LANGUAGE, COGNITION AND NEUROSCIENCE
2025, VOL. 40, NO. 1, 41–55
https://doi.org/10.1080/23273798.2024.2390003

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Check for updates

The role of semantic information in Chinese word segmentation

Ruqi Chen[a,b], Linjieqiong Huang[a], Manuel Perea [c,d] and Xingshan Li[a,b]

[a]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, People's Republic of China; [b]Department of Psychology, University of Chinese Academy of Sciences, Beijing, People's Republic of China; [c]Departamento de Metodología and ERI-Lectura, Facultad de Psicología, Universitat de València, Valencia, Spain; [d]Centro de Investigación Nebrija en Cognición, Universidad Nebrija, Madrid, Spain

ABSTRACT

Word segmentation is crucial for reading in Chinese, where the absence of explicit word boundaries poses a distinct challenge. Previous studies in Chinese have examined how lexical and sub-lexical variables affect word segmentation. The present study investigated whether higher-level semantic information affects word segmentation using a primed word segmentation task with Overlapping Ambiguous Strings (OAS). An OAS is a three-character string in Chinese (e.g. ABC [in Latin letters]) where the middle character can constitute a word with both the left (word AB) and right (word BC) characters. The OAS was preceded by a semantic or repetition prime (presented for 42, 83, or 200 ms, across participants), priming either AB or BC. The semantic priming effect occurred at the 200-ms Stimulus Onset Asynchrony (SOA), whereas the repetition priming effect occurred at both 83 and 200-ms SOAs. These findings demonstrate that semantic information can affect word segmentation in Chinese within 200 ms.

ARTICLE HISTORY
Received 18 February 2024
Accepted 30 July 2024

KEYWORDS
Word segmentation; Chinese reading; overlapping ambiguous string; semantic priming; repetition priming

1. Introduction

Although Chinese texts lack spaces to demarcate words, Chinese readers can nonetheless read fluently (Liversedge et al., 2016). Numerous experiments have demonstrated that the absence of inter-word spaces does not diminish the importance of words in Chinese reading (e.g. Li et al., 2014). Thus, examining how Chinese readers segment words is crucial for understanding the mechanisms of Chinese reading. Over the last two decades, many studies have investigated how Chinese readers segment words without the aid of inter-word spaces and substantial progress has been made in addressing this question (see Li & Pollatsek, 2020, for a model of Chinese reading). However, previous studies have primarily focused on how sub-lexical or lexical factors (e.g. word frequency) affect word segmentation. This leaves the contribution of semantic information unresolved. In the present study, we investigated the role of semantic information during word segmentation in Chinese.

To study how Chinese readers segment words, previous research has typically employed stimuli with special characteristics: overlapping ambiguous strings (OAS). An OAS is a character string with ambiguous word boundaries (e.g. Hsu & Huang, 2000). The middle character of an OAS can constitute two distinct words with the characters on its left and right. For instance, in the OAS "客运营", which is composed of three characters (ABC), the two characters on the left (AB) may constitute the word "客运" [passenger transport], whereas the two characters on the right (BC) may constitute the word "运营" [operation]. Note that the OAS does not necessarily constitute a meaningful linguistic unit itself (e.g. "客运营" does not make up a meaningful phrase by itself). Thus, when encountering an OAS, Chinese readers need to determine which characters belong to a given word. This can result in two segmentation types: AB-C or A-BC, depending on which two characters are grouped. This segmentation problem can be compared to encountering an ambiguous trimorphemic nonword in English. For example, the trimorphemic nonword *milkteabag* can be segmented into *milktea-bag* or *milk-teabag*.

Ma et al. (2014) examined how Chinese readers segment overlapping ambiguous strings (OAS) during sentence reading. They embedded an OAS in sentences where only the following context disambiguated the segmentation. Based on the following context, the OAS should be segmented as AB-C or A-BC (i.e. global segmentation). Furthermore, at a local level within the

CONTACT Linjieqiong Huang huangljq@psych.ac.cn; Xingshan Li lixs@psych.ac.cn CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing, China
Supplemental data for this article can be accessed online at https://doi.org/10.1080/23273798.2024.2390003.

OAS, the frequency of the first word AB was either higher or lower than the second word BC, yielding high-low or low-high conditions, respectively. Therefore, the global segmentation was either consistent or inconsistent with the local segmentation. Ma and colleagues found that when the following context was consistent with the local segmentation based on frequency, readers made fewer regressions into the OAS region than in the inconsistent conditions. They proposed that Chinese readers process OAS using a two-stage procedure during sentence reading. In the first stage, readers segment words using local information such as word frequency, with words of higher frequency being more likely to be segmented as a word. Subsequently, in a later *information integration stage*, readers check whether the initial segmentation fits with the context. If it does not, they would correct the error, usually by regressing their eyes to a previous part of the sentence.

In addition to word frequency, previous studies have found other factors affecting Chinese word segmentation. For example, several experiments have shown a left-side word advantage (Huang & Li, 2020; Ma et al., 2014, Experiment 3): other things being equal, the left-side word is more likely to be segmented as a word. Moreover, the statistical cues of the word boundary can also influence word recognition and segmentation. Other studies (Liang et al., 2015, 2017, p. 2023; Yen et al., 2012) have shown an effect of character positional probability (i.e. some Chinese characters are more likely to be the last character of a word, while others are more likely to be the first character). For example, Yen et al. (2012) found that the processing of a word is more difficult when the ending characters of the words are more frequently used as word beginnings than word endings. Conversely, other studies, such as Liang et al. (2023), showed that only the position probability of the final character influences word segmentation. In sum, prior research has identified several lexical and sub-lexical factors that can influence word segmentation in Chinese.

To model how Chinese readers segment words during sentence reading, Li and Pollatsek (2020) proposed an integrated model of word processing and eye-movement control during Chinese reading (Chinese Reading Model, [CRM]). Using the framework of the interactive activation model (McClelland & Rumelhart, 1981), the CRM consists of three levels of units: a visual perception level, a character level, and a word level. The input of the model is constrained by the perceptual span[1], which extends from one character to the left of fixation to three characters to the right of fixation (Inhoff & Liu, 1998). To address the word segmentation problem, the CRM assumes that word identification and word segmentation are a unified process. Specifically, all the words constituted by characters within the perceptual span are activated during reading. Spatially overlapping words compete until a word unit prevails. When the activation of a given word unit surpasses a threshold, it is identified and simultaneously segmented. As shown by Li and Pollatsek (2020), the CRM can effectively simulate numerous benchmark findings of Chinese reading. For example, it can simulate the influence of word processing on eye-movement control (e.g. effects of word frequency, predictability, and word length) and how the left-side word advantage and word frequency influence the initial stage of word segmentation.

A shortcoming of most previous studies on Chinese word segmentation is their focus on sub-lexical or lexical variables, leaving the role of higher-level semantic information in this process largely unexplored. Even though several recent studies have examined the role of prior context in word segmentation, the role of semantic information in word segmentation is still not fully understood. For example, Huang and Li (2020) examined how prior context influences the word segmentation of OAS during sentence reading. They selected OAS in which A was a one-character verb, and AB was a two-character noun, and they manipulated the prior context to constrain the part-of-speech of the following word. In the informative condition, the prior context constrained the following word to be a noun or a verb, thus supporting either AB-C or A-BC segmentation, and the post-target context was consistent with this segmentation. In the neutral condition, both a noun and a verb were equally favoured by the prior context, resulting in equal support for the AB-C and A-BC segmentation, while only the post context provided disambiguating information. They found that first-pass reading times were longer in the A-BC condition than in the AB-C condition under informative contexts, whereas this difference disappeared under neutral contexts. Huang and Li (2020) reasoned that when the context favoured the A-BC segmentation, competition increased because the word AB is favoured by a left-word advantage. Thus, their study suggested that prior context can rapidly influence first-pass reading. Critically, however, the sentence context in their study was a mixture of semantic relations between words and syntactic constraints. Therefore, their results cannot provide direct evidence for the impact of semantic information on word segmentation.

Although Huang and Li (2020) demonstrated that prior context could elicit effects on first-pass reading, their study did not conclusively determine whether the prior context only affects the initial stage of word segmentation. In subsequent work, Huang et al. (2021) further confirmed that prior context could influence the word segmentation stage and not only the subsequent integration stage. In their study, they manipulated the extent to which the word BC was plausible as an immediate continuation following the prior context, while the word AB was always plausible given the prior context. They found that, compared with a less plausible word BC, first-pass reading times on the OAS were longer with a more plausible word BC. They reasoned that, when word AB and BC were both supported by prior context (i.e. a more plausible BC condition), the competition between word AB and BC was stronger compared to when only word AB was supported by prior context (i.e. a less plausible BC condition). Their results suggest that prior context could influence word competition between words AB and BC during the word segmentation stage, with words supported by prior context being more likely to be segmented. However, the context constraint manipulation in their study was not designed to distinguish between different types of constraint. Therefore, previous research has not directly tested whether semantic information affects Chinese word segmentation.

To sum up, the role of semantic information in word segmentation in Chinese remains unclear. There are two possibilities regarding whether semantic information affects word segmentation. One possibility is that semantic information does not affect the initial word segmentation, as proposed by Ma et al. (2014). In this scenario, the initial stage of word segmentation would only involve orthographic processing, whereas semantic information would only serve to check the initial segmentation in a later integration stage (see Ma et al., 2014). In line with this, the CRM did not include an explicitly designed semantic component but could adequately simulate the initial stage of word segmentation.

A second possibility is that semantic information does affect the initial stage of word segmentation. The idea is that semantic and orthographic processing may not be independent but may interact during the word segmentation process in Chinese. The semantic level would then provide feedback to the word level, thereby affecting word segmentation. This possibility is in line with the interactive activation models proposed for alphabetic languages (Carreiras et al., 2014; McClelland, 2016; McClelland & Rumelhart, 1981; Stolz & Besner, 1996). In these models, the representation formed at each level may be influenced by representations from both higher and lower processing levels. The present experiment was designed to distinguish these two possibilities. Therefore, the findings of the present study will be important to further develop more comprehensive models of Chinese reading; furthermore, they may also be informative for understanding the role of semantics during word segmentation in other unspaced writing systems (e.g. Thai).

In the present study, we investigated how semantic information affects Chinese word segmentation. To achieve this goal, we designed a primed OAS segmentation paradigm that combines the semantic priming paradigm with an OAS segmentation task. Semantic priming refers to the phenomenon where the processing of a target word (e.g. in lexical decision or naming tasks) is faster and more accurate when the prime and target are semantically related (e.g. doctor – NURSE) than when they are unrelated (e.g. knife – NURSE, Meyer & Schvaneveldt, 1971). The semantic priming paradigm has been a valuable tool for investigating word recognition and the structure of the mental lexicon (e.g. Dehaene et al., 1998; Hutchison, 2003; Hutchison et al., 2013; Perea & Gotor, 1997).

Previous studies have typically examined semantic priming in a lexical decision task, where participants are asked to judge whether the target character string is a word or not. However, this task is not suitable for our goal because both the left two characters and the right two characters of the OAS are words in Chinese. Instead, we developed a primed OAS segmentation task. To implement the primed OAS segmentation paradigm, a masked two-character-word prime was presented for a short duration depending on the Stimulus Onset Asynchrony (SOA) condition. Then an OAS was presented. Participants were asked to report the first word they recognised. This task was inspired by the pilot experiments conducted by Huang et al. (2021). They presented an OAS in isolation on paper and asked participants to "indicate the word boundaries" with a pen. Their three experiments showed that readers tend to segment high-frequency words more often than low-frequency words. These findings reveal the feasibility of the OAS segmentation task to test the factors impacting word segmentation in Chinese.

To assess how presenting a prime word affects the segmentation of a following character string, we used three types of primes: repeated (i.e. either AB or BC), semantically related (i.e. a word related to either AB or BC), and unrelated primes (i.e. primes were not related to AB or BC). For example, for the OAS "护理财", there are two sets of primes in the repetition priming condition, with one set corresponding to the first word "护理" [nursing] and the other set corresponding to the second word "理财" [wealth management]; for the semantic priming condition, the primes could be "照料" [take care of] or "赚钱" [make money]; for the unrelated condition, the primes could be "前锋" [vanguard] or "宁静" [silence].

To examine the time course of the priming effects in the OAS segmentation task, we also manipulated the SOA (42, 83 ms, or 200 ms, across participants). Following previous studies (e.g. Rastle et al., 2000), we selected these durations to examine the effects across a range of situations where the processing degree of the prime varies. A 42-ms duration prevents explicit recognition of the primes. At the 83-ms duration, most participants are aware of the existence of the prime but cannot fully identify it, while the duration of 200 ms enables full recognition of primes. Previous studies have often manipulated SOA to examine the time course when different types of information become available during lexical access (Chen & Shu, 2001; Marslen-Wilson et al., 2008; Perea & Rosa, 2002; Rastle et al., 2000). For example, Zhou and Marslen-Wilson (2000) aimed to investigate the relative time course of semantic and phonological activation in Chinese reading. These studies showed that the semantic priming effects are typically very small at an SOA of 60–80 ms in lexical decision, while the effect is larger and more reliable at a longer SOA (e.g. 200 ms). Moreover, in the present experiment, we also compared the pattern of repetition and semantic priming across the various SOAs to facilitate comparison of the OAS primed segmentation task with previous studies with word recognition tasks (Holcomb & Grainger, 2009; Lee & Zhang, 2018; Tan & Yap, 2016).

The combination of several prime-target SOAs and the use of semantic and repetition pairs will help clarify the role of semantic information in Chinese word segmentation. If semantic information influences the segmentation of OAS, we would observe a semantic priming effect, where participants would choose more frequently the corresponding word preceded by a semantic prime than by an unrelated prime. Likewise, reaction times (RT) in the semantically related conditions would be faster compared to unrelated conditions. This could be due to one of the words being preactivated by the prime, accelerating the segmentation process. As stated above, previous studies have usually shown that semantic priming requires higher-level processing of the prime to emerge than repetition priming (see Fabre et al., 2007; Holcomb et al., 2005; Holcomb & Grainger, 2009). Thus, when the SOA is very short, there may be insufficient time for the semantic activation from the primes to spread within the semantic layer and influence the OAS segmentation process. As a result, semantic priming may be restricted to the longer rather than the shorter SOAs.

Furthermore, comparing the effect size between the repetition priming and semantic priming effects on Chinese word segmentation can shed some light on whether semantic information provides additional influence over and above the form information from the repetition priming condition. In the repetition priming condition, the prime and the word it supports share both form information (including visual form and phonological form) and semantic information; in contrast, in the semantic priming condition, they only share semantic information. Because orthographic information is available earlier than semantic information during word processing, it is expected that the repetition priming effect could be observed at very short SOAs, as occurs in other word recognition paradigms.

Critically, there is the possibility that semantic information alone does not provide extra cues for OAS segmentation in Chinese; in this case, we would expect no semantic priming effect in the primed OAS segmentation task regardless of SOA. Either of these outcomes will have important implications for Chinese word segmentation models. As noted above, CRM did not implement a semantic processing component. If the present study shows strong evidence that semantic information plays an important role in Chinese word segmentation, CRM would need to be updated to include a semantic processing component.

2. Materials & method

2.1. Participants

Participants were students from colleges around the Institute of Psychology, Chinese Academy of Sciences. All of them were native Chinese speakers and had normal or corrected-to-normal vision. They received a small amount of monetary compensation for their participation.

The sample size was determined by power analysis, using the *mixedpower* function from the *mixedpower* package (Kumle et al., 2021). First, we conducted a pilot study with 30 participants, with 10 participants for each SOA condition, and analysed the pilot data (word segmentation results) with general linear mixed-effect models (see Results section for details). Then based on the results, we explored how the power varies as a function of the number of participants under different conditions. As previous studies usually do not report semantic effects at 42-ms SOA in the masked priming paradigm (e.g. de Wit & Kinoshita, 2015; however, see Wong et al., 2014), we aimed to achieve 80% power for all the other effects. Thus, we made our decision based on the semantic priming effect at the SOA of 83 ms, which has a lower overall power than that at the SOA of 200 ms. The results showed that it takes 36 participants to reach a power of 89% for the semantic priming effect at 83-ms SOA, suggesting that 36 participants for each SOA condition were enough for a well-powered experiment (Kumle et al., 2021). Thus, a total of 108 participants (75 females and 33 males) were recruited, 36 at each SOA condition. Their ages ranged from 18 to 28 years (M = 22.10 years, SE = 0.35).

2.2. Materials and design

The study was a 3 (SOA: 42, 83 ms, or 200 ms) × 3 (prime type: repetition, semantic, or unrelated) × 2 (primed word: AB or BC) design, with SOA being a between-participant factor, and with priming type and primed word being within-participant factors. The sample stimuli and the descriptive statistics are presented in Table 1.

We used a between-participants design for the SOA factor, as did in some previous studies (e.g. Zhou & Marslen-Wilson, 2000), for two main reasons. First, according to previous studies, mixing visible and invisible primes within a block can produce undesirable effects. When participants notice primes in some trials, their attention might be drawn to the prime-target relationship, increasing the priming effect. Zimmerman and Gomez (2012) found greater priming effects for 48-ms primes in the mixed condition (48-ms and 64-ms SOA) compared to a condition with all the primes presented for 48-ms. Second, because it has been hard to find enough OASs, a between-participants design allows for more items per SOA condition.

A total of 180 OASs were selected as the target items. In an OAS, both AB and BC are two-character words (which are listed as words in the Lexicon of Common Words in Contemporary Chinese, 2008). However, the full three-character combination (ABC) did not form a recognised word or phrase in Chinese. To verify that the three-character strings did not resemble words or meaningful phrases, we recruited 18 Chinese speakers who did not participate in the experiment to assess each OAS, asking them to rate the extent to which the three-character strings looked like a word or phrase they encountered daily. The participants reported on a 7-point Likert-type scale (1 = *very unlike*, 7 = *very like*). All scores were under 4, suggesting the participants did not regard them as words or meaningful units.

Table 2. Properties of characters in OAS.

Properties	A *M*	A *SE*	B *M*	B *SE*	C *M*	C *SE*
Character frequency[1]	687	50	2326	175	692	53
Frequency of single-character words	483	33	1659[2]	120	468	34
Number of strokes	8.63	0.22	7.00	0.21	8.44	0.21

Note. 1. The units of character frequency and word frequency are occurrences per million.
2. Among the 180 characters of Character B, 11 were not standalone words. The word frequency in this cell was calculated based on the Character Bs that were composed of the character alone.

To eliminate any selection bias, we carefully matched the word frequency of words AB and BC. Both characters A and C constitute a one-character word on their own, so we not only matched the character frequency but also matched the word frequency of A and C as a one-character word.[2] The statistics are shown in Table 2. The frequency data were obtained from the Lexicon of common words in contemporary Chinese (Lexicon of Common Words in Contemporary Chinese Research Team, 2008).

Each OAS was matched with six types of primes (see Table 1 for examples). Half of these six types of primes were designed to pair with either AB or BC in the target OAS strings. The primes had different relations with the primed words in different conditions. For the repetition priming condition, the prime was identical to either AB or BC, depending on the primed word condition. For the semantic condition, the prime was semantically related to either AB or BC. To evaluate the semantic relatedness between the prime and the primed word in the semantic condition, we asked 17 participants who did not participate in the formal experiment to rate the semantic relatedness between the primed words (i.e. ABs and BCs) and their corresponding semantic primes on a 7-point Likert-type scale (1 = *unrelated in meaning*, 7 = *highly related*). The relatedness score between the word AB and its semantic prime (M = 5.91, SE = 0.037) was comparable to that of BC (M = 5.84, SE = 0.043; $t(179)$ = 1.34, p = .184). For unrelated priming conditions, the prime did not relate to either

Table 1. Stimuli sample and the descriptive statistics of primes.

	Prime						
	Primed word: AB			Primed word: BC			
Properties	Repetition	Semantic	Unrelated	Repetition	Semantic	Unrelated	Target
Examples	护理	照料	前锋	理财	赚钱	宁静	护理财
Translation	Nursing	Take care of	Vanguard	Wealth management	Make money	Silence	
Word frequency	9.75 (6.83)	9.90 (8.93)	9.79 (7.32)	9.84 (6.84)	9.79 (8.42)	9.67 (6.80)	
Stroke numbers	15.64 (3.82)	15.95 (3.96)	15.76 (2.79)	15.44 (3.96)	16.01 (3.71)	15.77 (2.96)	

Note: The unit of word frequency is times of occurrence per million. *SD*s are given in parentheses.

AB or BC (words AB and unrelated primes: $M = 2.56$, $SE = 0.026$; words BC and unrelated primes: $M = 2.54$, $SE = 0.022$). The relatedness score between the primed words and their semantic primes was greater than that of the unrelated primes, $F(1, 179) = 8909$, $p < .001$, and the discrepancy between the semantic and unrelated conditions did not differ between different primed word conditions, $F(1, 179) = 0.797$, $p = .373$. Testing the priming effects against these baselines allows us to rule out any pre-existing segmentation bias, either due to the preference of the participants or the material properties. The word frequency and stroke number were comparable between the six conditions, and repeated analysis of variance showed no significant differences between conditions. There were no duplicate words in all the primes used in the experiment.

A Latin square design was used to counterbalance the assignment of the primes to their target OAS. Consequently, there are six lists containing all the OASs and there are 30 OASs under each condition. An OAS was preceded by different types of primes in different lists. The raw data and word stimuli used can be found at https://www.scidb.cn/en/anonymous/TmphUUZq.

2.3. Apparatus

Each participant was tested individually in a darkened and quiet room. The materials were presented on a 21-inch CRT monitor (Sony Multiscan G520) with a 1,024 × 768-pixel resolution and a refresh rate of 120 Hz. The experiment was programmed by E-prime 2.0 software (Psychology Software Tools Inc, Pittsburgh, PA). The primes and OAS were shown in different font styles and sizes (40-point Boldface font for the primes and 42-point Song font for the OAS). We made this design to decrease the visual similarity of the stimuli so that any repetition priming effect that emerges in our experiment could not be merely attributed to the overlap in visual features. The participants were asked to place their chins on a chin rest to minimise head movements. Participants were seated 60 cm away from the monitor. At this distance, each character of the prime subtended a visual angle of approximately 1.7°, and one character for the OAS subtended about 1.9°. All the stimuli were in white on a black background.

2.4. Procedure

The participants were randomly assigned to one of the three SOA conditions. For each trial, a fixation point (+) was presented for 500 ms, followed by a 250-ms blank screen, and then a 500-ms mask consisting of four ※ signs. Then, a prime was presented for a different duration depending on different SOA conditions (42, 83, or 200 ms), and then an OAS was presented until the participant's response. The fixation point, the middle of the mask, the middle of the prime, and the middle character of OAS overlapped at the center of the screen (see Figure 1).

Before the start of the experiment, participants were informed that they should keep their eyes on the center of the screen from the onset of the fixation cross until a response was made. They were instructed to indicate the position of the first word they recognised by pressing a key on the keyboard using their index fingers. Participants pressed "f" if AB is a word and "j"

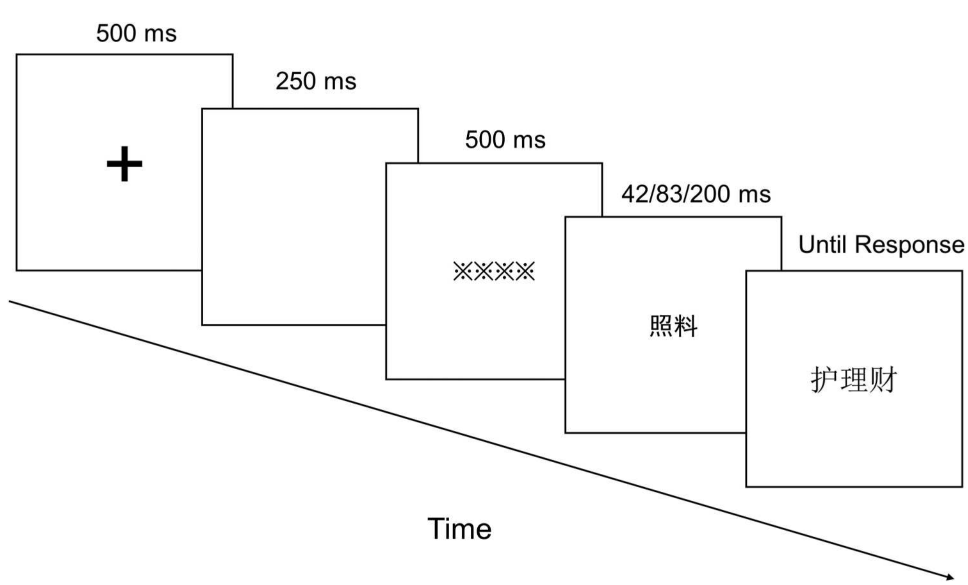

Figure 1. Example trial sequence of semantic priming condition.

if BC is a word. They were asked to respond as accurately as possible. According to previous research using the masked priming paradigm, the 42-ms and 83-ms SOA conditions could prevent participants from conscious awareness, although in the 83-ms condition, the prime was occasionally explicitly identified. In the 200-ms SOA condition, the prime would always be consciously identified. Following Rastle et al. (2000), we did not mention the presence of the prime to prevent conscious awareness in the first two conditions, while at the condition of 200 ms, we told the participants that there was a two-character word between the mask and the OAS.

The participants were randomly assigned to one of the six lists. The experimental trials were presented in a random order following 11 practice trials. Every participant read all of the 180 OASs once. There were three blocks with 60 trials each, and there was a one-minute break after practice and two breaks in the main experimental session.

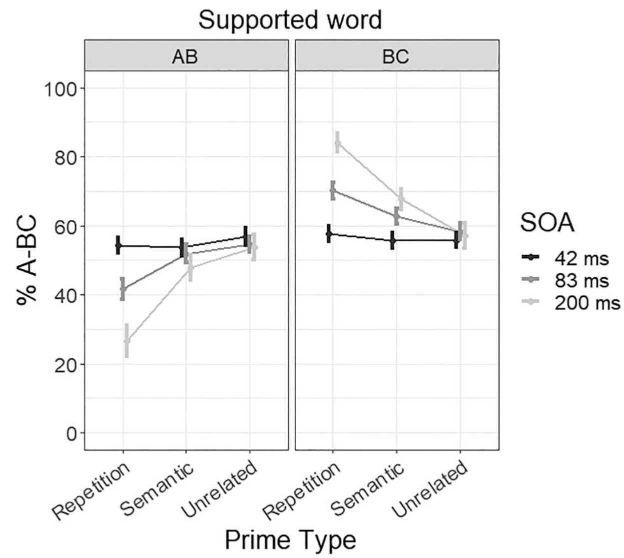

Figure 2. By-subject mean percentage of response A-BC. Note. The left panel shows the data under conditions when the word AB was primed. The right panel shows the data under conditions when the word BC was primed. The error bars represent the by-subject standard error of the means.

3. Results

The trials in which response latencies were shorter than 250 ms or longer than 2,000 ms were excluded from analyses. Trials with response latencies outside three standard deviations from the means of every condition of every participant were also excluded. In total, approximately 4.60% of trials were excluded. The rate of A-BC response choice across conditions is summarised in Figure 2. In addition to response choice, we also recorded RTs as a dependent variable. The RTs across conditions were summarised in Table 4.

3.1. Selection rate analysis

The most important outcome of the study is the result of OAS segmentation. We used the *glmer* function from the *lme4* package (Bates et al., 2015) to conduct generalised mixed-effects models (GLMM) and estimated the *p*-values using the *summary* function from the *lmerTest* package (Kuznetsova et al., 2017) in the R statistical package (R Development Core Team, 2020). The dependent variable of the models was the segmentation results (coding response A-BC as 1, AB-C as 0). Following Schad et al. (2020), we used planned contrasts to test specific hypotheses about the pattern of means directly in a single multiple regression model. Specifically, we used customised contrasts to directly test theoretically motivated hypotheses, including (a) two contrasts testing the main effects of SOA, comparing the 83-ms and 200-ms to the 42-ms SOA baseline condition, respectively; (b) three contrasts testing the effects of primed word under different SOA conditions, all of which compare primed BC condition against the baseline AB condition; (c) other contrasts testing whether the semantic or repetition conditions differ from the unrelated condition (see Table A1). As argued by Schad et al. (2020), directly testing hypotheses motivated contrasts in a single multiple regression model has some advantages over the ANOVA omnibus *F*-test. The *F*-test does not provide information about the source of a main effect or interaction, whereas planned comparisons can directly test hypotheses about means and avoid issues of multiple comparisons. Following Barr et al. (2013), we used the maximal model that could converge. We started with a maximal model, including all the random intercepts and slopes. As it failed to converge, we used a zero-correlation parameter model and discarded the random components that produced the smallest variances until the model converged (Bates et al., 2015). The random structure of the final model is presented in Appendix A2.

The fixed effects of the final model are shown in Table A1. The main effect of SOA did not reach significance. On average, the response choice did not differ between the 200-ms and 42-ms conditions, or between the 83-ms and 42-ms conditions. This outcome is reasonable, because the pattern of selection rate across different prime types was opposite when priming AB and BC (see Figure 2). Hence, there would be no main effect for SOA conditions. The effect of the primed word was not significant under the 42-ms SOA condition but was significant under the 83-ms and 200-ms SOA conditions. Specifically, participants chose A-BC significantly more frequently when the word BC was primed than when AB was primed under the 83-ms and 200-ms SOA conditions.

The critical effects of interests are the repetition and semantic priming effects. The priming effects here refer to the difference in selection rate between repetition/semantic and unrelated priming conditions. In the 42-ms SOA condition, regardless of whether the primed word was AB or BC, neither the repetition nor semantic priming effects approached significance. In the 83-ms SOA condition, the repetition priming effects were significant regardless of whether the primed word was AB or BC. However, the semantic priming effects were not significant when the primed word was AB, although the effect was close to significant when the primed word was BC. In the 200-ms SOA condition, both the repetition and semantic priming effects were significant.

To compare the effects at the largest SOA (200 ms) in terms of effect size, we converted the log odds ratio of the priming effects to the approximation of Cohen's d standardised effect size by multiplying the log odds ratio by $\frac{\sqrt{3}}{\pi}$ (Hasselblad & Hedges, 1995). The effect sizes of repetition priming at the 200-ms SOA condition for words AB and BC were − 0.74 and 0.87, respectively. The effect sizes of semantic priming at the 200-ms SOA condition for words AB and BC were − 0.14 and 0.29, respectively. Thus, the effect sizes were substantially larger for repetition priming than for semantic priming at the 200-ms SOA.

As stated earlier, most of the previous research in word segmentation demonstrated a left-side word advantage (Huang & Li, 2020; Li et al., 2009; Ma et al., 2014; Experiment 3). That is to say, when others are equal, the word on the left of the OAS is more likely to be segmented. However, Experiment 1 of Ma et al.'s (2014) study reflected a right-side word advantage. In their experiment, they also presented the OAS string in isolation and asked participants to fixate on the center as in our experiment. We now test if there was any segmentation bias in our study. We conducted by-item one-sample *t-tests* to examine any bias in word segmentation across the six neutral (unrelated) conditions, where the prime is assumed to not affect the OAS segmentation. We compared the A-BC selection rate under each unrelated condition to the chance level, i.e. a selection rate of 50%. Results showed that participants segmented the OAS as A-BC more frequently than the chance level under all six conditions (see Table 3). This pattern reveals a right-side word advantage, which we will return to in the Discussion section.

Table 3. Results of one-sample t test of A-BC selection rate in the unrelated conditions.

Primed word	SOA (ms)	*M*	*SE*	*t*(179)	*p*
AB	42	0.564	0.016	3.94	**<.001**
	83	0.547	0.018	2.55	**0.012**
	200	0.537	0.016	2.37	**0.019**
BC	42	0.558	0.016	3.73	**<.001**
	83	0.581	0.016	5.11	**<.001**
	200	0.572	0.016	4.56	**<.001**

Note: Significant effects are indicated in bold. The *A-BC* selection rates in each condition are compared to chance, i.e. a selection rate of 50%.

3.2. Reaction time analysis

For the RT data, we used the *lmer* function from the *lme4* package (Bates et al., 2015) to conduct linear mixed-effects models (LMM) and estimated the *p*-values using the *summary* function from the *lmerTest* package (Kuznetsova et al., 2017). The dependent variables of the models were RTs. Similar to segmentation analysis, we used planned contrasts to test specific hypotheses about the pattern of means. Similarly, we used customised contrasts to directly test theoretically motivated hypotheses, including (a) two contrasts testing the main effects of SOA; (b) three contrasts testing the effects of primed word under different SOA conditions; (c) 12 contrasts testing whether the semantic or repetition conditions differ from the unrelated condition. To gain the maximal model that could converge, we started with a model including all the random intercepts and slopes. Then we trimmed the random structure of the model, using the same method as segmentation analysis. The random effect structure of the final model is in Table A4.

The fixed effects of the final model are shown in Table A3. The critical effects of interests are the repetition and semantic priming effects. The pattern of results was similar to that of the response rate analyses. The priming effects here refer to the difference in RTs between repetition/semantic and unrelated priming conditions. In the 42-ms condition, neither the repetition nor semantic priming effects approached significance. In the 83-ms condition, the repetition priming effects were significant regardless of whether the primed word was AB or BC. However, the semantic priming effects were not significant, regardless of the primed word condition. In the 200-ms SOA condition, both the repetition and semantic priming effects were significant Table 4.

The main effect of SOA did not reach significance. RTs, on average, did not differ between the 200-ms and 42-ms, or between the 83-ms and 42-ms conditions. The

Table 4. Results of reaction times (ms) in the experimental conditions.

Condition		SOA					
		42 ms		83 ms		200 ms	
		M	*SE*	*M*	*SE*	*M*	*SE*
AB	Repetition	698	31	715	30	610	23
	Semantic	692	30	758	30	707	33
	Unrelated	694	30	761	31	739	37
BC	Repetition	703	32	700	32	600	25
	Semantic	703	32	748	30	698	32
	Unrelated	705	29	759	30	729	37

position of the primed word (i.e. which word in the OAS the prime paired with) affected RTs significantly under the 42-ms and 200-ms SOA condition, but not under the 83-ms SOA condition. Specifically, under the 42-ms SOA condition, the participants responded significantly faster when the prime was paired with word AB than with BC; under the 200-ms SOA condition, the participants responded significantly faster when the prime was paired with word BC than with AB. The faster responses observed when the prime was paired with word BC compared to AB in the 200-ms SOA condition are straightforward to interpret. Readers likely segmented the OASs as A-BC more readily when the prime corresponded to word BC. Additionally, responses made with the right hand are generally quicker than those made with the left hand, which may account for the response time patterns observed in the 200-ms SOA condition. However, the faster responses when the prime was paired with word AB compared to BC in the 42-ms SOA condition were unexpected. We do not have a clear explanation for this outcome. The 9 ms difference observed in this condition could potentially be a type I error, reflecting a statistical anomaly rather than a meaningful effect.

4. Discussion

One of the challenges for readers of unspaced writing systems is segmenting continuous characters into words during reading. In the present study, we used a primed OAS segmentation task to investigate whether semantic information is used during Chinese word segmentation. We manipulated the SOA between the prime and target to examine the time course of orthographic and semantic activation. For the OAS segmentation results on the response choice at the 200-ms SOA, we observed significant semantic and repetition priming effects, with repetition priming being larger than semantic priming. At the 83-ms SOA, we found a significant repetition priming effect; for semantic priming, there was only a non-significant trend when the primed word was BC ($p = .051$) and the effect was far from significant when the primed word was AB. At the 42-ms SOA, neither priming effect approached significance. We found a similar pattern in the latency data. At the 200-ms SOA, we observed both repetition and semantic priming effects; at the 83-ms SOA, we only observed a repetition priming effect; and at 42-ms SOA, neither repetition nor semantic priming was significant. In the next subsections, we examine the implications of the semantic priming effects on OAS segmentation and the development of models of Chinese reading, along with a brief examination of the role of the fixation position and ambiguous morphemic boundaries.

4.1. The repetition and semantic priming effect on OAS segmentation

We observed sizable semantic priming effects at the 200-ms SOA in the two dependent variables. Specifically, when the OAS was preceded by a semantically related prime rather than an unrelated prime, the participants segmented the primed word more frequently, and their responses were faster. Thus, the semantic information elicited by the prime facilitates the processing of a semantically related target word – the semantically related primes were not phonologically or orthographically related to the primed word. Since the prime provides evidence supporting the primed word within the OAS, the primed word will be more likely to win the competition between the two spatially overlapping words and does so more quickly. Indeed, the latency data showed that presenting a semantically related word preceding the target OAS makes word segmentation faster than presenting an unrelated word. Therefore, the current findings support the hypothesis that semantic information can affect Chinese word segmentation, at least at a 200-ms SOA. It should be noted that some degree of preactivation in semantically related primes is necessary to influence the target words, and that semantic feedback may require time to take effect. This explains why we only observed a significant semantic priming effect at the 200-ms SOA. At the shorter SOAs (e.g. 42 and 83 ms), there is not enough time to fully process the primed word semantically to affect the OAS segmentation decision. Further research may pinpoint the approximate prime duration necessary for semantic priming to occur in the primed OAS segmentation paradigm.

Note that unlike semantic priming, which only occurred at the 200-ms SOA, repetition priming occurred not just at the 200-ms SOA but also at the 83-ms SOA. It should be noted that the semantic priming effects at short SOAs are elusive and small even in the lexical decision task, so it is no surprise that the evidence here only occurred clearly at the 200 ms SOA (Fabre et al., 2007; Holcomb et al., 2005; Holcomb & Grainger, 2009). While the 83-ms prime duration may be sufficient for repetition primes to activate the character and lexical representations of the target stimuli, it may not have been sufficient to activate the semantic representation of the primed word in the OAS segmentation task.

In addition to the overall difference in time course between repetition and semantic priming, at the 200-ms SOA, the effect size was larger for repetition priming than semantic priming. This is again consistent with the findings in the word recognition literature, where repetition priming effects are larger and more stable than semantic priming effects (e.g. Tan & Yap, 2016). This phenomenon may have two sources. First, repetition primes facilitate not only the form representation but also the semantic component of the primed word. According to the entry-opening model of masked repetition priming (Forster & Davis, 1984), the prime not only opens the form-based entry for a word but also initiates the extraction of related semantic information. Second, regarding the semantic relatedness between the prime and the primed word, the repetition primes fully overlap with the primed word. In contrast, the semantic prime does not completely overlap with the target in meaning.

Finally, we did not find any priming effects (i.e. neither repetition nor semantic priming) at the 42-ms SOA. This appears to diverge from previous studies on visual word recognition, which can detect repetition priming effects at very short SOAs in tasks such as the lexical decision task (Forster & Davis, 1984; Luo et al., 2004). We believe that these differences may be attributed to the nature of the OAS segmentation task. This task requires a decision between different segmentation types, involving the activation of two-word units and the competition between them. This process of word competition may require additional time to segment them from the character string. In contrast, repetition priming effects in the lexical decision task mainly reflect the activation of individual units.

As reviewed in the Introduction, several studies have investigated how prior sentence context affects word segmentation in Chinese (Huang et al., 2021; Huang & Li, 2020). Our study goes beyond these studies because we directly examined the role of semantic information on Chinese word segmentation by excluding other confounding factors. Huang and Li (2020) only examined the role of syntactic information, and Huang et al.'s (2021) manipulation did not disentangle semantic and syntactic constraints. In contrast, we directly tested the role of semantic information using a primed word segmentation task with semantically related pairs. In a sentence reading scenario, it is difficult to distinguish which stage the context affects during sentence reading since the word segmentation and integration stages are difficult to disentangle. In contrast, the present experiment employed isolated OAS stimuli. As there is no demand for integration in this scenario, the results would reflect the word segmentation stage.

It is worth noting that some theorists have argued that the locus of the semantic priming effect is at a post-access stage in alphabetic orthographies (e.g. Neely et al., 1989). In this view, the participants would use the semantic relation after lexical access but before making their decision (e.g. in a lexical decision task). However, this hypothesis seems unlikely in a primed word segmentation task in Chinese such as the one used in the present experiment. Once the word in the OAS has been identified, it should have already been segmented. If semantic priming occurs after word identification, it should no longer affect word segmentation. However, semantic priming effects did emerge in our study. Therefore, it cannot be accounted for by post-lexical processes alone.

Nevertheless, we acknowledge that the semantic information in the OAS segmentation task may not be entirely equivalent to the semantic information encountered during natural sentence reading. Firstly, during sentence reading, semantic information is often provided by the entire prior context, which typically consists of more than one word, and is often intermixed with syntactic or pragmatic information to constrain the upcoming text. In the present study, semantic information was provided by a single semantically prime word, which may not fully capture the complexity and richness of semantic processing during sentence reading. Second, the semantic priming procedure may not function in the same manner as the prior semantic information during natural sentence reading. Primes were presented briefly and sometimes subliminally. However, in sentence reading, semantic information unfolds incrementally and dynamically as readers integrate information from multiple words and contexts. Therefore, while the present results offer insights into the effects of semantic information on word segmentation in Chinese, the experimental manipulation in the current experiment may not fully mirror the natural reading process.

4.2. Insights into models of Chinese word segmentation

The findings of the present study have important implications for models of word segmentation in Chinese reading. In this section, we present a potential modification to the CRM to account for semantic priming effects. To recapitulate the CRM, it is built based on the principles of the influential interactive activation model (McClelland & Rumelhart, 1981). The word processing module is composed of three levels of units: a visual perception level, a character level, and a word level. The between-level connections (feedforward and feedback links) can be inhibitory or excitatory, while all within-level connections are assumed to be inhibitory. To solve the word segmentation problem during reading, the model posits that spatially overlapping words compete for a single winner, which is then recognised and segmented simultaneously.

To accommodate semantic priming in our experiment, the CRM has to include a semantic component (Figure 3). Moreover, we need some extra assumptions to allow the prime to pre-activate the primed word. We can make some assumptions similar to those used in Stolz and Besner (1996) within the spirit of the interactive activation framework. First, the activation of the lexical units will not only feedforward to the corresponding semantic units but will also activate related semantic units, though less intensively. Second, the within-level competition does not eliminate the activation of weaker candidates, so the related semantic representation can remain activated and provide feedback to lower levels.

This extended version of the CRM can capture the observed semantic priming effect in the present study. For example, as illustrated in Figure 4, the presentation of a prime (e.g. 照料 [take care of]) would activate its character units and word unit. The word unit of the prime not only feeds forward to its semantic unit but also the

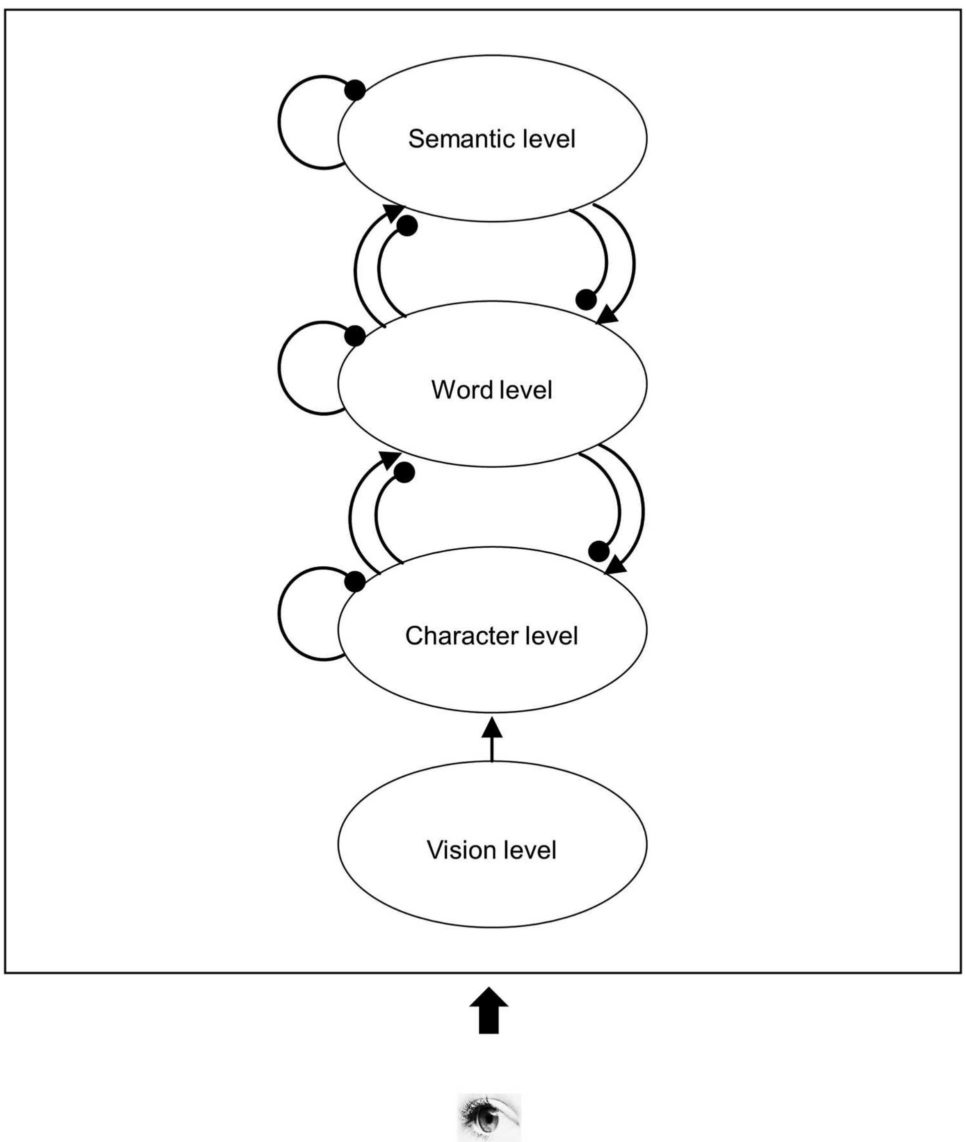

Figure 3. Proposed model of word segmentation with semantic processing.
Note. Arrows on lines indicate excitatory connections, while circles on lines represent inhibitory links.

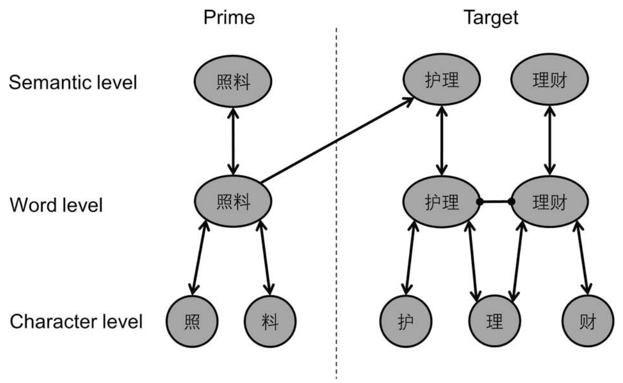

Figure 4. Semantic priming effect on word segmentation in the proposed model.
Note. Presentation of the prime (照料 [take care of]) activates the character and the word units. Then the activation feeds forward to the semantic units, including that of the primed word (护理 [nursing]). The semantic units provide feedback to the lower levels. Therefore, the primed word will have some advantage in the competition after the overlapping ambiguous string (护理财) appears. Arrows on lines indicate excitatory connections, while circles on lines represent inhibitory links.

semantically related units, including the semantic units of the primed word (e.g. 护理 [nursing]). Then, the semantic units provide feedback to the word and character levels. Thereby the word and character units of the primed word need less bottom-up processing to be fully activated when the OAS appears. Once the OAS (e.g. 护理财) appears, all the word units are activated and compete for a winner. Since the representations of the primed word have been preactivated in three levels, the word will be more likely to win the competition and, thus, more likely to be segmented as a word. As activation at the semantic level takes time, feedback from the semantic level to the word and character levels aiding OAS segmentation for semantically primed words may be negligible at very brief prime durations.

In the context of our study, one might be concerned about the possibilities of segmenting out the first (A) and the last character (C) first instead of AB or BC. These one-character words can also be activated and then compete with two-character words AB or BC. However, it is less likely to segment out A or C first instead of AB or BC. Based on CRM, all the words formed by the characters within the perceptual span are activated. When processing the OAS, words such as A, B, C, AB, and BC are all activated. Two-character words like AB and BC hold certain advantages over one-character words such as A and C, as they receive excitatory feedforward links from two character units, while one-character words only receive excitatory feedforward links from one character units. Thus, despite A and C having higher word frequencies than those of AB and BC, the latter hold higher priorities during the word competition.

4.3. The influence of fixation position on word segmentation

While not central for the purposes of our experiment, we observed a right-side word advantage. At first sight, this finding seems at odds with previous empirical evidence of a left-side word advantage in the studies of processing isolated character strings (Li et al., 2009), and also some studies in sentence reading (Huang & Li, 2020; Huang et al., 2021; Ma et al., 2014, Experiment 3). Nonetheless, the present study is consistent with Experiment 1 of Ma et al. (2014). We believe the differences across studies may be caused by the presentation method. In the present study, as in Ma et al.'s Experiment 1 (2014), participants had to fixate on the middle character position of the OAS centre before the OAS appeared. In their case, the OAS was presented briefly, and the participants had to choose the pronunciation name of the middle character, which is pronounced differently when it belongs to the left or the right word. Because readers have been familiar with reading from left to right, they may tend to allocate more attention to the words on the right side of fixation (see Liu et al., 2020). Therefore, the right side of the word may have some advantage when the eyes fixate on the middle character. In contrast, in Li et al.'s (2009) study showing a left-side word advantage, participants were asked to read the strings from the leftmost to the right. In this scenario, the left side of the word may have some advantage when the eyes fixate on the first character of the OAS. Future experiments should directly investigate this hypothesis (e.g. by manipulating the location of the initial fixation across blocks).

4.4. Ambiguous morphemic boundary in English

Although there are spaces between English words, similar segmentation problems also exist for ambiguous novel compounds. These are words that can be segmented in two ways. An example is *clamprod*, which can be segmented as *clam-prod* or *clamp-rod*. Research on these compounds has shown that both segmentation types are initially available (Libben, 1994, 2005; Libben et al., 1999). Libben (1994) found that lexical decision RTs for ambiguous novel compounds were longer than those for non-ambiguous novel compounds. Based on these findings, Libben argued that all legal lexical entries in a multimorphemic string are automatically activated during visual word recognition. Additionally, Libben et al. (1999) demonstrated that encountering ambiguous novel compounds aided in recalling related semantic associations of constituents, suggesting automatic activation of all potential constituents. In sum, these studies suggest that when processing English novel compounds with multiple parses, multiple constituents are available initially.

The concept that multiple initial parses are available aligns with the basic idea of the Chinese Reading Model (CRM), designed to simulate how Chinese readers segment words without spaces. According to CRM, all possible words formed by characters within the perceptual span are initially activated and compete for recognition. When a word prevails in this competition, it is both identified and segmented from the rest. This initial activation of all potential words in CRM parallels Libben's (2005) assumption that multiple parses are available at the outset. However, there is a notable difference: Libben suggests that all segmentation patterns, such as AB-C and A-BC, are initially available. In contrast, CRM posits that all potential words are activated, not all segmentation patterns. Only after a word wins the competition is the segmentation determined. For instance, when processing an ambiguous string like ABC, CRM activates all possible words: A, B, C, AB, and BC. Since AB and BC spatially overlap, they compete directly, with only one emerging as the winner. If AB wins, BC is inhibited. However, word C does not spatially overlap with AB and is therefore not inhibited by it. Consequently, both AB and C can be recognised simultaneously, leading to the segmentation of AB-C. From this perspective, A-BC does not emerge as an option during the process, differing from Libben's notion of multiple available parses.

5. Conclusion

The present study demonstrated that the semantic information from a prime stimulus affects a primed segmentation task with overlapping ambiguous strings (e.g. "客运营" can be segmented as "客运" [passenger transport] or "运营" [operation]), particularly at a 200-ms SOA. This finding underscores the role of semantic information in the word segmentation stage of Chinese word segmentation. Our findings are essential not only for understanding the dynamics of Chinese word segmentation but also for refining word segmentation models to incorporate semantic information. Furthermore, our study introduces a novel paradigm for exploring the intricacies of Chinese word segmentation.

Notes

1. This is the amount of information that can be effectively processed when the eyes fixate on a single position. Chinese reading perceptual span was measured by Inhoff and Liu (1998) using a gaze contingent moving window paradigm (McConkie & Rayner, 1976). They asked participants to read texts within the "window" around a fixation while masking the texts outside. By varying the size of the window, they determined the size of perceptual span as the minimum window size needed to maintain reading performance equivalent to natural reading.
2. It should be noted that although we controlled the properties of Characters A and C very well, the properties of character B was not controlled. As shown in Table 2, the character frequency and word frequency as a one-character word were higher and the number of strokes were smaller than the other two characters. Moreover, although all Characters A and C were words by themselves, 11 out of 180 Characters B were not words by themselves. We did not intentionally control the properties of Character B because it plays different roles in word segmentation according to CRM. According to CRM, when processing an OAS like ABC, CRM activates all possible words: A, B, C, AB, and BC. CRM assumes that words compete with each other when they are spatially overlapping. Therefore, word B competes with both AB and BC, and it has a very small chance to win the competition because it was supported by one character, while words AB and BC were supported by bottom-up activation from two characters. Moreover, because word B overlaps with both words AB and BC, they should affect AB and BC equally so that the properties of B should not affect the outcome word segmentations. Therefore, CRM assumes that the properties of Characters B does not influence word segmentation outcomes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by a grant from the National Natural Science Foundation of China (32371156). Linjieqiong Huang was supported by China Postdoctoral Science Foundation (2022M723362) and the Scientific Foundation of Institute of Psychology, Chinese Academy of Sciences (E2CX6625CX). Manuel Perea was supported by a grant from the Valencian Government (CIAICO/2021/172).

Data availability statement

The data, code, and materials from the present study are publicly available at the Psychological Science Data Bank (PsyDB): https://www.scidb.cn/anonymous/TmphUUZq

ORCID

Manuel Perea http://orcid.org/0000-0002-3291-1365

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv. https://doi.org/10.48550/arXiv.1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Chen, H.-C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin & Review*, *8*(3), 511–518. https://doi.org/10.3758/BF03196186

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.-F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. https://doi.org/10.1038/26967

de Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1062–1075. https://doi.org/10.1037/xlm0000074

Fabre, L., Lemaire, P., & Grainger, J. (2007). Attentional modulation of masked repetition and categorical priming in young and older adults. *Cognition*, *105*(3), 513–532. https://doi.org/10.1016/j.cognition.2006.10.011

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698. https://doi.org/10.1037/0278-7393.10.4.680

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, *117*(1), 167–178. https://doi.org/10.1037/0033-2909.117.1.167

Holcomb, P. J., & Grainger, J. (2009). ERP effects of short interval masked associative and repetition priming. *Journal of Neurolinguistics*, *22*(3), 301–312. https://doi.org/10.1016/j.jneuroling.2008.06.004

Holcomb, P. J., Reder, L., Misra, M., & Grainger, J. (2005). The effects of prime visibility on ERP measures of masked priming. *Cognitive Brain Research*, *24*(1), 155–172. https://doi.org/10.1016/j.cogbrainres.2005.01.003

Hsu, S.-H., & Huang, K.-C. (2000). Interword Spacing in Chinese Text Layout. *Perceptual and Motor Skills*, *91*(2), 355–365. https://doi.org/10.2466/pms.2000.91.2.355

Huang, L., & Li, X. (2020). Early, but not overwhelming: The effect of prior context on segmenting overlapping ambiguous strings when reading Chinese. *Quarterly Journal of Experimental Psychology*, *73*(9), 1382–1395. https://doi.org/10.1177/1747021820926012

Huang, L., Staub, A., & Li, X. (2021). Prior context influences lexical competition when segmenting Chinese overlapping ambiguous strings. *Journal of Memory and Language*, *118*, 104218. https://doi.org/10.1016/j.jml.2021.104218

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. https://doi.org/10.3758/BF03196544

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C.-S., Yap, M. J., Bengson, J. J., Niemeyer, D., & Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099–1114. https://doi.org/10.3758/s13428-012-0304-z

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 20–34. https://doi.org/10.1037/0096-1523.24.1.20

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, C.-Y., & Zhang, Y. (2018). Processing lexical and speaker information in repetition and semantic/associative priming. *Journal of Psycholinguistic Research*, *47*(1), 65–78. https://doi.org/10.1007/s10936-017-9514-y

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

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: a systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, *143*(2), 895. https://doi.org/10.1037/a0033580

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

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology*, *58*(4), 525–552. https://doi.org/10.1016/j.cogpsych.2009.02.003

Liang, F., Blythe, H. I., Bai, X., Yan, G., Li, X., Zang, C., & Liversedge, S. P. (2017). The role of character positional frequency on Chinese word learning during natural reading. *PLoS One*, *12*(11), e0187656. https://doi.org/10.1371/journal.pone.0187656

Liang, F., Blythe, H. I., Zang, C., Bai, X., Yan, G., & Liversedge, S. P. (2015). Positional character frequency and word spacing facilitate the acquisition of novel words during Chinese children's reading. *Journal of Cognitive Psychology*, *27*(5), 594–608. https://doi.org/10.1080/20445911.2014.1000918

Liang, F., Gao, Q., Li, X., Wang, Y., Bai, X., & Liversedge, S. P. (2023). The importance of the positional probability of word final (but not word initial) characters for word segmentation and identification in children and adults' natural Chinese reading.. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(1), 98–115. https://doi.org/10.1037/xlm0001116

Libben, G. (1994). How is morphological decomposition achieved? *Language and Cognitive Processes*, *9*(3), 369–391. https://doi.org/10.1080/01690969408402124

Libben, G. (2005). Everything is psycholinguistics: material and methodological considerations in the study of compound processing. *Canadian journal of linguistics/revue canadienne de linguistique*, *50*(1-4), 267–283. https://doi.org/10.1017/S000841310000373X

Libben, G., Derwing, B. L., & de Almeida, R. G. (1999). Ambiguous novel compounds and models of morphological parsing. *Brain and Language*, *68*(1-2), 378–386. https://doi.org/10.1006/brln.1999.2093

Liu, W., Inhoff, A. W., & Li, X. (2020). Attention shifting during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(9), 979–990. https://doi.org/10.1037/xhp0000755

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, *147*, 1–20. https://doi.org/10.1016/j.cognition.2015.10.013

Luo, Q., Peng, D., Jin, Z., Xu, D., Xiao, L., & Ding, G. (2004). Emotional valence of words modulates the subliminal repetition priming effect in the left fusiform gyrus: an event-related fMRI study. *NeuroImage*, *21*(1), 414–421. https://doi.org/10.1016/j.neuroimage.2003.09.048

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*(3), 1046–1059. https://doi.org/10.1037/a0035389

Marslen-Wilson, W. D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes*, *23*(3), 394–421. https://doi.org/10.1080/01690960701588004

McClelland, J. L. (2016). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 25–58). Erlbaum. https://doi.org/10.4324/9781315630427-11

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*(5), 365–368. https://doi.org/10.3758/BF03335168

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003–1019. https://doi.org/10.1037/0278-7393.15.6.1003

Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition*, *62*(2), 223–240. https://doi.org/10.1016/S0010-0277(96)00782-2

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, *66*(3), 180–194. https://doi.org/10.1007/s00426-002-0086-5

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(4-5), 507–537. https://doi.org/10.1080/01690960050119689

R Development Core Team. (2020). *R: A language and environment for statistical computing (Version 4.0.0) [Computer software]*. R Foundation for Statistical Computing. https://www.R-project.org

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Stolz, J. A., & Besner, D. (1996). Role of set in visual word recognition: Activation and activation blocking as nonautomatic processes. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1166–1177. https://doi.org/10.1037/0096-1523.22.5.1166

Tan, L. C., & Yap, M. J. (2016). Are individual differences in masked repetition and semantic priming reliable? *Visual Cognition*, *24*(2), 182–200. https://doi.org/10.1080/13506285.2016.1214201

Wong, A. W. K., Wu, Y., & Chen, H. C. (2014). Limited role of phonology in reading Chinese two-character compounds: Evidence from an ERP study. *Neuroscience*, *256*, 342–351. https://doi.org/10.1016/j.neuroscience.2013.10.035

Yen, M.-H., Radach, R., Tzeng, O. J. L., & Tsai, J.-L. (2012). Usage of statistical cues for word boundary in reading Chinese sentences. *Reading and Writing*, *25*(5), 1007–1029. https://doi.org/10.1007/s11145-011-9321-z

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in Reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1245–1265. https://doi.org/10.1037/0278-7393.26.5.1245

Zimmerman, R., & Gomez, P. (2012). Drawing attention to primes increases inhibitory word priming effects. *The Mental Lexicon*, *7*(2), 119–146. https://doi.org/10.1075/ml.7.2.01zim