Prior Context Influences Lexical Competition When Segmenting Chinese Overlapping Ambiguous Strings

Linjieqiong Huang[1,2], Adrian Staub[3], and Xingshan Li[1]

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China
[3]Department of Psychological and Brain Sciences, University of Massachusetts

Author Note

This research was supported by a grant from the National Natural Science Foundation of China (31970992). This work was also jointly funded by the National Natural Science Foundation of China (NSFC) and the German Research Foundation (DFG) in Project Crossmodal Learning, NSFC 61621136008/DFG TRR-169. The data, code and materials from the present study are publicly available at the Open Science Framework website: https://doi.org/10.17605/OSF.IO/2YF39

Correspondence should be addressed to Xingshan Li, 16 Lincui Road, Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China. Email: lixs@psych.ac.cn.

Abstract

We report three eye-movement experiments that investigated the effect of prior sentence context on the processing of overlapping ambiguous strings (OAS) during Chinese reading. An OAS is a Chinese character string (ABC) in which the middle character can form a distinct word with both the character on its left (word AB) and on its right (word BC). In three experiments, we manipulated the extent to which the right-side word (BC) was plausible as an immediate continuation following the prior context; the left-side word AB was always plausible given the prior context, and the sentence continued in a manner that was compatible with word AB. Compared with a less plausible word BC, first-pass reading times on the OAS were longer with a more plausible word BC. The results suggest that in reading of Chinese strings with ambiguous word boundaries, plausibility influences an early stage of competition between words, rather than only a later checking process that occurs after the initial segmentation.

Keywords: Chinese reading, eye movements, word segmentation, prior context, word competition

Prior Context Influences Lexical Competition When Segmenting Chinese Overlapping Ambiguous Strings

Unlike alphabetic languages like English, there are no inter-word spaces to explicitly indicate word boundaries in Chinese text. Successful comprehension requires Chinese readers to accurately segment continuous text into words. In recent decades, much has been learned regarding how Chinese readers segment words (Hsu & Huang, 2000a, 2000b; Huang & Li, 2020; Inhoff & Wu, 2005; Li et al., 2009; Ma et al., 2014; Ma et al., 2017; Perfetti & Tan, 1999; Yen et al., 2012). In some situations, word boundaries are ambiguous and character strings can be segmented in multiple ways. Ambiguity is one of the challenges in Chinese word segmentation, and how Chinese readers deal with ambiguous boundaries remains a question that needs further study.

One type of ambiguity occurs with an *overlapping ambiguous string* (OAS; Luo et al., 2002). An OAS is a string of characters (ABC) in which the middle character can form distinct words with the characters on both its left (word AB) and its right (word BC) (Gan et al., 1996; Hsu & Huang, 2000a, 2000b; Li et al., 2003; Yen et al., 2012). For example, in the OAS "选手套" (pronounced *xuan shou tao*), the first two characters constitute the word "选手" (meaning *player*), while the last two characters constitute another word "手套" (meaning *glove*). Thus, each OAS can be segmented into two structures: AB-C (e.g., 选手-套) or A-BC (e.g., 选-手套). In order to determine which structure is correct in a given sentence, Chinese readers need to use contextual information.

Previous studies have examined how Chinese readers process OASs (Huang & Li, 2020; Inhoff & Wu, 2005; Li et al., 2009; Ma et al., 2014; Ma et al., 2017; Perfetti & Tan, 1999). Perfetti and Tan (1999) proposed a two-character assembly strategy which assumes that Chinese readers prefer to segment Chinese characters as 2-character words because there are more 2-character words in the Chinese lexicon; according to the Lexicon of Common Words in Contemporary Chinese Research Team (2008), 72% of Chinese words (by type frequency) are 2-character words. Perfetti and Tan assume that the first two characters of the OAS have absolute priority during word segmentation, so that the middle character is usually assigned to the word on the left, and does not participate in the processing of the remaining words. However, this assumption was challenged by Inhoff and Wu (2005), who suggested that readers might not process characters in a strictly serial order, so that the middle character of the OAS is not always assigned to the word on the left. To support this argument, they designed a study in which participants read sentences containing a four-character string (ABCD) whose first two characters constituted a word (AB) and last two characters constituted another word (CD). In the ambiguous condition, the central two characters also constituted a two-character word (BC; e.g., in the string "专科学生", meaning *college student*, three words can be formed, "专科", "科学", and "学生"). In contrast, in the control condition the central two characters did not constitute a word (e.g., in the string "专科毕业", meaning *college graduation*, only two words may be formed, "专科" and "毕业"). First-pass reading times and total reading times were found to be longer in the ambiguous condition than in the control condition. Based on these findings, they proposed the *multiple activation hypothesis* to explain how Chinese readers segment words. According to this hypothesis, all the possible word candidates in the perceptual span are activated during reading. In the example above, the three words (i.e., “专科”, “科学” and “学生”) are all activated, so that readers need extra time to decide whether the character “科” belongs to “专科” or “科学”.

What would happen if all possible words in the perceptual span are activated? Ma et al. (2014) suggested that these words would compete with each other for a single winner. Once a word unit wins the competition, this word is identified and segmented from text. They referred to this as the *competition hypothesis*. In their study, the OASs could be segmented as either AB-C or A-BC. Disambiguating information consistent with one segmentation followed the OAS region. Meanwhile, they also manipulated the word frequencies of AB and BC to construct a high-low frequency condition (i.e., AB frequency was higher than BC frequency) or a low-high frequency condition (i.e., BC frequency was higher than AB frequency). Thus, segmentation based on word frequency was either consistent or inconsistent with the following context. They found that second-pass reading times were shorter and readers made fewer regressions to the OAS when the following context was consistent with a segmentation favoring the higher-frequency word. They concluded in favor of a competition hypothesis holding that any word in the perceptual span has a chance to win the competition if its activation is high enough, and word frequency is an important factor in determining which word wins the competition. They did, however, also conclude in favor of a left-side word advantage, as first-pass reading times were longer when the right-side word was higher in frequency, suggesting increased competition between the two possible segmentations. In Ma et al. (2014), as prior sentence context did not provide sufficient information to segment the OAS, readers could only use word frequency and the left-side advantage to make an initial segmentation. Later, when they read the disambiguating information, they might find that the initial segmentation was incorrect, so that they needed extra time or went back to the earlier part of the text to correct this initial segmentation.

Recently, Li and Pollatsek (2020) proposed a new computational model, the *Chinese Reading Model* (CRM), to account for how Chinese readers segment words and how they control their eye movements. The competition hypothesis was implemented within this model, on the basis of the interactive activation model (McClelland & Rumelhart, 1981), with some additional assumptions regarding how Chinese readers segment words. According to this model, all the characters in the perceptual span are activated, and all of the words formed by these characters are activated. The words that are spatially overlapping compete with each other, and only one of these words can win the competition. Once a word wins the competition, it is identified, and it is also segmented from text simultaneously. The model successfully predicts how Chinese readers segment overlapping ambiguous strings as reported by Ma et al. (2014) during first-pass reading.[1]

[1] The CRM model can simulate Ma et al.'s (2014) major finding that a high-frequency word is more likely to be segmented as a word during first-pass reading. However, because CRM does not have a semantic processing component, it does not simulate the finding regarding second-pass reading.

What happens when the prior context provides information that biases toward a particular segmentation of an OAS? Huang and Li (2020) addressed this question. In their informative condition, the prior contexts were constructed to generate syntactic constraints that supported either a left-word segmentation (AB-C) or a right-word segmentation (A-BC). In the neutral condition, the prior context did not provide any biasing syntactic information, while the post-target region disambiguated toward either left-word (AB-C) or right-word (A-BC) segmentation. They found that there were lower skipping rates, longer reading times on early measures (first fixation duration, gaze duration), and more regressions out on the OAS region in the informative A-BC condition than in the informative AB-C condition. However, no such first-pass differences between the AB-C and A-BC segmentation types were found in the neutral condition. In both the informative and neutral conditions, however, regression-in probabilities were lower in the AB-C condition than in the A-BC condition. These results suggest a general bias toward AB-C segmentation, but in addition an effect of prior context that operates rapidly enough to affect first pass reading.

Although Huang and Li (2020) showed that prior context has a rapid effect on word segmentation, it is not clear at which stage this effect is taking place. According to Li and Pollatsek (2020), when processing OASs all of the possible words in the perceptual span are activated and compete with each other for a single winner. The question is whether prior sentence context is involved in the competition and affects the results of initial segmentation. One possibility is that sentence context directly affects the initial stage of word segmentation. Alternatively, readers might not use sentence context information during the competition stage, but instead at a stage of evaluating whether the initial segmentation fits the sentence context. In the current study, we report three experiments investigating whether prior context exerts its effect during or after the word competition stage when segmenting OASs.

According to the view that context influences only a relatively late evaluation stage, readers initially segment the OAS based on word frequency and the left-side word advantage. Only when one of the words wins the competition will readers check whether the winning word fits the prior sentence context. We refer to this possibility as the *post-competition* hypothesis. A close analogy appears in the literature on lexical ambiguity resolution, where the integration model (Rayner & Frazier, 1989) makes a similar assumption that preceding context does not influence the lexical access stage, but rather the post-lexical integration stage by promoting the integration of context-related meaning. The order in which the meanings of an ambiguous word are accessed is only affected by the meaning frequency, and hence a high-frequency meaning will be accessed more quickly. When prior context favors the subordinate meaning, the initial process of integrating the dominant meaning fails, and the reader must then access the subordinate meaning, predicting longer reading times than when prior context favors the dominant meaning.

According to the *competitive constraint* hypothesis, on the other hand, all available information, including prior context, word frequency, and the left-side word advantage may impact the word competition stage itself, influencing how an OAS is initially segmented. The more information supports one word relative to an alternate word, the faster this word reaches activation and the faster the competition ends. In studies of lexical ambiguity resolution, the reordered access model (Duffy et al., 1988) proposed a similar view that preceding context information and meaning frequency both determine the access order of meanings. Specifically, the meaning supported by the preceding context is accessed faster, and the meaning that is higher in frequency is accessed faster. When the preceding context and meaning frequency cause two or more meanings to be available simultaneously, the competition between meanings will result in longer reading times.

In the present study, readers' eye movements were recorded as they read sentences containing OASs in three experiments. The correct segmentation of the OAS was always the left-word segmentation (AB-C), as confirmed by the following context. As we explain more fully below, the A-BC segmentation was actually not grammatical even at the point of reading the OAS itself, because on this analysis word A would not fit into the sentence syntactically or semantically. But nevertheless, both word AB and word BC might be activated and compete with each other as the reader processes the OAS. Our interest was whether prior context information could affect the word competition stage. Specifically, when prior context information supports word BC, does word BC compete more strongly with word AB?

Thus, the critical manipulation concerned the extent to which the right-side word (BC) was plausible as an immediate continuation following the preceding context, without word A. In the more plausible condition, both the left-side word (AB) and the right-side word (BC) of the OASs were plausible given the preceding context, while in the less plausible condition, word AB was plausible but word BC was less plausible. Previous studies have shown that when segmenting OASs, a higher frequency word will win the competition more often among activated words and word AB will have a processing advantage over word BC (Huang & Li, 2020; Ma et al., 2014). Thus, in Experiment 1, in order to reduce the processing advantage of the left-side word AB, the frequency of word BC was manipulated to be higher than that of word AB.

The two hypotheses we proposed above make different predictions regarding first-pass reading times in the OAS region. According to the post-competition hypothesis, readers would determine the initial segmentation based on word frequency and the left-side word advantage. Since the higher frequency word is the right-side word, neither word AB nor word BC has absolute priority and each word has a chance to win the competition. When one of the words wins the competition, the reader will check whether the winning word fits the prior sentence context. If this check happens rapidly enough to affect first-pass reading, then first-pass reading might be slowed when word BC is less plausible. On the other hand, if this check happens too late to affect first-pass reading, then first-pass reading time should not be affected at all by the plausibility of word BC. In neither case would the post-competition hypothesis predict that first-pass reading is slowed when word BC is *more* plausible.

The competitive constraint hypothesis predicts a different pattern of results. Based on this hypothesis, the plausibility of a word given the prior context is assumed to influence the word competition stage. In the condition in which both word AB and word BC are plausible, word AB is supported by the left-side word advantage, while word

BC is supported by word frequency; plausibility does not favor one word over the other. However, in the less plausible condition, word AB is also supported by prior context, but word BC is not, and hence word AB would presumably win after a quicker competition. In summary, the competitive constraint hypothesis predicts longer first-pass reading times on the OAS in the more plausible condition than the less plausible condition.

In Experiments 2 and 3, we manipulated word AB to have frequencies that were either higher or lower than word BC, to investigate whether the prior context influences word competition when the left-side word AB is also a higher frequency word. We will describe the logic of Experiments 2 and 3 in the respective sections below.

Experiment 1

Method

Participants. Forty participants (24 female and 16 male) were recruited to participate in Experiment 1. Given the number of trials in each condition, this yielded 1,480 total observations per condition, which is very close to the recommendation of Brysbaert and Stevens (2018) for well-powered within-subjects designs. All were native Chinese speakers and had normal or corrected-to-normal vision. Their ages ranged from 19 to 28 years (*M* = 22.15 years, *SE* = 0.38).

Apparatus. Participants' eye movements were recorded using an SR Research Eyelink 1000 eye-tracking system with a sampling rate of 1,000 Hz. The materials were presented on a 21-inch CRT monitor (resolution: 1024 × 768 pixels; refresh rate: 150 Hz) connected to a Dell PC. Each sentence was displayed on a single line in Song 20-

point font and the characters were shown in black (RGB: 0, 0, 0) on a gray background (RGB: 128, 128, 128). A chin rest and forehead rest were used to minimize head movement during the experiment. Participants were seated 58 cm from the computer screen; at this distance, one character subtended a visual angle of approximately 0.7°. For each participant, the viewing was binocular, but only the right eye was monitored.

Materials and Design. Seventy-four OASs were selected. The frequency data of all the items was obtained from the Lexicon of common words in contemporary Chinese (Lexicon of Common Words in Contemporary Chinese Research Team, 2008). The word frequency of the right-side word (BC; $M = 85.12$ occurrences per million, $SE = 12.01$, ranging between 11.04 and 755.46) was manipulated to be significantly higher than that of the left-side words (AB; $M = 17.43$ occurrences per million, $SE = 2.90$, ranging between 0.06 and 159.45, $t(73) = -5.48$, $p < .001$). Within each item, the frequency of the word BC was higher than that of the word AB, with a minimum difference of 8.04 occurrences per million. The stroke number did not differ significantly between the first ($M = 7.54$, $SE = 0.31$) and the third character ($M = 6.93$, $SE = 0.30$, $t(73) = 1.42$, $p = .157$). To examine how readers segment the OASs without context, we presented them in isolation to another 12 participants, and asked them to indicate where the word boundaries were. Results suggested that Chinese readers segmented the OASs as A-BC more frequently than chance level ($M = .58$, $SE = .04$, $t(73) = 2.05$, $p = .042$). These results indicate that a higher frequency word had an advantage of being segmented as a word, so that the right-side word was more likely to be segmented as a word in the experimental material.

Each OAS was embedded into two sentences (see Table 1 for examples). The left-word segmentation construction (AB-C) was plausible with the prior sentence context in all of the sentences. We manipulated the extent to which the word BC was plausible given the sentence context preceding the OASs. Specifically, the plausibility of word BC was manipulated as a potential continuation of the preceding context without word A. In the more plausible condition, both word AB and word BC were plausible in the preceding context, while in the less plausible condition, word AB was plausible but word BC was less plausible. Regardless of the initial plausibility of word BC, the correct segmentation was always the AB-C segmentation, as revealed by the following context. For example, in the more plausible condition, given a preceding context “张铭说他有” (meaning *Zhang Ming said he had*), both AB “带头” (meaning *taken the lead*) and BC “头发” (meaning *hair*) of the OAS were plausible. In contrast, for the less plausible condition, given a preceding context “有经验的张铭” (meaning *Zhang Ming who was experienced*), AB “带头” (meaning *took the lead*) was plausible but BC “头发” (meaning *hair*) was less plausible. But as shown in Table 1, the following context was only consistent with the meaning of *taken the lead*, regardless of the context before the OAS.

Importantly, adopting the A-BC segmentation of the OAS would actually result in an ungrammatical sentence even prior to encountering the following context, because word A cannot be integrated into the sentence. For the example in Table 1, word A (“带”), meaning *band* or *belt*, would not fit into the sentence if word BC (“头发”, *hair*) were, in fact, taken to be the continuation of the pre-A context. Thus, it would arguably be surprising, with these materials, for the A-BC segmentation to be entertained at all, let alone for this segmentation to sometimes win the initial competition. We return to this issue in discussing the results.

To measure whether the plausibility of word BC as an immediate continuation following the preceding context was manipulated effectively, we displayed the preceding part of the sentences to the left of the OAS region and word BC to 32 participants who did not participate in the main experiment and asked them to rate plausibility on a 7-point scale (1 = *very implausible*, 7 = *very plausible*). The plausibility values were significantly higher in the more plausible condition (M = 5.35, SE = 0.07, ranging between 3.81 and 6.63) than those in the less plausible condition (M = 3.46, SE = 0.10, ranging between 1.63 and 5.94, $t(73)$ = 15.40, $p < .001$). Within each item, the plausibility value in the more plausible condition was higher than that in the less plausible condition.

Another 32 Chinese speakers who did not participate in the eye-tracking experiment were recruited to assess the plausibility of the whole sentence. The plausibility values did not significantly differ by condition (more plausible: M = 5.72, SE = 0.08, ranging between 3.63 and 6.88; less plausible: M = 5.90, SE = 0.08, ranging between 2.81 and 6.75; $t(73) = -1.64$, p = .104). We also displayed the prior contexts and word AB to an additional 32 participants and asked them to rate their plausibility. No significant difference between the more plausible condition (M = 5.17, SE = 0.06, ranging 3.81–6.50) and the less plausible condition (M = 5.31, SE = 0.06, ranging 3.81–6.13) was found ($t(73) = -1.55$, p = .125). To measure predictability, we displayed the parts of the sentences to the left of the OAS region to 20 participants who did not participate in the main experiment and asked them to write down the words they predicted would come up next. The predictability of words A, AB, or BC in the OAS was 0, 0.01 and 0.01 in the more plausible condition, and 0, 0.01, 0 in the less plausible condition, respectively. The sentence length was comparable under different conditions (16.27 characters for the more plausible condition, 16.51 characters for the less plausible condition, $t(73) = -0.59$, $p = .555$).

Table 1

Examples of stimuli in Experiment 1

Condition	Instruction	Example/Translation
More plausible	Stimuli	张铭说他有**带头-发**传单给路人。
	The whole sentence	Zhang Ming said he had *taken the lead* in *handing out* leaflets to passers-by.
	Prior context + word AB	Zhang Ming said he had *taken the lead*…
	Prior context + word BC	Zhang Ming said he had *hair*…
Less plausible	Stimuli	有经验的张铭**带头-发**传单给路人。
	The whole sentence	Zhang Ming who was experienced *took the lead* in *handing out* leaflets to passers-by.
	Prior context + word AB	Zhang Ming who was experienced *took the lead*…
	Prior context + word BC	Zhang Ming who was experienced *hair*…

Note. The OASs are in bold and the hyphens are added for illustrative purposes, but the characters were not bolded or segmented in the experiment. The OASs, word AB and word BC in the corresponding translation are in italics.

Procedure. When participants came into the lab, they were given the experimental instructions and a brief description of the apparatus. The eye tracker was calibrated at the beginning of the experiment and again during the experiment as needed. A three-point calibration and validation procedure were used, and the maximal error of validation was below 0.5° in visual angle. Each sentence appeared after participants fixated on a character-sized box at the location of the first character of each sentence. Next, each participant read five sentences for practice, followed by 74 experimental sentences and 74 filler sentences in a random order. Participants were asked to read the sentences silently and to answer the questions following the sentences. After reading each sentence, they pressed a response button to start the next trial.

Results and Discussion

The mean accuracy of the comprehension questions was 92%, indicating that the participants understood the sentences well. Since eye blinks cause noise in eye-movement data, trials in which participants made more than three blinks while reading the entire sentence or made one or more blinks on the target word were excluded from analysis, resulting in the exclusion of 5.03% of the trials. Fixations with durations longer than 1,000 ms or shorter than 80 ms (approximately 0.34%) were also excluded from analyses. We primarily analyzed the following eye movement measures on the OAS region: (a) *first fixation duration* (the duration of the first fixation on the OAS region during the first-pass reading); (b) *first-pass reading time* (the summed duration of all first-pass fixations on the OAS region before moving on to another word); (c) *go-past time* (the summed duration starting when entering the OAS region until this region's right boundary is crossed); (d) *regression-in probability* (the percentage of regressions made back to the OAS region after leaving it).

Data were analyzed using *generalized mixed-effects models* (GLMMs). Generalized linear mixed-effect models (GLMM) avoid issues associated with the use of either raw or transformed latency measures in mixed-effects models; the models allow interpretation with respect to the original response scale, but also avoid the issue of non-Normal residuals (Lo & Andrews, 2015). For fixation duration measures, raw untransformed data and the gamma distribution were adopted. For regression-in probability, raw data and the binomial distribution were adopted. Plausibility was entered as a fixed effect, specifying the participants and items as crossed random effects, including intercepts and slopes (Baayen et al., 2008). Following Barr et al. (2013), we used the maximal model that could converge. We first constructed a model with a maximal random factor structure. When the maximal model failed to converge, we used a zero-correlation parameter model and dropped the random components that generated the smallest variances. The *glmer* function from the *lme4* package (Bates et al., 2015) was used within the *R* Environment for Statistical Computing (R Development Core Team, 2020). We report regression coefficients (*b*s, which estimate the effect size), standard errors (*SE*s), *t*-values (for durations), *z*-values (for binary dependent variables), and corresponding *p*-values. We estimated and reported the *p*-values for the effects using the *summary* function from the *lmerTest* package (Kuznetsova et al., 2017). Detailed eye movement measures and fixed-effect estimates from the GLMMs for all measures are shown in Table 2[2].

Table 2

Eye movement measures in the OAS region, and results of the generalized linear mixed-effects models in Experiment 1

Eye-movement	More plausible	Less plausible	*B*	*SE*	*t/z*	*p*

[2] We also analyzed first fixation duration and gaze duration on the BC region in all three experiments. The reliable plausibility effects of BC were observed in gaze duration, with patterns consistent with that of the OAS region, with the exception that the effect of plausibility reached significance only in the low-high frequency condition for gaze duration in Experiment 3.

measures						
First fixation duration	281 (5)	277 (6)	−4.13	5.76	−0.72	.474
First-pass reading time	569 (24)	521 (19)	**−48.28**	**9.06**	**−5.99**	**< .001**
Go-past time	809 (39)	719 (31)	**−94.55**	**13.75**	**−6.88**	**< .001**
Regression-in probability	.62 (.02)	.53 (.02)	**−0.38**	**0.09**	**−4.18**	**< .001**

Note. First fixation duration, first-pass reading time and go-past time were measured in milliseconds. Standard errors are given in parentheses. Significant effects are indicated in bold.

The effect of plausibility was not significant for first fixation duration, but was significant for all the other measures. Specifically, first-pass reading times and go-past times were significantly longer in the more plausible condition than in the less plausible condition. Compared with the less plausible condition, readers made more regressions into the OAS region in the more plausible condition.

Though for each pair of experimental sentences the plausibility of word BC was higher in the more plausible condition than that in the less plausible condition, the distributions of plausibility values were overlapping between the more plausible and less plausible conditions. To assess whether the effects of plausibility that we reported above are similar if plausibility is treated as a continuous variable, we conducted additional supplementary analyses with plausibility values as a (centered) continuous predictor, rather than as a two-level categorical predictor. As shown in Appendix A, the results were similar to what has been reported above.

The pattern of first-pass reading time could be easily accounted for by the competitive constraint hypothesis. Based on this hypothesis, the prior context influences competition during the word segmentation process. In the more plausible condition, words AB and BC were both plausible at the point they appeared. Thus, there was strong competition between them. In contrast, in the less plausible condition, word AB was plausible but word BC was less plausible. Hence, word AB would win the competition relatively quickly. The strong competition in the more plausible condition would result in longer first-pass reading times than in the less plausible condition.

On the contrary, the results appear to be inconsistent with the post-competition hypothesis. According to this hypothesis, either word AB or word BC has a chance to win the competition, based on the left-side word advantage (favoring word AB) and word frequency (favoring word BC). Since the prior context does not influence the word competition stage, a more plausible word BC would not increase first-pass reading times compared with a less plausible one. Indeed, any difference in first-pass reading times should take the form of longer reading times in the less plausible condition, as the post-competition hypothesis would predict that a less plausible word BC does sometimes win the competition, but is then difficult to integrate into the preceding context.

Readers also made more regressions into the OAS region in the more plausible condition than in the less plausible condition. This suggests that compared with the less plausible condition, readers were more likely to make an A-BC segmentation in the more plausible condition that was later contradicted by the following context, resulting in a regression into the OAS region to correct the initial segmentation.

We noted above that it might be unexpected for readers to entertain the A-BC segmentation at all; while word BC could be a plausible continuation of the pre-OAS context, word A would not fit into the resulting interpretation. The results from this experiment suggest, however, that not only did the A-BC segmentation initially compete with the AB-C segmentation, as shown by the effect of the plausibility of BC on first-pass reading time, but in addition it appears that readers sometimes did adopt this segmentation when word BC was more plausible, indicated by the increase in regressions back into the OAS in the more plausible condition. We return to a discussion of this pattern in the General Discussion.

Experiment 2

Experiment 1 suggested that the prior context exerts its influence during word competition when segmenting the OAS. The main goal of Experiment 2 was to investigate whether the prior context still influences word competition when the left-side word AB is a higher frequency word. Hence, in Experiment 2, the frequency of word AB was manipulated to be either higher or lower than that of word BC, yielding *high-medium* frequency and *medium-high* frequency conditions, respectively; we explain these condition names below. The stimuli in the medium-high frequency condition were similar to those in Experiment 1. Thus, we expected to replicate the findings in Experiment 1 for these conditions. In the high-medium frequency condition, the effect of plausibility might be smaller or even non-existent; because the frequency of word BC was lower, it may not be activated sufficiently for its plausibility to influence the competition.

Method

Participants. Forty-four participants (29 female and 15 male) were recruited to participate in Experiment 2, yielding 1,364 observations per condition. No participant had participated in Experiment 1. All were native Chinese speakers and had normal or corrected-to-normal vision. Their ages ranged from 18 to 28 years (*M* = 22.15 years, *SE* = 0.39).

Apparatus. The apparatus was identical to that used in Experiment 1.

Materials and Design. A total of 124 OASs were chosen as target words. For half of the OASs (the high-medium frequency condition), the frequency of word AB (*M* = 98.64 occurrences per million, *SE* = 15.93, ranging 5.76–679.96) was significantly higher than that of word BC (*M* = 16.82 occurrences per million, *SE* = 5.00, ranging 0.09–258.07; $t(61) = 4.90$, $p < .001$). Within each item, word frequency of word AB was higher than that of word BC, and the minimum difference was 4.22 occurrences per million. For the other half (the medium–high frequency condition), the frequency of word AB (*M* = 16.26 occurrences per million, *SE* = 2.43, ranging 0.06–83.32) was significantly lower than that of word BC (*M* = 83.38 occurrences per million, *SE* = 13.05, ranging 14.86–755.46; $t(61) = -5.06$, $p < .001$). Within each item, word frequency of word BC was higher than that of word AB, and the minimum difference was 9.06 occurrences per million. The condition names we use for this experiment are motivated by the fact that the means for the lower frequency words fall in the range that in the literature is referred to as medium-frequency (Hauk & Pulvermüller, 2004; Perea & Pollatsek, 1998), and by the need to distinguish these word frequencies from the even lower-frequency words used in Experiment 3.

The stroke number was matched between the first character and the third character for both the high-medium frequency condition ($t(61) = 0.001$) and the medium-high frequency condition (t(61) = 1.62, p = .108). As in Experiment 1, we asked 15 participants how they segmented the OASs without context, finding that in the high-medium frequency condition, Chinese readers segmented the OASs as AB-C more often than chance level (M = .73, SE = .03; t(61) = 9.00, p < .001), while in the medium-high frequency condition, participants segmented the OASs as A-BC marginally more often than chance level (M = .57, SE = .03; t(61) = −1.94, p = .058). This pattern was consistent with the word frequency contrast.

Each OAS was embedded into two sentences (see Table 3 for examples). As in Experiment 1, the context following the OAS indicated that the correct segmentation of the OAS was the left-word segmentation (AB-C). Also, as in Experiment 1, if the reader did initially adopt the A-BC segmentation, word A would have to be ignored due to its lack of syntactic or semantic fit. We manipulated the extent to which word BC was plausible given the preceding (pre-word A) context, resulting in a more plausible condition and a less plausible condition. Thus, the design was a 2 (plausibility: more plausible vs. less plausible) × 2 (word frequency: high-medium frequency vs. medium-high frequency) within-participant design.

As in Experiment 1, we conducted three norming studies to assess the plausibility of the sentences. A total of 108 native Chinese speakers who did not participate in the main experiment were recruited for these norming studies. They were assigned to one of two counterbalanced lists and were asked to rate their plausibility on a 7-point scale (1 = *very implausible*; 7 = *very plausible*). The results showed that the plausibility of the right-side words (BC) was rated significantly higher in the more plausible conditions than those in the less plausible conditions, and within each item, the plausibility value in the more plausible condition was higher than that in the less plausible condition. However, the plausibility of the left-side words (AB), the plausibility of the whole sentence, and sentence length were comparable across conditions (see Table 4 for details). We also asked 20 participants who did not participate in the main experiment to write down the words they predicted after reading the preceding contexts. The predictability of words A, AB, or BC in the OAS was close to zero (all were 0 in the more plausible high-medium frequency condition; 0, 0.01 and 0 respectively in the less plausible high-medium frequency condition; 0, 0.01 and 0 respectively in the more plausible medium-high frequency condition; all were 0 in the less plausible medium-high frequency condition).

Table 3

Examples of stimuli in Experiment 2

Condition	Instruction	Example/Translation
More plausible, high-medium frequency	Stimuli	他把不要的**东西-装**了满满一箱子。
	The whole sentence	He *packed* a box full of *what* he didn't want.
	Prior context + word AB	He took *what* he didn't want…
	Prior context + word BC	He took the *suit* he didn’t want…
Less plausible, high-medium frequency	Stimuli	他把要吃的**东西-装**了满满一袋子。
	The whole sentence	He *packed* a bag full of *what* he wanted to eat.
	Prior context + word AB	He took *what* he wanted to eat…
	Prior context + word BC	He took the *suit* he wanted to eat…
More plausible, medium-high frequency	Stimuli	张铭说他有**带头-发**传单给路人。
	The whole sentence	Zhang Ming said he had *taken the lead* in *handing out* leaflets to passers-by.
	Prior context + word AB	Zhang Ming said he had *taken the lead*…
	Prior context + word BC	Zhang Ming said he had *hair*…
Less plausible, medium-high frequency	Stimuli	有经验的张铭**带头-发**传单给路人。
	The whole sentence	Zhang Ming who was experienced *took the lead* in *handing out* leaflets to passers-by.

Prior context + word AB	Zhang Ming who was experienced *took the lead*…
Prior context + word BC	Zhang Ming who was experienced *hair*…

Note. The OASs are in bold and the hyphens are added for illustrative purposes, but the characters were not bolded or segmented in the experiment. The OASs, word AB and word BC in the corresponding translation are in italics.

Table 4

ANOVA results of norming studies and sentence length in Experiment 2

	High-medium frequency		Medium-high frequency		Plausibility	Word frequency	Interaction
	More plausible	Less plausible	More plausible	Less plausible			
Plausibility of the whole sentence	5.53 (0.08) Range: 3.67–6.44	5.65 (0.06) Range: 4.61–6.61	5.42 (0.07) Range: 4.28–6.44	5.55 (0.07) Range: 3.50–6.50	$F(1, 61) = 2.35$, $p = .130$	$F(1,61) = 2.59$, $p = .113$	$F(1, 61) = 0.002$, $p = .969$
Plausibility: prior context + word AB	5.12 (0.07) Range: 3.71–6.06	5.14 (0.06) Range: 4.00–6.00	5.06 (0.07) Range: 3.94–5.94	5.18 (0.08) Range: 3.59–6.41	$F(1, 61) = 1.06$, $p = .308$	$F(1, 61) = 0.02$, $p = .894$	$F(1, 61) = 0.66$, $p = .419$
Plausibility: prior context + word BC	5.19 (0.10) Range: 2.94–6.61	3.63 (0.10) Range: 1.39–5.00	5.12 (0.07) Range: 4.00–6.50	3.41 (0.09) Range: 2.22–5.39	$F(1, 61) = 277.46$, $p < .001$	$F(1,61) = 2.46$, $p = .122$	$F(1, 61) = 0.87$, $p = .355$
Sentence length	15.55 (0.3)	15.40 (0.3)	15.05 (0.3)	15.44 (0.3)	$F(1, 61) = 0.26$, $p = .609$	$F(1, 61) = 0.35$, $p = .557$	$F(1, 61) = 1.71$, $p = .196$

Note. There were 18 subjects per list. Standard errors are given in parentheses.

Procedure. The procedure was identical to that used in Experiment 1. Each participant read five sentences for practice, followed by 124 experimental sentences and 124 filler sentences in a random order. Participants were asked to read the sentences silently and to answer comprehension questions following approximately one-third of the sentences.

Results and Discussion

The mean accuracy on the comprehension questions was 95%, indicating that the participants understood the sentences well. Trials in which participants made more than three blinks while reading the entire sentence or made one or more blinks on the target word were excluded from analyses, resulting in the exclusion of 4.49% of the trials. Fixations with durations longer than 1,000 ms or shorter than 80 ms (approximately 0.45%) were also excluded from analysis.

The main question in this experiment is whether there is a plausibility effect in the high-medium frequency condition. Therefore, we used the generalized linear mixed-effects models to directly test theoretically motivated hypotheses using the following three customized contrasts (Schad et al., 2020): 1) assessing the effect of word frequency (high-medium frequency condition vs. medium-high frequency condition), 2) testing whether the two levels of plausibility differ significantly for the high-medium frequency condition, and 3) testing whether the two levels of plausibility differ significantly for the medium-high frequency condition. As in Experiment 1, when a maximal model failed to converge, we used a zero-correlation parameter model and dropped the random components that generated the smallest variances. Detailed eye-movement measures are shown in Table 5, and fixed-effect estimates from the GLMMs for all measures are shown in Table 6.

First Fixation Duration. First fixation durations in the high-medium frequency condition (*M* = 246 ms, *SE* = 4.87) were comparable to those in the medium-high frequency condition (*M* = 252 ms, *SE* = 4.73). No significant difference was found between the more plausible and less plausible conditions in either the medium-high frequency condition or the high-medium frequency condition.

First-pass reading time. First-pass reading times in the high-medium frequency condition (*M* = 374 ms, *SE* = 15.64) were shorter than those in the medium-high frequency condition (*M* = 422 ms, *SE* = 20.12). In addition, first-pass reading times were significantly longer in the more plausible than the less plausible condition in both the high-medium frequency condition and the medium-high frequency condition.

Go-Past Time. Go-past times were longer in the medium-high frequency condition (*M* = 542 ms, *SE* = 32.91) than in the high-medium frequency condition (*M* = 456 ms, *SE* = 25.51). In the high-medium frequency condition, no significant difference was found between the more plausible and less plausible conditions. But in the medium-high frequency condition, go-past times were significantly longer in the more plausible condition than the less plausible condition.

Regression-In Probability. In the high-medium frequency condition (*M* = 0.39, *SE* = 0.02), readers made fewer regressions into the OAS region than in the medium-high frequency condition (*M* = 0.48, *SE* = 0.02). Furthermore, there were lower regression-in probabilities in the less plausible condition than the more plausible condition in both the high-medium frequency condition and the medium-high frequency condition.

Table 5

Eye movement measures in the OAS region in Experiment 2

Eye-movement measures	High-medium frequency		Medium-high frequency	
	More plausible	Less plausible	More plausible	Less plausible
First fixation duration	250 (5)	244 (5)	253 (5)	251 (5)
First-pass reading time	386 (17)	362 (15)	443 (24)	399 (18)
Go-past time	460 (26)	451 (26)	574 (35)	510 (33)
Regression-in probability	.41 (.03)	.36 (.02)	.50 (.02)	.45 (.02)

Note. First fixation duration, first-pass reading time and go-past time were measured in milliseconds. Standard errors are given in parentheses.

Table 6

Results of the generalized linear mixed-effects models in Experiment 2

Eye-movement measures	Fixed effects	*B*	*SE*	*t/z*	*p*
First fixation duration	Word frequency	5.72	4.28	1.34	.181
	H-M: More plausible vs Less plausible	−6.02	4.71	−1.28	.201
	M-H: More plausible vs Less plausible	−1.47	4.30	−0.34	.733
First-pass reading time	Word frequency	**42.70**	**6.81**	**6.27**	**< .001**
	H-M: More plausible vs Less plausible	**−23.81**	**5.28**	**−4.51**	**< .001**
	M-H: More plausible vs Less plausible	**−34.32**	**5.50**	**−6.24**	**< .001**
Go-past time	Word frequency	**77.81**	**4.97**	**15.65**	**< .001**
	H-M: More plausible vs Less plausible	−7.14	5.94	−1.20	.230
	M-H: More plausible vs Less plausible	**−57.82**	**6.84**	**−8.45**	**< .001**
Regression-in probability	Word frequency	**0.40**	**0.12**	**3.41**	**< .001**
	H-M: More plausible vs Less plausible	**−0.24**	**0.09**	**−2.57**	**.010**
	M-H: More plausible vs Less plausible	**−0.25**	**0.09**	**−2.81**	**.005**

Note. Significant effects are indicated in bold. H-M = high-medium frequency condition; M-H = medium-high frequency condition.

As for Experiment 1, we also conducted additional analyses in which plausibility was treated as a continuous predictor (see Appendix A). The results of those analyses are consistent with the results presented above, with the exception that the effect of plausibility reached significance only in the first-pass reading time measure, but was again significant at both levels of frequency.

Experiment 2 replicated the main findings of Experiment 1 when word BC was higher in frequency than word AB: first-pass reading times were longer in the more plausible condition than the less plausible condition, suggesting that prior context influenced word competition as predicted by the competitive constraint hypothesis. The new finding is that when word BC was lower in frequency than word AB, the pattern was similar. First-pass reading times were longer in the more plausible condition than the less plausible condition. The results indicate that the effect of plausibility is not eliminated when word AB is higher in frequency than word BC.

We also found a significant effect of word frequency. Compared with the high-medium frequency condition, first-pass reading times were longer in the medium-high frequency condition. This result is consistent with the idea that when the left-side word advantage is supported by a difference in word frequency, word AB tends to win the competition with word BC especially quickly.

Finally, readers again made more regressions into the OAS region in the more plausible condition than in the less plausible condition, suggesting that a more plausible word BC sometimes won the initial competition, and subsequent context revealed this segmentation to be incorrect. There was also an effect of the frequency manipulation on regressions-in, suggesting that when word BC was higher in frequency, this also sometimes resulted in word BC winning the initial competition.

Before describing Experiment 3, we present a *post hoc* analysis of Experiments 1 and 2 that is relevant to the design of the following experiment[3]. As we have described, character A could be a word by itself in our design. It is possible, then, that the observed plausibility effect of word BC, with longer reading times when word BC is plausible given the pre-A context, is actually caused by differences in plausibility of word A, given the preceding context. Hence, we investigated the plausibility of word A in Experiments 1 and 2. We displayed the prior contexts and word A to another 80 participants who did not participate in the main experiment and asked them to rate the plausibility of the fragment on a 7-point scale (1 = *very implausible*, 7 = *very plausible*). The rated plausibility was indeed significantly different between the more plausible word BC and less plausible word BC conditions (for Experiment 1, $t(73) = -3.10$, $p = .002$; for Experiment 2: $F(1, 61) = 15.57$, $p < .001$). Compared with the less plausible condition (for Experiment 1, $M = 4.71$, $SE = 0.07$; for Experiment 2, $M = 5.04$, $SE = 0.07$ in the high-medium frequency condition, $M = 5.00$, $SE = 0.07$ in the medium-high frequency condition), the plausibility of word A was lower in the more plausible condition (for Experiment 1, $M = 4.42$, $SE = 0.06$; for Experiment 2, $M = 4.83$, $SE =$

[3] We thank one of the reviewers for emphasizing the potential role of word A's plausibility.

0.07 in the high-medium frequency condition, $M = 4.75$, $SE = 0.05$ in the medium-high frequency condition). Hence, the plausibility of word A was not perfectly controlled in Experiments 1 and 2. Though the differences between conditions in the plausibility of word A are quite modest, it is possible that the lower plausibility of word A led to longer first-pass reading times when word BC was more plausible in the pre-A context.

However, we then conducted analyses with the plausibility of word A as a (centered) continuous predictor (see Appendix B for detail). The patterns of results in Experiments 1 and 2 reported above did not change, suggesting that the observed effects are not likely caused by differences in the contextual plausibility of word A. We will further address this question in Experiment 3.

Experiment 3

In Experiment 2, we replicated the effect of the plausibility of word BC on first-pass reading time, and found that the effect of plausibility was still present when word AB was higher in frequency than word BC. In Experiment 3, we employed low and high frequency words to further investigate whether prior context affects word competition when the word frequency difference is larger.

Method

Participants. Forty-eight participants (30 female and 18 male) were recruited to participate in Experiment 3, yielding 1,344 observations per condition. None of them had participated in Experiments 1 or 2. All were native Chinese speakers with normal or corrected-to-normal vision. Their ages ranged from 19 to 26 years (*M* = 22.14 years, *SE* = 0.27).

Apparatus. The apparatus was identical to that used in Experiment 1.

Materials and Design. The design was identical to Experiment 2, except that we employed low-frequency and high-frequency words in this experiment. A total of 112 OASs were selected. As in Experiment 2, we manipulated the word frequency contrast of words AB and BC. In the high-low frequency condition, the left-side words (AB) were high-frequency words (*M* = 141.48 occurrences per million, *SE* = 11.37, ranging 50.39−439.77) and the right-side words (BC) were low-frequency words (*M* = 1.58 occurrences per million, *SE* = 0.11, ranging 0.09−2.91, $t(55) = 12.30$, $p < .001$). In the low-high frequency condition, the left-side words (AB) were low-frequency words (*M* = 1.44 occurrences per million, *SE* = 0.11, ranging 0.05−2.96), and the right-side words (BC) were high-frequency words (*M* = 167.39 occurrences per million, *SE* = 19.25, ranging 52.98−755.46, $t(55) = -8.62$, $p < .001$). The stroke number was matched between the first and third characters for both the high-low frequency condition ($t(55) = -1.60$, $p = .114$) and the low-high frequency condition ($t(55) = 1.24$, $p = .217$). As in Experiment 2, when the OASs were presented in isolation, Chinese readers (n = 15) segmented them as AB-C more often than chance level (*M* = .88, *SE* = .02, $t(55) = 18.99$, $p < .001$) in the high-low frequency condition, but as A-BC more often than chance level (*M* = .62, *SE* = .03, $t(55) = -3.96$, $p < .001$) in the low-high frequency condition. We note that with a large frequency disparity between the two words, there is now an apparent asymmetry in the segmentation bias, presumably reflecting the left-side word advantage; in the high-low frequency condition the segmentation contradicted the frequency difference only 12% of the time, while in the low-high frequency condition it did so 38% of the time.

We conducted four norming studies identical to those in Experiment 2, finding that the rated plausibility of the right-side word (BC) was significantly higher in the more plausible condition than the less plausible condition, and within each item, the plausibility value in the more plausible condition was higher than in the less plausible condition. Other factors did not result in differences in plausibility (see Table 7). The predictability of words A, AB, or BC in the OAS was close to zero (all were 0 in the more plausible high-low frequency condition; 0, 0.01 and 0 respectively in the less plausible high-low frequency condition; 0, 0.01 and 0 respectively in the more plausible low-high frequency condition; 0, 0.01 and 0 respectively in the less plausible low-high frequency condition), as indicated by the results for another 20 participants.

Unlike in Experiments 1 and 2, the plausibility ratings of word A were not significantly different across conditions in Experiment 3, as indicated in Table 7. Thus, this experiment provides a further opportunity to assess whether the observed plausibility effect of word BC is actually caused by differences in plausibility of word A.

Table 7

ANOVA results of norming studies and sentence length in Experiment 3

	High-low frequency		Low-high frequency		Plausibility	Word frequency	Interaction
	More plausible	Less plausible	More plausible	Less plausible			
Plausibility of the whole sentence	5.58 (0.06) Range: 4.65–6.53	5.56 (0.07) Range: 3.59–6.35	5.53 (0.06) Range: 4.71–6.35	5.55 (0.05) Range: 4.65–6.53	$F(1, 55) = 0.003$, $p = .958$	$F(1, 55) = 0.21$, $p = .649$	$F(1, 55) = 0.19$, $p = .664$
Plausibility: prior context + word AB	5.15 (0.06) Range: 3.89–6.17	5.28 (0.07) Range: 3.94–6.10	5.16 (0.05) Range: 3.85–5.72	5.24 (0.06) Range: 4.28–6.67	$F(1, 55) = 2.54$, $p = .117$	$F(1, 55) = 0.08$, $p = .779$	$F(1, 55) = 0.15$, $p = .697$
Plausibility: prior context + word A	4.76 (0.08) Range: 3.60–5.85	4.70 (0.08) Range: 3.35–5.85	4.57 (0.07) Range: 3.30–5.65	4.71 (0.09) Range: 3.10–6.35	$F(1, 55) = 0.17$, $p = .687$	$F(1, 55) = 1.04$, $p = .313$	$F(1, 55) = 2.03$, $p = .160$
Plausibility: prior context + word BC	5.42 (0.08) Range: 3.61–6.11	3.72 (0.10) Range: 2.17–5.78	5.38 (0.06) Range: 4.44–6.22	3.70 (0.11) Range: 2.28–5.56	$F(1, 55) = 508.09$, $p < .001$	$F(1, 55) = 0.09$, $p = .763$	$F(1, 55) = 0.001$, $p = .976$
Sentence length	15.61 (0.3)	15.75 (0.3)	15.64 (0.3)	15.70 (0.3)	$F(1, 55) = 0.34$, $p = .564$	$F(1, 55) = 0.001$, $p = .976$	$F(1, 55) = 0.06$, $p = .801$

Note. There were 18 subjects per list. Standard errors are given in parentheses.

Procedure. The procedure was identical to that used in Experiment 2. Each participant read five sentences for practice, followed by 112 experimental sentences and 112 filler sentences in a random order. Participants were asked to read the sentences silently and answer comprehension questions following approximately one third of the sentences.

Results and Discussion

The mean accuracy on the comprehension questions was 95%, indicating that the participants understood the sentences well. Trials in which participants made more than three blinks while reading the entire sentence or made one or more blinks on the target word were excluded from analyses, resulting in exclusion of 4.82% of the trials. Fixations with durations longer than 1,000 ms or shorter than 80 ms (approximately 0.34%) were also excluded from analyses. Eye movement measures and data analyses were identical to those in Experiment 2. Detailed eye movement measures are shown in Table 8, and fixed-effect estimates from the GLMMs are shown in Table 9 for all measures.

First Fixation Duration. First fixation durations in the high-low frequency condition (M = 250 ms, SE = 5.74) were shorter than in the low-high frequency condition (M = 259 ms, SE = 6.03). No significant difference was found between the more plausible and less plausible conditions in either the high-low frequency condition or the low-high frequency condition.

First-pass reading time. First-pass reading times were shorter in the high-low frequency condition (M = 380 ms, SE = 16.85) than in the low-high frequency condition

(*M* = 415 ms, *SE* = 18.23). First-pass reading times were significantly longer in the more plausible condition than in the less plausible condition in both the high-low frequency condition and the low-high frequency condition.

Go-Past Time. Go-past times were longer in the low-high frequency condition (*M* = 534 ms, *SE* = 25.48) than the high-low frequency condition (*M* = 463 ms, *SE* = 20.72). In the high-low frequency condition, no significant difference was found by plausibility condition. In the low-high frequency condition, go-past times were significantly longer in the more plausible condition than those in the less plausible condition.

Regression-In Probability. In the high-low frequency condition (*M* = 0.40, *SE* = 0.03), the regression-in probabilities were comparable to those in the low-high frequency condition (*M* = 0.43, *SE* = 0.02). In the high-low frequency condition, there was no significant difference by plausibility condition. But in the low-high frequency condition, readers made fewer regressions into the OAS region in the less plausible condition than in the more plausible condition.

Table 8

Eye movement measures in the OAS region in Experiment 3

Eye-movement measures	High-low frequency		Low-high frequency	
	More plausible	Less plausible	More plausible	Less plausible
First fixation duration	252 (6)	248 (6)	259 (7)	259 (6)
First-pass reading time	385 (19)	374 (16)	426 (20)	404 (17)
Go-past time	457 (22)	468 (21)	553 (31)	512 (23)
Regression-in probability	.40 (.03)	.39 (.03)	.46 (.03)	.40 (.02)

Note. First fixation duration, first-pass reading time and go-past time were measured in milliseconds. Standard errors are given in parentheses.

Table 9

Results of the generalized linear mixed-effects models in Experiment 3

Eye-movement measures	Fixed effects	*B*	*SE*	*t/z*	*p*
First fixation duration	Word frequency	**9.28**	**4.43**	**2.09**	**.036**
	H-L: More plausible vs Less plausible	−4.43	4.18	−1.06	.290
	L-H: More plausible vs Less plausible	0.004	3.88	0.001	.999
First-pass reading time	Word frequency	**37.18**	**6.13**	**6.06**	**< .001**
	H-L: More plausible vs Less plausible	**−13.09**	**6.38**	**−2.05**	**.040**
	L-H: More plausible vs Less plausible	**−26.59**	**7.54**	**−3.53**	**< .001**
Go-past time	Word frequency	**69.99**	**6.50**	**10.78**	**< .001**
	H-L: More plausible vs Less plausible	3.34	5.49	0.61	.543
	L-H: More plausible vs Less plausible	**−44.51**	**6.00**	**−7.42**	**< .001**
Regression-in probability	Word frequency	0.17	0.12	1.49	.137
	H-L: More plausible vs Less plausible	−0.07	0.09	−0.70	.483
	L-H: More plausible vs Less plausible	**−0.31**	**0.09**	**−3.34**	**< .001**

Note. Significant effects are indicated in bold. H-L = high-low frequency condition; L-H = low-high frequency condition.

For this experiment, we also conducted additional analyses in which plausibility was treated as a continuous predictor (see Appendix A). The results were similar to what has been reported above, with the exception that the effect of plausibility reached significance only in the low-high frequency condition for first-pass reading time.

The results are consistent with the previous experiments. For the low-high frequency condition, a more plausible word BC resulted in longer first-pass reading times than a less plausible one, which is consistent with the findings in Experiments 1 and 2. The regressions-in data also indicated that when word BC was higher in frequency than word AB, readers were more likely to make an incorrect segmentation in the more plausible condition compared with the less plausible condition, leading to the need to regress back into the word from later in the sentence. Moreover, for the high-low frequency condition where the left-side word advantage and a strong frequency combined to favor word AB, we also find a plausibility effect on first-pass reading time. Thus, even when word frequency difference is larger, prior context can influence the word competition stage.

In this experiment, the plausibility of word A was controlled across conditions, and we still found a similar effect of plausibility of word BC as in Experiments 1 and 2. Thus, these results further exclude the possibility that the observed effects are caused by differences in the contextual plausibility of word A.

Finally, we again found a significant effect of relative word frequency. First fixation duration and first-pass reading times were longer in the low-high frequency condition than that in the high-low frequency condition, suggesting that when the left-side word AB was high in frequency, it could win the competition quickly.

General Discussion

In the current study, three eye-tracking experiments were conducted to investigate whether the prior context exerts its effect after or during word competition when segmenting Chinese OASs.

Summary of Results

In Experiment 1, when words AB and BC were both plausible, first-pass reading times were significantly longer than when word BC was less plausible as an immediate continuation following the prior context. In addition, readers made more regressions into the OAS region when word BC was more plausible. In Experiment 2, the results replicated the main findings of Experiment 1 for the conditions in which word BC was higher in frequency than word AB, and found that when word AB was higher in frequency, the pattern was quite similar. First-pass reading times were longer and regression-in probabilities were higher when word BC was more plausible. We increased the word frequency difference by employing low-frequency and high-frequency words in Experiment 3. For the low-high frequency condition, a more plausible word BC resulted in longer first-pass reading times and higher regression-in probability on the OAS region than a less plausible one. For the high-low frequency condition, we also found an effect of plausibility on first-pass reading time.

The Effect of Prior Context on Word Competition

In the present study, we found that first-pass reading times were longer when words AB and BC were both plausible in the preceding context than when word BC was less plausible. This suggests that the plausibility of word BC influences the competition between word AB and word BC. To recapitulate the logic laid out in the Introduction, the competitive constraint account predicts longer first-pass reading times on the OAS when both words AB and BC are plausible because this increases the level of competition between the two words, only one of which can be selected. On the other hand, if context influences only a post-competition stage, the plausibility of word BC should either have no effect on first-pass reading times on the OAS (if first-pass reading time reflects only the duration of the competition stage itself) or a more plausible word BC should actually *decrease* first-pass reading times (if first-pass reading time reflects trials on which a less plausible word BC wins the competition, and then the implausibility is rapidly detected).

The finding that prior context exerts its effect during word competition is consistent with the one previous study that investigated how prior sentence context affects segmentation of an OAS (Huang & Li, 2020). They found that when the prior context supported AB-C segmentation, there were higher skipping rates and shorter first fixation durations than when prior context supported A-BC segmentation. Assuming that AB-C segmentation is also supported by the left-side word advantage, then contextual support for A-BC segmentation would have created more intense competition than contextual support for AB-C segmentation.

It is notable that in the current study, unlike in Huang and Li (2020), the context never fully supported A-BC segmentation; while word BC was, in one set of conditions, relatively plausible following the pre-OAS context, it was not possible to integrate both word A and word BC with the preceding context, as we have discussed above. Thus, the finding in the present study that a plausible word BC increased reading time on the OAS offers the intriguing suggestion that the word segmentation process in Chinese reading is sensitive to potentially plausible combinations of even non-adjacent words.

According to the CRM model proposed by Li and Pollatsek (2020), word frequency and the left-side advantage can have very early effects, but prior sentence context may have an effect somewhat later. Sentence context can only exert an effect when the activation of the words passes an initial threshold, which is distinct from the final threshold for selection. As a result, when the effects of other factors like word frequency and the left-side word advantage are overwhelming, the prior context effect may be overridden. However, in Experiment 3, where we employed quite low frequency BC words, we still saw an effect of the plausibility of word BC, though only on first pass reading times and not on regressions in. Thus, this aspect of the CRM model requires further investigation.

In the three experiments, readers made more regressions into the OAS region in the more plausible condition than in the less plausible condition (except, as just noted, in the high-low frequency condition of Experiment 3). It appears that compared with a less plausible word BC, a more plausible word BC was more likely to win the initial competition, with the resulting analysis later contradicted by the subsequent context. Thus, when readers read the post-target context in the more plausible condition, they were relatively likely to make more regressions into the OAS region to correct the initial segmentation. Again, this is arguably surprising, given that in order to interpret word BC as the continuation of the pre-word-A context, word A itself would have to be entirely ignored. More research is necessary to investigate whether readers did sometimes fully ignore word A, treating word BC as the continuation of the context until they encountered later material that contradicted this analysis.

Despite the fact that the A-BC segmentation was implausible given the preceding context, we still observed a plausibility effect of word BC. One possible mechanism for this effect is that the plausibility of all the activated words whose activations pass a threshold are taken into account during the competition stage of the word segmentation process. When the activation level of word BC reaches the threshold, the plausibility of word BC can influence the competition, so that word BC is more likely to be segmented as a single word when it is plausible than implausible with the pre-OAS context.

Because the competition happens before the OAS is segmented, the plausibility of word BC affects the competition even though it is not contiguous with pre-OAS context. After the OAS is initially segmented, readers will further integrate the A-BC or AB-C segmentation with preceding context. Only at that stage would readers find out that the A-BC segmentation is not plausible.

The Early Effect of Plausibility in Chinese Reading

The early effect of plausibility on competition between words produced by different segmentation of characters is consistent with previous studies of Chinese reading, and contrasts with some previous studies of English reading. In Staub et al. (2007), the left constituent of a spaced English compound (e.g., *cafeteria manager*) was manipulated to be either plausible or implausible as a head noun at the point where it appeared, whereas the compound as a whole was always plausible. For example, after *The new principal talked to the,* the noun *cafeteria* is implausible, while the compound *cafeteria manager* is plausible; but after *The new principal visited the,* either continuation is plausible. When the head noun analysis of the left constituent was implausible, reading times on this word increased, beginning with the first fixation. This result suggests that the implausibility of the head noun analysis did not prevent readers from initially adopting this analysis, in preference to the (correct) compound analysis.

On the other hand, Yang et al. (2012) employed a similar design in Chinese reading. In their experiment, the verb before either a one- or two-character target word was manipulated so as to affect the plausibility of the following noun. When the target word was one character, a robust plausibility effect was observed. When the target word was two characters, this word was always plausible, but the verb manipulation resulted in the first character being either plausible or implausible as a separate word. For example, a two-character word meaning *gatekeeper*, whose first character means *door*, was preceded either by the verb *kicked* (making either *door* or *gatekeeper* plausible) or *entreated* (making *door* implausible, but *gatekeeper* plausible). The critical result was that there was no plausibility effect in the comparison of these conditions; the implausibility of the analysis on which the first character was a separate word appears to have prevented that analysis from being adopted. Thus, the Yang et al. study provides independent evidence that for Chinese readers, plausibility has a very early effect on word segmentation; while English readers initially adopt an analysis on which, e.g., *the principal talked to the cafeteria*, rather than *the cafeteria manager*, there is no indication that Chinese readers adopt a similar analysis.

This cross-linguistic difference may be understood as reflecting a difference in the degree of parallelism of lexical processing across the perceptual span. Models of eye movements when reading spaced, alphabetic orthographies such as English (Engbert et al., 2005; Reichle et al., 1998; Reichle et al., 2009) have debated whether multiple words are processed simultaneously, but recent behavioral and neuroscientific evidence (White, et al., 2018; White et al., 2019) has suggested that for readers of English, truly parallel processing of multiple words may be impossible. On the other hand, the demands of efficiently processing Chinese orthography, in which identifying word boundaries is a critical task, and must be accomplished early in the stream of processing, seems to lead to a great deal of parallelism in activation of the multiple, partially overlapping words that are visible in the perceptual span (Li et al., 2009; Li & Pollatsek, 2020). Moreover, it now appears that contextual information affects the competition between words.

In addition to the demands of lexical segmentation, the difference in length between a Chinese OAS and an English compound word may play a role in this cross-linguistic difference. The OAS consists of three characters and can fall within foveal vision. Thus, parallel processing may be relatively easy in this situation. However, a long English compound (e.g., “cafeteria manager”) is not easily read in a single fixation, and hence parallel processing of multiple words may not be feasible.

Even though multiple words are not likely processed in parallel in English reading, there is evidence that the semantic information of part of a word and the semantic information of the whole word are processed in parallel. Bowers et al. (2005) presented participants with target words that contain embedded words (i.e., *hatch*, which contains *hat*) or that are embedded words within longer words (i.e., *bee*, part of *beer*), and asked them to perform a semantic categorization task for the target words. In the congruent condition, the embedded or embedding word was associated with the same response, while in the incongruent condition, it was associated with a conflicting response. They found that the semantic categorization of targets was slower and less accurate when a higher-frequency embedded or embedding word was associated with a conflicting response. These semantic congruence effects suggested that the semantic representations of the whole word and its embedded word are both activated during the course of identifying non-compound words. Thus, it appears that English readers do engage in semantic processing of orthographic substrings in words, which provides a parallel to the results of the current study with Chinese readers.

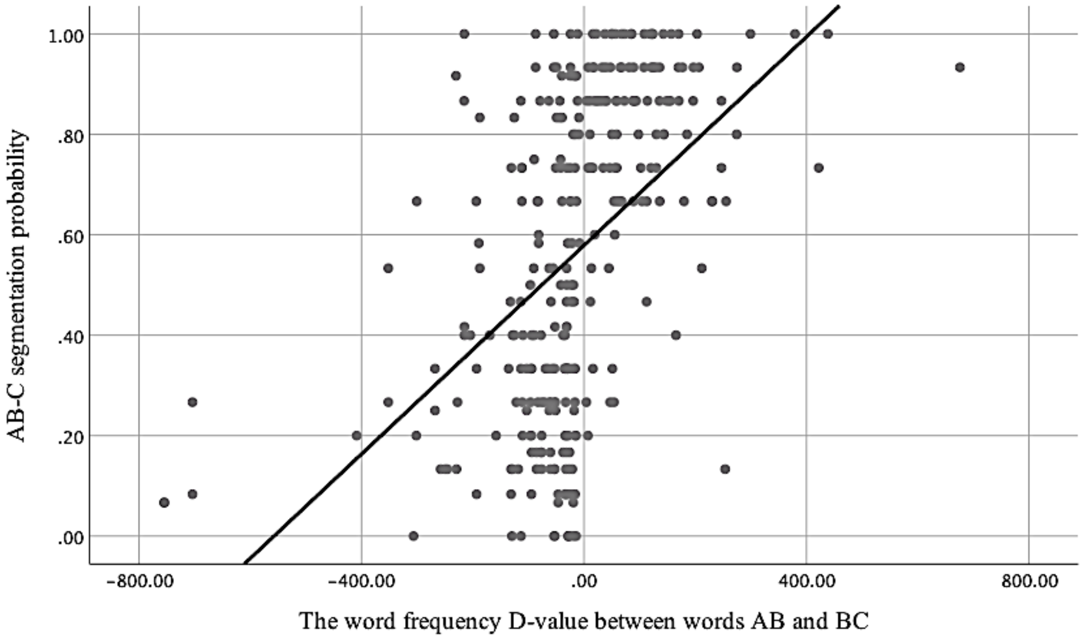
The early effect of plausibility observed in the current study for Chinese reading is consistent with findings showing an early effect of plausibility in English reading. Some studies have adopted the gaze-contingent boundary paradigm to investigate whether there is a plausibility preview benefit in English reading (e.g., Brothers & Traxler, 2016; Schotter & Jia, 2016; Veldre & Andrews, 2017, 2018a, 2018b). The results showed that compared with implausible previews, when previews are plausible continuations of the sentence, there were higher skipping rates and longer first-pass reading times on the target word. Thus, it appears that readers do process the semantic fit of a parafoveal word in the sentence, at least to some degree.

The Bias of the Overlapping Ambiguous String Itself

In these three experiments, when we prepared the material, we also tested participants' segmentation of OASs that were presented in isolation (see the "Materials and design" section for details). We found that in both the high-medium and high-low frequency conditions, Chinese readers segmented the OASs as AB-C more often than chance level, while the opposite occurred in the medium-high and low-high frequency conditions. We conducted a correlation analysis for all items in three experiments, and found that the relative frequencies of words AB and BC were correlated with AB-C segmentation probabilities (r = .49, p < .01; see Figure 1). There is a clear tendency to segment an OAS to make a higher-frequency two-character word, and this bias increases as the frequency difference increases.

Figure 1

Relationship between the relative frequencies of the words AB and BC and the segmentation bias across three experiments



Note. Each dot represents an individual item. Word frequency D-value = word frequency of AB minus word frequency of BC.

In addition, there is also a general left-side word advantage. In Figure 1, this is evident in the fact that the intercept of the regression line is not at .5, but at about .6; an OAS with equally matched frequencies between word AB and BC is predicted to be segmented as AB-C about 60% of the time. We assume this is due to the direction of reading. Since Chinese reading is from left to right, there might be more visual attention initially allocated to the characters on the left than the right. When the first and second characters can constitute a word, readers might tend to segment the first two characters as words.

Implications for Readability of Text

It should be noted from the present study that contextual information can sometimes also lead to segmentation errors for Chinese readers. Thus, a well-written text should avoid causing potential segmentation errors. When there is segmentation error, the text should provide clues for readers to detect and correct the error as soon as possible.

Conclusion

In summary, the results of three experiments in the current study clearly support the competitive constraint hypothesis, suggesting that prior context influences word competition when processing overlapping ambiguous strings. Plausibility given prior context, word frequency, and the left-side word advantage each provide support for an AB-C segmentation or an A-BC segmentation. The results provide further evidence that plausibility has a strong and early influence on word segmentation processes in Chinese reading.

Supplementary Material

The data, code and materials from the present study are publicly available at the Open Science Framework website: https://doi.org/10.17605/OSF.IO/2YF39

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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., 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

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a “hat” in “that”? *Journal of Memory and Language*, *52*(1), 131–143. https://doi.org/10.1016/j.jml.2004.09.003

Brothers, T., & Traxler, M. J. (2016). Anticipating syntax during reading: Evidence from the boundary change paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1894–1906. https://doi.org/10.1037/xlm0000257

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446. https://doi.org/10.1016/0749-596x(88)90066-6

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. https://doi.org/10.1037/0033-295x.112.4.777

Gan, K. W., Palmer, M., & Lua, K. T. (1996). A statistically emergent approach for language processing: Application to modeling context effects in ambiguous Chinese word boundary perception. *Computational Linguistics*, *22*(4), 531–553.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090–1103. https://doi.org/10.1016/j.clinph.2003.12.020

Hsu, S.-H., & Huang, K.-C. (2000a). Effects of word spacing on reading Chinese text from a video display terminal. *Perceptual and Motor Skills*, *90*(1), 81–92. https://doi.org/10.2466/pms.2000.90.1.81

Hsu, S.-H., & Huang, K.-C. (2000b). 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

Inhoff, A. W., & Wu, C. (2005). Eye movements and the identification of spatially ambiguous words during Chinese sentence reading. *Memory & Cognition*, *33*(8), 1345–1356. https://doi.org/10.3758/bf03193367

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

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

Li, M., Gao, J., Huang, C., & Li, J. (2003, July 11–12). *Unsupervised training for overlapping ambiguity resolution in Chinese word segmentation* [Paper presentation]. Proceedings of the Second SIGHAN Workshop on Chinese Language Processing, Sapporo, Japan. https://doi.org/10.3115/1119250.1119251

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*. Advance online publication. 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

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, Article 1171. https://doi.org/10.3389/fpsyg.2015.01171

Luo, X., Sun, M., & Tsou, B. K. (2002). Covering ambiguity resolution in Chinese word segmentation based on contextual information. In *Proceedings of the 19th international conference on computational linguistic* (Vol. 1., pp. 1–7). Association for Computational Linguistics.

https://doi.org/10.3115/1072228.1072283

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

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*(1), 158–166. https://doi.org/10.1037/xlm0000298

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

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 767–779. https://doi.org/10.1037/0096-1523.24.3.767

Perfetti, C. A., & Tan, L. H. (1999). The constituency model of Chinese word identification. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 115–134). Lawrence Erlbaum Associates Publishers.

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

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 779–790. https://doi.org/10.1037/0278-7393.15.5.779

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. https://doi.org/10.1037/0033-295x.105.1.125

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21. https://doi.org/10.3758/pbr.16.1.1

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*, Article 104038. https://doi.org/10.1016/j.jml.2019.104038

Schotter, E. R., & Jia, A. (2016). Semantic and plausibility preview benefit effects in English: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *42*(12), 1839–1866. https://doi.org/10.1037/xlm0000281

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-

noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1162–1169. https://doi.org/10.1037/0278-7393.33.6.1162

Veldre, A., Andrews, S. (2017). Parafoveal preview benefit in sentence reading: Independent effects of plausibility and orthographic relatedness. *Psychonomic Bulletin & Review*, *24*(2), 519–528. https://doi.org/10.3758/s13423-016-1120-8

Veldre, A., Andrews, S. (2018a). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. *Journal of Memory and Language*, *100*, 1–17. https://doi.org/10.1016/j.jml.2017.12.002

Veldre, A., Andrews, S. (2018b). How does foveal processing difficulty affect parafoveal processing during reading? *Journal of Memory and Language, 103*, 74–90. https://doi.org/10.1016/j.jml.2018.08.001

White, A. L., Palmer, J., & Boynton, G. M. (2018). Evidence of serial processing in visual word recognition. *Psychological Science*, *29*(7), 1062–1071. https://doi.org/10.1177/0956797617751898

White, A. L., Palmer, J., Boynton, G. M., & Yeatman, J. D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proceedings of the National Academy of Sciences*, *116*, 10087–10096. https://doi.org/10.1073/pnas.1822137116

Yang, J., Staub, A., Li, N., Wang, S., & Rayner, K. (2012). Plausibility effects when reading one- and two-character words in Chinese: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*(6), 1801–1809. https://doi.org/10.1037/a0028478

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

Appendix A. Supplementary Analyses With Plausibility of Word BC

Experiment 1

Plausibility was entered as a (centered) continuous predictor, specifying the participants and items as crossed random effects, including intercepts and slopes. When a maximal model failed to converge, we used a zero-correlation parameter model and dropped the random components that generated the smallest variances. Fixed-effect estimates from the GLMMs for all measures are shown in Table A.1.

The effect of plausibility was not significant for first fixation duration. First-pass reading times and go-past times were longer when plausibility values were larger. Moreover, readers made more regressions into the OAS region when plausibility values were larger.

Table A.1

Results of the generalized linear mixed-effects models in Experiment 1

Eye-movement measures	*b*	*SE*	*t/z*	*p*
First fixation duration	1.30	1.66	0.79	.431
First-pass reading time	**12.51**	**4.61**	**2.71**	**.007**
Go-past time	**30.66**	**5.37**	**5.71**	**<.001**
Regression-in probability	**0.17**	**0.04**	**4.03**	**<.001**

Note. Significant effects are indicated in bold.

Experiment 2

We used generalized linear mixed-effects models to directly test theoretically motivated hypotheses using the following three customized contrasts (Schad et al., 2020): 1) assessing the effect of word frequency (high-medium frequency condition vs. medium-high frequency condition), 2) testing whether the effect of plausibility was significant for the high-medium frequency condition, and 3) testing whether the effect of plausibility was significant for the medium-high frequency condition. When a maximal model failed to converge, we used a zero-correlation parameter model and dropped the random components that generated the smallest variances. Plausibility was entered as a (centered) continuous predictor. Fixed-effect estimates from the GLMMs for all measures are shown in Table A.2.

First Fixation Duration. First fixation durations in the high-medium frequency condition (*M* = 246 ms, *SE* = 4.87) were shorter than those in the medium-high frequency condition (*M* = 252 ms, *SE* = 4.73). No significant effect of plausibility was found in either the high-medium frequency condition or the medium-high frequency condition.

First-pass reading time. First-pass reading times in the high-medium frequency condition (*M* = 374 ms, *SE* = 15.64) were shorter than those in the medium-high frequency condition (*M* = 422 ms, *SE* = 20.12). Additionally, first-pass reading times were significantly longer when plausibility values were larger in both the high-medium frequency condition and the medium-high frequency condition.

Go-Past Time. Go-past times were longer in the medium-high frequency condition (*M* = 542 ms, *SE* = 32.91) than in the high-medium frequency condition (*M* = 456 ms, *SE* = 25.51). Moreover, no significant effect of plausibility was found in either the high-medium frequency condition or the medium-high frequency condition.

Regression-In Probability. In the high-medium frequency condition (*M* = 0.39, *SE* = 0.02), readers made fewer regressions into the OAS region than that in the medium-high frequency condition (*M* = 0.48, *SE* = 0.02). No significant effect of plausibility was found in either the high-medium frequency condition or the medium-high frequency condition.

Table A.2

Results of the generalized linear mixed-effects models in Experiment 2

Measures	Fixed effect	*b*	*SE*	*t/z*	*p*
First fixation duration	Word frequency	5.61	4.25	1.32	.187
	H-M: Plausibility	2.42	1.52	1.59	.112
	M-H: Plausibility	0.66	1.54	0.43	.669
First-pass reading time	Word frequency	**49.57**	**7.79**	**6.36**	**< .001**
	H-M: Plausibility	**6.48**	**3.52**	**1.84**	**.066**
	M-H: Plausibility	**11.03**	**3.59**	**3.07**	**.002**
Go-past time	Word frequency	**89.34**	**8.08**	**11.06**	**< .001**
	H-M: Plausibility	10.08	7.56	1.33	.182
	M-H: Plausibility	−5.02	7.61	−0.66	.510
Regression-in probability	Word frequency	**0.41**	**0.12**	**3.52**	**< .001**
	H-M: Plausibility	0.13	0.10	1.28	.200
	M-H: Plausibility	0.09	0.14	0.65	.516

Note. Significant effects are indicated in bold. H-M = high-medium frequency condition; M-H = medium-high frequency condition.

Experiment 3

Data analyses were identical to those in Experiment 2 (in the Appendix section). Fixed-effect estimates from the GLMMs are shown in Table A.3 for all measures.

First Fixation Duration. First fixation durations in the high-low frequency condition (*M* = 250 ms, *SE* = 5.74) were shorter than in the low-high frequency condition (*M* = 259 ms, *SE* = 6.03). No significant difference was found between the plausible and less plausible conditions in either the high-low frequency condition or the low-high frequency condition

First-pass reading time. First-pass reading times were shorter in the high-low frequency condition (*M* = 380 ms, *SE* = 16.85) than in the low-high frequency condition (*M* = 415 ms, *SE* = 18.23). No significant plausibility effect was found in the high-low frequency condition. But in the low-high frequency condition, first-pass reading times were significantly longer when plausibility values were larger.

Go-Past Time. Go-past times were longer in the low-high frequency condition (*M* = 534 ms, *SE* = 25.48) than the high-low frequency condition (*M* = 463 ms, *SE* = 20.72). In addition, no significant plausibility effect was found in the high-low frequency condition. But in the low-high frequency condition, go-past times were significantly longer when plausibility values were larger.

Regression-In Probability. In the high-low frequency condition (*M* = 0.40, *SE* = 0.03), the regression-in probabilities were comparable to those in the low-high frequency condition (*M* = 0.43, *SE* = 0.02). In the high-low frequency condition, there was no significant effect of plausibility. However, in the low-high frequency condition, readers made more regressions into the OAS region when plausibility values were larger.

Table A.3

Results of the generalized linear mixed-effects models in Experiment 3

Measures	Fixed effect	*b*	*SE*	*t/z*	*p*
First fixation duration	Word frequency	**9.23**	**4.37**	**2.11**	**.035**
	H-L: Plausibility	1.08	1.51	0.71	.476
	L-H: Plausibility	1.57	1.55	1.01	.312
First-pass reading time	Word frequency	**36.99**	**4.98**	**7.42**	**< .001**
	H-L: Plausibility	3.96	2.98	1.33	.183
	L-H: Plausibility	**7.38**	**3.04**	**2.42**	**.015**
Go-past time	Word frequency	**70.35**	**7.06**	**9.96**	**<.001**

	H-L: Plausibility	−5.61	4.40	−1.28	.202
	L-H: Plausibility	**12.70**	**4.03**	**3.15**	**.002**
Regression-in probability	Word frequency	0.18	0.12	1.51	.132
	H-L: Plausibility	0.01	0.04	0.22	.828
	L-H: Plausibility	**0.20**	**0.05**	**4.11**	**< .001**

Note. Significant effects are indicated in bold. H-L = high-low frequency condition; L-H = low-high frequency condition.

Appendix B. Supplementary Analyses With Plausibility of Word A

We conducted additional supplementary analyses with plausibility value of word A as a (centered) continuous predictor (see Table B.1, B.2 for details). It indicated that the patterns of results in Experiments 1 and 2 were consistent with that of original models.

Table B.1

Results of the generalized linear mixed-effects models in Experiment 1

Eye-movement measures	Fixed effects	*B*	*SE*	*t/z*	*p*
First fixation duration	Plausibility	–4.02	5.09	–0.79	.429
	Rating (prior context + A)	–0.37	4.52	–0.08	.935
First-pass reading time	Plausibility	**–43.04**	**9.24**	**–4.66**	**< .001**
	Rating (prior context + A)	**–18.63**	**8.59**	**–2.17**	**.030**
Go-past time	Plausibility	**–80.01**	**7.19**	**–11.13**	**< .001**
	Rating (prior context + A)	**–52.13**	**8.70**	**–5.99**	**< .001**
Regression-in probability	Plausibility	**–0.38**	**0.13**	**–2.89**	**.004**
	Rating (prior context + A)	–0.07	0.14	–0.50	.617

Note. Significant effects are indicated in bold.

Table B.2

Results of the generalized linear mixed-effects models in Experiment 2

Eye-movement measures	Fixed effects	*B*	*SE*	*t/z*	*p*
First fixation duration	Word frequency	5.04	4.10	1.23	.219
	H-M: More plausible vs Less plausible	–6.16	4.34	–1.42	.156
	M-H: More plausible vs Less plausible	–0.66	4.87	–0.14	.892
	Rating (prior context + A)	–3.17	2.62	–1.21	.225
First-pass reading time	Word frequency	**41.97**	**6.43**	**6.53**	**< .001**
	H-M: More plausible vs Less plausible	**–19.89**	**7.52**	**–2.65**	**.008**
	M-H: More plausible vs Less plausible	**–29.46**	**5.50**	**–5.36**	**< .001**
	Rating (prior context + A)	**–19.56**	**5.53**	**–3.54**	**< .001**
Go-past time	Word frequency	**89.02**	**5.61**	**15.88**	**< .001**
	H-M: More plausible vs Less plausible	**–10.47**	**5.33**	**–1.97**	**.049**
	M-H: More plausible vs Less plausible	**–64.06**	**5.80**	**–11.04**	**< .001**
	Rating (prior context + A)	**–16.05**	**4.76**	**–3.38**	**< .001**
Regression-in probability	Word frequency	**0.39**	**0.12**	**3.31**	**< .001**
	H-M: More plausible vs Less plausible	**–0.19**	**0.09**	**–2.05**	**.040**

M-H: More plausible vs Less plausible	**–0.20**	**0.09**	**–2.13**	**.033**
Rating (prior context + A)	**–0.21**	**0.08**	**–2.70**	**.007**

Note. Significant effects are indicated in bold. H-M = high-medium frequency condition; M-H = medium-high frequency condition.