193–199.

Hua, H., Gu, J., Lin, N., & Li, X. (2017). Character position coding in visual word recognition. *Advances in Psychological Science*, 25, 1132–1138. doi: 10.3724/SP.J.1042.2017.01132

Huang, B., & Liao, X. (2007). *Modern Chinese*. Beijing: Higher Education Press.

Li, X., Liu, P., & Ma, G. (2011). A review of the cognitive mechanisms of word segmentation in Chinese reading. *Advances in Psychological Science*, 19, 459–470. doi: 10.3724/SP.J.1042.2011.00459

Meng, X., & Shu, H. (1999). Research on reading disabilities in Chinese children. *Psychological Development and Education*, 15, 54–57. doi: CNKI:SUN:XLFZ.0.1999-04-011

Peng, D., Ding, G., Wang, C., Taft, M., & Zhu, X. (1999). Processing of Chinese reversed-order words: The role of morphemes in word processing. *Acta Psychologica Sinica*, 31, 36–46. doi: 10.1088/0256-307X/16/12/025

Tian, X., Wang, C., Li, T., Meng, X., & Zhou, X. (2006). Mental lexicon storage of Chinese two-character words: Evidence from a developmental dyslexia case. *Psychological Development and Education*, 22, 146–157. doi: 10.3969/j.issn.1001-4918.2006.02.009

Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology*, 99, 245–264. http://dx.doi.org/10.1348/000712607X224478

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

Blythe, H. I., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2014). Reading transposed text: Effects of transposed letter distance and consonant-vowel status on eye movements. *Attention, Perception, & Psychophysics*, 76, 2424–2440. http://dx.doi.org/10.3758/s13414-014-0707-2

Beyersmann, E., McCormick, S. F., & Rastle, K. (2013). Letter transpositions within morphemes and across morpheme boundaries. *The Quarterly Journal of Experimental Psychology*, 66, 1524–1535. http://dx.doi.org/10.1080/17470218.2013.782326

Christianson, K., Johnson, R. L., & Rayner, K. (2005). Letter transpositions within and across morphemes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1327–1335. http://dx.doi.org/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, 408–415. http://dx.doi.org/10.1016/j.cognition.2006.12.001

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2014). Revisiting letter transpositions within and across morphemic boundaries. *Psychonomic Bulletin & Review*, 21, 1557–1575. http://dx.doi.org/10.3758/s13423-014-0609-2

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.

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: Human Experimental Psychology*, 39(A), 211–251. http://dx.doi.org/10.1080/14640748708401785

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

Gu, J., & Li, X. (2015). The effects of character transposition within and across words in Chinese reading. *Attention, Perception, & Psychophysics*, 77, 272–281. http://dx.doi.org/10.3758/s13414-014-0749-5

Gu, J., Li, X., & Liversedge, S. P. (2015). Character order processing in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 127–137. http://dx.doi.org/10.1037/a0038639

Inhoff, A. W., & Liu, W. M. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 20–34. http://dx.doi.org/10.1037//0096-1523.24.1.20

Inhoff, A. W., & Wu, C. (2005). Eye movements and the identification of spatially ambiguous words during Chinese sentence reading. *Memory & Cognition*, 33, 1345–1356. http://dx.doi.org/10.3758/BF03193367

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, UK: 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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.actpsy.2012.09.013

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. http://dx.doi.org/10.1037/0096-1523.33.1.209

Juhasz, B. J., Inhoff, A. W., & Rayner, K. (2005). The role of interword spaces in the processing of English compound words. *Language and Cognitive Processes*, 20, 291–298. http://dx.doi.org/10.1080/01690960444000133

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. http://dx.doi.org/10.1080/20445911.2013.832682

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. http://dx.doi.org/10.1037/a0030337

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

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*. Advance online publication. http://dx.doi.org/10.1037/rev0000248

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, 263–278. http://dx.doi.org/10.1016/j.jml.2009.01.002

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1046–1059. http://dx.doi.org/10.1037/a0035389

Ma, G., Pollatsek, A., Li, Y., & Li, X. (2017). Chinese readers can perceive a word even when it' s composed of noncontiguous characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 158–166. http://dx.doi.org/10.1037/xlm0000298

Masserang, K. M., & Pollatsek, A. (2012). Transposed letter effects in prefixed words: Implications for morphological decomposition. *Journal of Cognitive Psychology*, 24, 500–512. http://dx.doi.org/10.1080/20445911.2012.658037

Pagán, A., Blythe, H. I., & Liversedge, S. P. (2016). Parafoveal preprocessing of word initial trigrams during reading in adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 411–432. http://dx.doi.org/10.1037/xlm0000175

Perea, M., & Acha, J. (2009). Does letter position coding depend on consonant/vowel status? Evidence with the masked priming technique. *Acta Psychologica*, 130, 127–137. http://dx.doi.org/10.1016/j.actpsy.2008.11.001

Perea, M., & Carreiras, M. (2008). Do orthotactics and phonology constrain the transposed-letter effect? *Language and Cognitive Processes*, 23, 69–92. http://dx.doi.org/10.1080/01690960701578146

Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). Transposed-letter priming effects for close versus distant transpositions. *Experimental Psychology*, 55, 384–393. http://dx.doi.org/10.1027/1618-3169.55.6.384

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

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7, 65–81. http://dx.doi.org/10.1016/0010-0285(75)90005-5

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, 38, 1129–1144. http://dx.doi.org/10.1016/S0042-6989(97)00274-5.5

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. http://dx.doi.org/10.3758/BF03206448

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

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

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. http://dx.doi.org/10.1080/01690960344000198

Sánchez-Gutiérrez, C., & Rastle, K. (2013). Letter transpositions within and across morphemic boundaries: Is there cross-language difference? *Psychonomic Bulletin & Review*, 20, 1145–1151. http://dx.doi.org/10.3758/s13423-013-0425-0

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology Section A*, 57, 745–765. http://dx.doi.org/10.1080/02724980343000477

Taft, M., & Nillsen, C. (2013). Morphological decomposition and the transposed-letter (TL) position effect. *Language and Cognitive Processes*, 28, 917–938. http://dx.doi.org/10.1080/01690965.2012.679662

Tiffin-Richards, S. P., & Schroeder, S. (2015). Children' s and adults' parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology*, 27, 531–558. http://dx.doi.org/10.1080/20445911.2014.999076

Taft, M., & Zhu, X. (1997). Sub-morphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 761–775. http://dx.doi.org/10.1037/0278-7393.23.3.761

White, S., Johnson, R. L., Liversedge, S., & 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. http://dx.doi.org/10.1037/0096-1523.34.5.1261

Winskel, H., & Perea, M. (2013). Consonant/vowel asymmetries in letter position coding during normal reading: Evidence from parafoveal previews in Thai. *Journal of Cognitive Psychology*, 25, 119–130. http://dx.doi.org/10.1080/20445911.2012.753077

Winskel, H., Perea, M., & Ratitamkul, T. (2012). On the flexibility of letter position coding during lexical processing: Evidence from Thai. *Experimental Psychology*, 59, 216–223. http://dx.doi.org/10.1027/1618-3169/a000127

Witzel, N., Qiao, X., & Forster, K. (2011). Transposed letter priming with horizontal and vertical text in Japanese and English readers. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 914–920. http://dx.doi.org/10.1037/a0022194

Yang, J. (2013). Preview effects of plausibility and character order in reading Chinese transposed words: Evidence from eye movements. *Journal of Research in Reading*, 36, 383–402. http://dx.doi.org/10.1111/j.1467-9817.2013.01553.x

Ziegler, J. C., Bertrand, D., Lété, B., & Grainger, J. (2014). Orthographic and phonological contributions to reading development: Tracking developmental trajectories using masked priming. *Developmental Psychology*, 50, 1026–1036. http://dx.doi.org/10.1037/a0035187

Zhou, J., Ma, G., Li, X., & Taft, M. (2017). The time course of incremental word processing during Chinese reading. *Reading and Writing*, 31, 607–625. http://dx.doi.org/10.1007/s11145-017-9800-y