Monosyllabic focus verbs disrupt reading fluency in Mandarin: evidence from eye-tracking

Meiyuan Zhang ● *

Chinese Studies, Centre for Languages and Literature, Lund University, Lund, Sweden

Focus is a core component of information structure that highlights the most prominent element in a sentence. While pitch and duration are well-established prosodic markers of focus in Mandarin Chinese, the role of word length has received less attention. Due to historical developments, many Mandarin words exhibit elastic length, appearing in both monosyllabic and disyllabic forms. In modern Chinese, however, there is a strong prosodic preference for disyllabic words as the minimal prosodic unit. This study tested whether using monosyllabic verbs in focus position disrupts reading fluency due to prosodic mismatch. Thirty-seven native Mandarin speakers read sentences silently while their eye movements were recorded. The study employed a 2 × 2 factorial design that crossed focus status (focus vs. no focus) with word length (monosyllabic vs. disyllabic). Linear mixed-effects models were used to analyze gaze duration, first fixation duration, first pass duration, regression path duration, regression count, fixation count, and skipping probability. The results show that monosyllabic verbs in focus positions attracted longer gaze durations, more fixations, and more regressions than disyllabic verbs, indicating a processing cost linked to prosodic mismatch. These findings reveal how prosodic and information-structural cues jointly guide real-time reading and confirm the processing advantage of disyllabic verbs in focus contexts.

KEYWORDS

focus, word length, mandarin, eye-tracking, reading fluency, monosyllabic verbs, disyllabic verbs, prosodic mismatch

1 Introduction

Focus, a core concept in information structure, refers to the most prominent element in a sentence — the part that updates shared knowledge or signals contrast among alternatives (Lambrecht, 1994). Cross-linguistically, languages use different strategies to encode focus, including prosodic cues [e.g., English pitch expansion (Gussenhoven, 1983); Selkirk (1985), syntactic means (e.g., Hungarian clefting; (Kiss, 2015)], and morphological markers (e.g., Wolof, 2024). These differences show that focus realization depends on language-specific prosodic and syntactic systems that shape sentence comprehension.

In Mandarin Chinese, focus is traditionally described as marked mainly through syntactic strategies and particles, with prosody playing a secondary role (Lee and Sun, 2024). However, recent psycholinguistic research challenges this view, showing that prosodic cues, such as pitch range and duration can be as effective, or more so, than syntactic clefting for marking focus in spoken and silent reading (Yan and Calhoun, 2019; Ouyang and Kaiser, 2015; Arnhold, 2024). Importantly, studies of implicit prosody suggest that readers mentally assign a prosodic contour to written text while reading silently (Breen and Clifton, 2011; Fodor, 2002; Stolterfoht et al., 2007). Building on this, the present study

tests whether word length, specifically, the contrast between monosyllabic and disyllabic forms, functions as an additional cue to focus on Mandarin reading. In this study, we treat word length as a prosodic factor rather than a purely morphological property. As in Mandarin, disyllabic forms typically correspond to the preferred minimal prosodic unit, whereas monosyllabic forms are prosodically lighter.

The impact of syllable number, often referred to as the word length effect, is robustly documented in eye-tracking and reading research across languages. Classic studies (Rayner, 1998, 2009; Zang et al., 2018) show that longer words generally demand greater cognitive resources, resulting in longer fixation durations and lower skipping rates. These effects are typically linked to increased phonological and morphological complexity (Ferrand and New, 2003; Balota et al., 2004). In Mandarin, research has further shown that plausibility and context interact with word length, shaping how readers process one- and two-character words (Yang et al., 2012). Neuroimaging studies using ERP and fMRI have likewise demonstrated that longer words engage additional visual and phonological processing (Hauk and Pulvermuller, 2004; Schuster et al., 2016). Given that focus prominence during language production often relies on prosodic cues such as pitch expansion, it remains an open question whether longer words, words with more syllables, might also provide similar advantages for focus processing during reading.

Although the impact of word length on general reading time is well-established, its specific role in focus processing has received little empirical attention. This question is particularly relevant for Mandarin, where about 60–70% of words can appear in either monosyllabic or disyllabic forms that are semantically equivalent (Huang and Duanmu, 2013). Choosing between these alternatives often poses persistent challenges for L2 learners.

Mandarin has been described as a syllable-timed language with a strong prosodic bias for disyllabic units (Duanmu, 2007; Shi, 2023). Corpus analyses show that while monosyllabic words are highly frequent, disyllables dominate the range of unique word types, supporting their role as default prosodic carriers (Stolterfoht et al., 2007). This asymmetry suggests a possible trade-off: monosyllables are frequent and efficient but may be less distinctive as carriers of prosodic prominence due to their short duration and high homophony (Cai and Brysbaert, 2010; Karlgren, 1918).

Recent findings indicate that readers do integrate prosodic cues when predicting focus during silent reading (Li and Thompson, 1981). In Mandarin, a disyllabic word may match readers' prosodic expectations for focused elements, while an unexpected monosyllable in a focus position may trigger additional processing costs due to this mismatch.

In this study, we test whether Mandarin readers show a preference for disyllabic forms as prosodic carriers of focus. Specifically, we ask whether monosyllabic words appearing in focus positions lead to longer reading times and more regressions compared to disyllabic words, and whether these effects are reduced when the word is not in focus. By combining word length and focus manipulations with eye-tracking, we aim to clarify how prosodic templates shape online sentence processing in Mandarin.

2 Materials and methods

2.1 Participants

Thirty-seven native Mandarin speakers (28 female, mean age = 24.5 years, SD = 2.2) were recruited among Lund university international students. All participants had normal or corrected-to-normal vision and no reported reading disorders. Each participant received a small monetary compensation for their participation

Written informed consent was obtained from all participants prior to the study, which was approved by the Swedish Ethical Review Authority [2025-03192-01].

2.2 Materials and design

The stimulus materials consisted of a set of target words embedded in natural sentence contexts, manipulated according to two experimental factors: Focus (*Focus* vs. *No Focus*) and Word Length (*Disyllabic* vs. *Monosyllabic*). This resulted in a 2 × 2 design, crossing information-structural status with word length.

Eighty pairs of semantically equivalent verbs (e.g., 学 *xué* vs. 学习 *xuéxí* "to study"; 印 *yìn* vs. 复印 *fùyìn* "to copy"; 开 *kāi* vs. 打开 *dǎkāi* "to open"; 帮 *bāng* vs. 帮忙 *bāngmáng* "to help) were selected from a class of so-called elastic words as critical words (Dong, 2015). Elastic words are defined as word pairs consisting of a short (monosyllabic) and a long (disyllabic) form that share the same morpheme root and convey the same meaning. These forms are widely regarded as morphological variants of a single lexical item (Huang and Duanmu, 2013; Cai and Brysbaert, 2010; Karlgren, 1971; Guo, 1938; Chao, 1948). In this study, elastic verb pairs were selected as they provide an ideal test case for elastic length variation while minimizing potential confounds from rhythmic patterns that can arise with other word classes during sentence processing (Luo et al., 2010; Luo and Zhou, 2010; Long et al., 2022). Moreover, only verbs were included because prior research has demonstrated distinct processing profiles for nouns and verbs in Mandarin Chinese, with verbs offering clearer comparability in prosodic and syntactic behavior (Xia et al., 2016).

Verb pair selection was guided by two key criteria to ensure comparability and experimental control. First, semantic equivalence was verified: 80 candidate verb pairs were initially selected, each consisting of free morphemes that could stand alone in sentence contexts. Twenty native speakers then confirmed that each pair conveyed the same meaning. Second, frequency equivalence was assessed using log-transformed values from the SUBTLEX-CH corpus (Cai and Brysbaert, 2010). A paired-samples t-test showed that monosyllabic verbs were significantly more frequent than their disyllabic counterparts, $t_{(76)} = -8.14$, $p < 0.001$; therefore, word frequency was included as a covariate in all statistical analyses to control for this difference. All 80 pairs were used in the experiment, but three pairs with missing frequency data were excluded from the final statistical models, resulting in 77 pairs in the reported analyses.

Each verb pair was embedded in a short discourse passage corresponding to one of the four experimental conditions. The passages were adapted from the Peking University Corpus (Peking University Center for Chinese Linguistics, 2023). The critical verb

(monosyllabic or disyllabic) always appeared in sentence-final position within the passage, but not at the very end, in order to maintain syntactic consistency and align with Mandarin's default focus position. In all cases, the verbs occurred without overt objects, which is fully natural in Mandarin: some verbs are inherently intransitive (e.g., 睡觉 *shuìjiào* "to sleep"), while others permit conventional ellipsis (e.g., 学 *xué* in 自习室里边太吵了, 我去图书馆学 "The study room is too noisy; I'll go study in the library." Where the object of study is left implicit). The absence of an overt object therefore does not imply intransitive usage.

Focus assignment was manipulated across conditions. In the focused conditions, the target verb was discourse-new, with the disyllabic form in Condition A and the monosyllabic form in Condition B. In the unfocused conditions, the preceding context rendered the verb discourse-given, with the disyllabic form in Condition C and the monosyllabic form in Condition D. This design allowed us to test word-length effects while holding discourse structure constant. To preserve coherence and prevent unintended focus, each passage also included a natural-sounding follow-up sentence. Sentence length ranged from 24 to 72 characters.

The full stimulus set comprised 320 target sentences, evenly distributed across the four conditions (see Table 1 for an example). Sentence presentation was counterbalanced across participants to ensure equal exposure to all conditions. Each participant read a total of 120 sentences, including 40 filler sentences, presented in a randomized order across four sessions. To encourage attention to meaning and sentence-level processing, participants were asked to read aloud a subset of stimuli at the end of each session.

An independent group of 37 university students provided naturalness ratings for all stimulus sentences prior to the main experiment. To test whether naturalness differed across conditions, we fit a linear model to the ratings. This analysis showed that word length had a significant effect on perceived naturalness, with monosyllabic words rated as less natural than disyllabic words [$b = -0.67$, $SE = 0.03$, $t_{(2915)} = -24.12$, $p < 0.001$]. To control for this variation in perceived naturalness, the mean naturalness rating for each sentence was included as a covariate in a separate linear mixed-effects models analyzing the eye-tracking data (see Table 2).

2.3 Apparatus

The experiment was conducted at The Humanities Lab Digital Classroom (Niehorster et al., 2024). Visual stimuli were presented using Tobii Pro Lab on an Eizo FlexScan EV2451 24″ monitor (1920 × 1080 resolution) with a Tobii Pro TX300 eye tracker, sampling at 1.200 Hz. Each trial displayed in SimSun 72 font, with each character subtending approximately 1.8 ° of visual angle. All stimuli were displayed in a consistent font and line spacing, and calibration was monitored to ensure average error remained below 0.5 ° of visual angle throughout the session.

2.4 Procedure

Before the experiment, each participant completed a standard nine-point calibration procedure to ensure accurate gaze recording.

TABLE 1 Experimental conditions and example stimuli.

Focus	Length	Example sentences
Focus	Disyllabic	王莉问,"你去图书馆干什么?" 刘强说:"我去图书馆 [学习], 马上毕业了, 工作不好找, 我打算考研了。" Wang Li wen, "Ni qu tushuguan gan shenme?" Liu Qiang shuo: "Wo qu tushuguan [xuexi], mashang biye le, gongzuo bu hao zhao, wo dasuan kao yan le."
	Monosyllabic	王莉问,"你去图书馆干什么?" 刘强说:"我去图书馆 [学], 马上毕业了, 工作不好找, 我打算考研了。" Wang Li wen, "Ni qu tushuguan gan shenme?" Liu Qiang shuo: "Wo qu tushuguan [xue], mashang biye le, gongzuo bu hao zhao, wo dasuan kao yan le." Wang Li asked, *"Why are you going to the library?"* Liu Qiang replied, *"I'm going to study there. I'm graduating soon, and since jobs are hard to find, I've decided to prepare for graduate school."*
No focus	Disyllabic	李彤说:"自习室里边太吵了, 根本呆不下去, 我去图书馆 [学习], 你们吃饭的时候叫我一声。" Li Tong shuo: "Zixishi libian tai chao le, genben dai bu xia qu, wo qu tushuguan [xuexi], nimen chi fan de shihou jiao wo yi sheng."
	Monosyllabic	李彤说:"自习室里边太吵了, 根本呆不下去, 我去图书馆 [学], 你们吃饭的时候叫我一声。" Li Tong shuo: "Zixishi libian tai chao le, genben dai bu xia qu, wo qu tushuguan [xue], nimen chi fan de shihou jiao wo yi sheng." Li Tong said, "The study room is too noisy, it's impossible to stay. I'll go study at the library instead. Call me when it's time to eat."

TABLE 2 Naturalness ratings across conditions.

Focus	Length	Naturalness (SD)
Focus	Disyllabic	1.75 (0.20)
	Monosyllabic	1.10 (0.25)
No focus	Disyllabic	1.60 (0.18)
	Monosyllabic	0.95 (0.22)

Values are means with standard deviations in parentheses.

Recalibration was performed whenever the average calibration error exceeded 0.5 ° of visual angle. During the experiment, participants were instructed to read each sentence silently for comprehension and to press the space bar to advance to the next sentence. At the end of each session, participants were asked to read aloud a subset of the stimulus sentences to monitor attention and comprehension (Rayner, 1978). These sentences were not part of the critical items analyzed and were randomly selected across participants to avoid repetition effects. This read-aloud task served solely as an engagement check and was introduced only after the silent reading trials to minimize interference with natural reading behavior. Prior to the main experiment, participants completed 10 practice trials to become familiar with the procedure. The entire experimental session lasted approximately 45 min.

2.5 Data processing and statistics analysis

Fixations shorter than 80 ms or longer than 1,200 ms were removed. Gaze duration (GD), first fixation duration (FFD), first pass duration (FPD), regression path duration (RPD), fixation count (FC), regression count (RC), and skipping probability (SP) were then calculated. All time measures were log-transformed to reduce skewness.

Outlier removal was conducted following standard practice (Staub and Rayner, 2007). For each condition, condition-specific means and standard deviations were calculated for all fixation duration measures. Trials with values exceeding ± 3 standard deviations from the condition mean were excluded to minimize the influence of extreme outliers. This trimming was performed separately for each condition to avoid introducing systematic bias. Trials with a GD of zero were removed for reading-time analyses, as these indicate either skipped words or missing fixations. However, these trials were retained and coded separately to estimate skipping probability (SP) using a binomial mixed-effects model. After all data cleaning steps, approximately 80% of the original trials were retained for analysis, yielding 2,340 valid observations across all conditions. Due to natural reading behavior, monosyllabic words in non-focus conditions were skipped more frequently, resulting in proportionally fewer retained trials in this cell after trimming. This is expected and theoretically meaningful because short, low-salience words are more prone to skipping during natural reading (Rayner, 1978). Descriptive statistics for each measure by condition are provided in Table 3.

To examine the effects of focus, word length, lexical frequency, and perceived naturalness on gaze behavior, linear mixed-effects models were fitted using the lme4 package (Version 1.1.36) in R (Version 4.4.1; Bates et al., 2015). For each eye-tracking measure, two sets of models were compared. The first model included Focus, Word Length, their interaction, and word frequency as fixed effects. The second model additionally included mean naturalness ratings as a covariate to control for variation in sentence acceptability. Random intercepts were specified for participants and items to account for individual differences and item-level variability. A more complex random-effects structure, including by-item random slopes for Focus and Word Length, was also tested but did not significantly improve model fit compared to the random-intercepts-only model, $\chi^2_{(6)} = 4.51$, $p = 0.608$. Therefore, the simpler model was retained for interpretation and reporting.

The internal structure of disyllabic target words (Verb–Verb, Verb–Object, Other) was coded to allow follow-up analyses testing whether word length effects might be partly explained by structural differences.

3 Results

3.1 Main effects

Since word length strongly influenced naturalness ratings, we compared two models: one excluding naturalness as a covariate and one including it to account for potential overlap. Two model summary statistics are shown in Tables 4, 5.

The main eye movement measures include early measures, such as GD, FFD, and FPD, which reflect initial lexical processing, as well as late measures, such as RPD, RC, FC, and SP to capture later rereading and integrative processes (Rayner, 1998; Clifton et al., 2007).

Table 4 presents the fixed effects estimates for the early measures GD, FFD and FPD. For GD, there was a significant main effect of word length, indicating that monosyllabic verbs were associated with longer reading times than disyllabic verbs ($b = 0.25$). The interaction between focus and length was also significant ($b = -0.61$), showing that the length effect was reduced when the word was not in focus (see Figure 1A). For FPD and FFD, neither focus nor length showed significant main effects, but both measures showed significant interactions (FPD: $b = -0.17$; FFD: $b = -0.09$), suggesting that monosyllabic verbs were processed more quickly when not in focus.

Moreover, Table 5 shows the estimates for the late measures, including RPD, RC, FC, and Skipping (probability). A similar interaction pattern emerged for RPD, where monosyllabic verbs tended to elicit slightly longer durations ($b = 0.02$), and the Focus × Length interaction was again negative and significant ($b = -0.28$, $p < 0.001$). For RC, there was a significant main effect of Length ($b = 2.43$) and a significant negative interaction ($b = -2.31$), indicating that monosyllabic verbs triggered more regressions when in focus. FC showed the same pattern: more fixations for monosyllabic verbs ($b = 0.65$) and a significant negative interaction ($b = -1.31$; see Figure 1B).

Finally, the model for Skipping revealed a significant positive Focus × Length interaction ($b = 1.26$, $p < 0.001$), indicating that monosyllabic verbs were more likely to be skipped when not in focus (see Figure 1C).

Across all measures, word frequency did not show a consistent significant effect ($ps > 0.05$), suggesting that lexical access was not a primary driver of the observed length and focus effects. Taken together, these results indicate that both focus and word length and especially their interaction systematically influenced eye movement behavior during reading. Monosyllabic verbs generally attracted longer processing times and more regressions when in focus but were more likely to be skipped when not in focus.

3.2 Follow-up analysis: naturalness

To test whether the effects could be explained by perceived naturalness, we re-ran all models with mean naturalness ratings as a covariate. The main effect of word length for GD was reduced and became non-significant ($b = 0.10$, $p > 0.05$), but the critical Focus × Length interaction remained robust ($b = -0.47$, $p < 0.001$). For FFD and FPD, all main effects became non-significant except a marginal Focus × Length interaction for FPD ($b = -0.11$, $p = 0.06$). For RPD, the Focus × Length interaction stayed significant ($b = -0.18$, $p < 0.05$). For FC, the length effect decreased but stayed significant ($b = 0.42$, $p < 0.05$) and the interaction held ($b = -1.08$, $p < 0.001$). For RC, the length effect remained significant ($b = 2.53$, $p < 0.05$) but the interaction was only marginal ($b = -2.41$, $p = 0.06$). For SP, the Focus × Length interaction stayed significant ($b = 1.08$, $p < 0.01$).
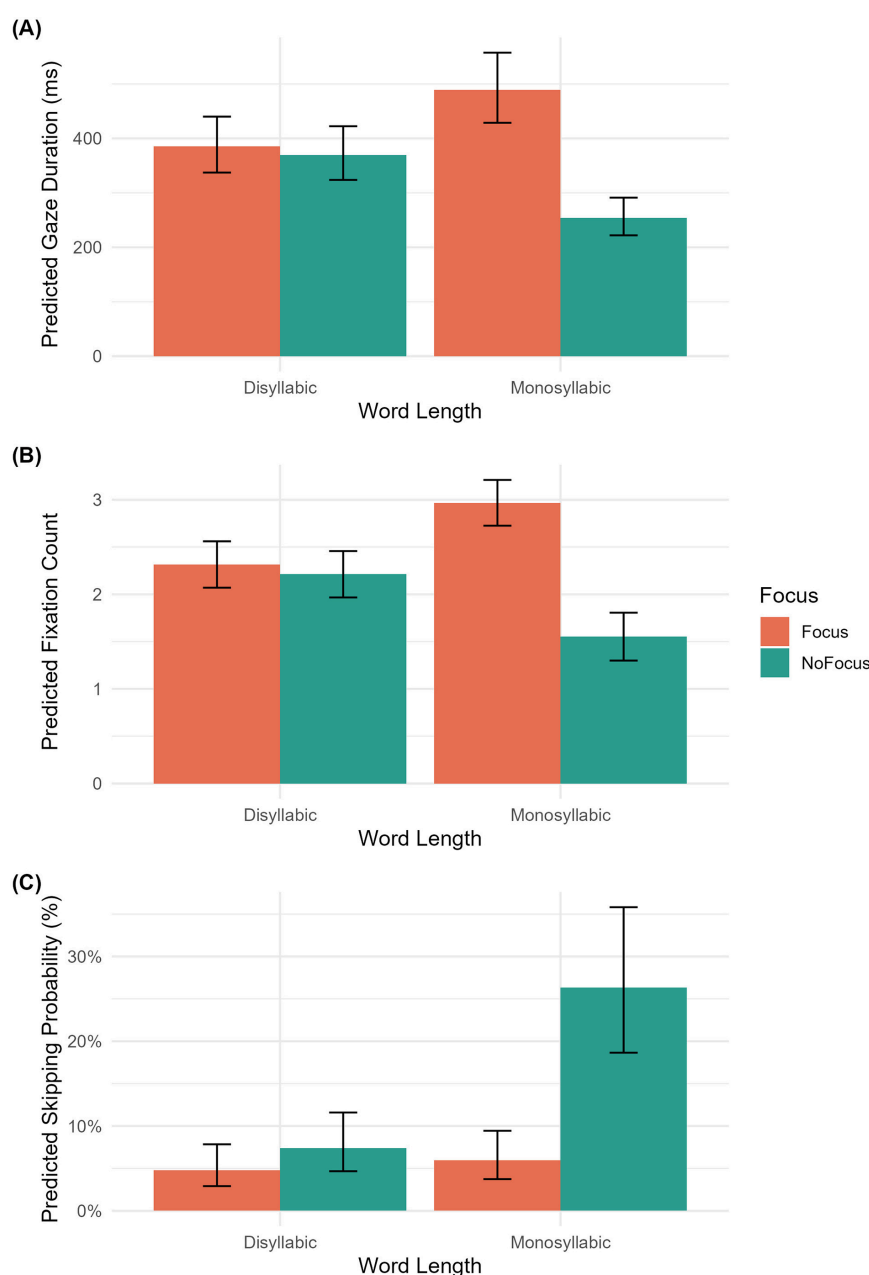
FIGURE 1
Predicted eye-tracking measures by focus and word length. Estimated marginal means (±95% confidence intervals) from linear mixed-effects models are shown for three reading measures: **(A)** Gaze Duration (GD), **(B)** Fixation Count (FC), and **(C)** Skipping Probability (SP). Bars represent predicted values for monosyllabic and disyllabic verbs under Focus and No Focus conditions. GD is in milliseconds, FC is in raw counts, and SP is plotted as predicted probability (%). Error bars indicate model-derived 95% confidence intervals.

Model comparisons showed that adding naturalness marginally improved fit for GD [$\chi^2_{(1)} = 4.30$, $p = 0.04$] but not for other measures ($ps > 0.05$). For example, the AIC decreased only slightly ($\Delta \text{AIC} \approx -1$ to -2) and the BIC differences were mixed, suggesting the added parameter did not meaningfully enhance explanatory power.

Taken together, these results indicate that the processing costs associated with mismatches between focus and word length cannot be explained by subjective naturalness alone, effectively ruling out the possibility that the observed effects are merely artifacts of awkward phrasing or sentence acceptability. This supports the interpretation that the prosodic expectations linked to word length play a genuine role in Mandarin silent reading.

3.3 Follow-up analysis: internal structure

As a follow-up to the main word length effect, we examined whether it could be explained by differences in the internal structure of the disyllabic target verbs (Roelofs, 2002). Following

TABLE 3 Descriptive statistics of eye movement measures across conditions.

Focus	Length	Trials	GD	FFD	FP	RPD	FC	RC	SP
Focus	Disyllabic	643	488 (361)	208 (86)	256 (148)	369 (283)	2.3 (1.5)	3 (5.6)	11.3%
Focus	Monosyllabic	662	679 (525)	219 (108)	268 (174)	391 (310)	3 (2.1)	5.4 (7.6)	10.6%
No focus	Disyllabic	616	477 (314)	208 (86)	254 (139)	309 (184)	2.3 (1.3)	3.8 (5.2)	12.2%
No focus	Monosyllabic	468	331 (228)	196 (88)	215 (113)	237 (135)	1.6 (0.9)	4.1 (6)	32.3%

Means are presented with standard deviations in parentheses. Trials Remaining indicates the number of valid observations per condition after trimming for outliers and missing values.

TABLE 4 Fixed effects estimates for early eye-tracking measures.

Measure	Predictor	b	SE	CI	t/z	d	Sig*
GD	Focus	−0.04	0.07	[−0.17, 0.09]	−0.63	−0.07	
	Length	0.24	0.07	[0.11, 0.37]	3.52	0.41	***
	Freq	0.01	0.11	[−0.22, 0.23]	0.05	0.01	
	Focus × Length	−0.61	0.09	[−0.79, −0.43]	−6.64	−1.05	***
FFD	Focus	0.00	0.03	[−0.05, 0.05]	−0.11	−0.01	
	Length	0.02	0.03	[−0.04, 0.07]	0.62	0.04	
	Freq	0.01	0.05	[−0.07, 0.10]	0.31	0.04	
	Focus × Length	−0.09	0.04	[−0.17, −0.02]	−2.56	−0.24	*
FPD	Focus	0.00	0.03	[−0.07, 0.06]	−0.13	−0.01	
	Length	0.02	0.04	[−0.05, 0.09]	0.67	0.05	
	Freq	−0.05	0.06	[−0.17, 0.07]	−0.77	−0.10	
	Focus × Length	−0.18	0.05	[−0.27, −0.08]	−3.56	−0.37	***

b, unstandardized estimate; SE, standard error; CI, 95% confidence interval; d, effect size. Freq, Word frequency; Significance: *p < 0.05, **p < 0.01, ***p < 0.001.

established linguistic classifications (Packard, 2000; Robson, 2018), we categorized the verbs into three primary structural types: Verb-Verb (VV), Verb-Object (VO), and other forms [including Verb-Complement (VC), Morpheme-Morpheme (MM), and Subject-Verb (SV) structures]. Within Focus–Disyllabic condition, target words consisted of VV ($n = 251$ observations, 26 unique words), VO ($n = 365$, 43 unique), and other disyllabic forms (VC, MM, SV combined; $n = 84$, 11 unique). The No Focus–Disyllabic condition contained similar proportions (VV: $n = 223$; VO: $n = 407$; Other: $n = 76$). As expected, monosyllabic items were only present in the Focus–Monosyllabic ($n = 777$) and No Focus–Monosyllabic ($n = 734$) conditions.

To test whether internal structure could account for the word length effect, we compared two mixed-effects models for each reading measure: one with Focus × Length and one with Focus × Structure as predictors. Across all measures, the structure models did not improve model fit compared to the simpler length-based models (all ps >0.46). For Skipping Probability, no disyllabic targets were skipped, and in the full dataset, the structure model did not improve fit ($p = 1$).

Within the structure models, VV and VO forms often showed numerically fewer fixations than monosyllabic words (e.g., Fixation Count: VV $b = −0.81$; VO $b = −0.61$), with significant Focus × Structure interactions (VV: $b = 1.43$; VO: $b = 1.30$). However, pairwise comparisons within the disyllabic subset showed no

significant differences among VV, VO, and other structures in either focus condition for any measure (all Tukey-adjusted ps >0.51). Mean fixation counts remained similar under focus (VV: $M = 2.19$; VO: $M = 2.34$; Other: $M = 2.52$) and no focus (VV: $M = 2.19$; VO: $M = 2.25$; Other: $M = 2.09$).

Together, these results indicate that the processing cost associated with focus and word length is not explained by internal word structure, supporting the interpretation that the length effect reflects a genuine prosodic cue rather than a morpho-syntactic difference.

4 Discussion

This study examined how focus and word length interact to shape eye movement behavior during silent reading in Mandarin. As predicted, monosyllabic words in focus positions elicited longer gaze durations, more fixations, and more regressions compared to their unfocused counterparts, while unfocused monosyllabic words were more likely to be skipped. These results are consistent with established findings that word length robustly influences both early and late eye movement measures (Rayner, 1998; Kliegl et al., 2004) and extend this effect by demonstrating that discourse-level focus systematically moderates how readers allocate visual attention.

TABLE 5 Fixed effects estimate for late eye-tracking measures and skipping.

Measure	Predictor	b	SE	CI	t/z	d	Sig*
RPD	Focus	−0.13	0.05	[−0.22, −0.03]	−2.56	−0.23	*
	Length	0.02	0.05	[−0.08, 0.12]	0.43	0.04	
	Freq	0.00	0.09	[−0.17, 0.17]	−0.02	0.00	
	Focus × Length	−0.28	0.07	[−0.41, −0.14]	−3.94	−0.51	***
FC	Focus	−0.10	0.13	[−0.36, 0.15]	−0.79	−0.08	
	Length	0.65	0.14	[0.38, 0.92]	4.78	0.48	***
	Freq	−0.02	0.23	[−0.47, 0.43]	−0.10	−0.02	
	Focus × Length	−1.31	0.19	[−1.68, −0.95]	−7.01	−0.97	***
RC	Focus	0.50	0.72	[−0.91, 1.92]	0.70	0.11	
	Length	2.43	0.75	[0.96, 3.91]	3.23	0.55	**
	Freq	−1.07	1.25	[−3.53, 1.38]	−0.86	−0.24	
	Focus × Length	−2.31	1.02	[−4.30, −0.31]	−2.27	−0.52	*
Skipped	Focus	0.46	0.24	[−0.01, 0.94]	1.90	0.46	
	Length	0.23	0.26	[−0.27, 0.73]	0.91	0.23	
	Freq	0.15	0.40	[−0.64, 0.93]	0.37	0.15	
	Focus × Length	1.26	0.32	[0.63, 1.90]	3.91	1.26	***

b, unstandardized estimate (log-odds for Skipping); SE, standard error; CI = 95% confidence interval; d, effect size (standardized coefficient). Freq, word frequency (log-transformed). Significance: *p < 0.05, **p < 0.01, ***p < 0.001. The Skipping model estimates the log odds of a word being skipped.

As word length functions as a prosodic factor in Mandarin, the significant interaction between focus and word length supports the claim that readers generate prosodic expectations during silent reading (Fodor, 2002). The increased processing cost for focused monosyllabic verbs suggests that Mandarin readers disfavor monosyllabic forms as focus carriers, preferring disyllabic or even longer forms to mark prosodic prominence. This is consistent with the language's preference for binary foot structures (Duanmu, 2007). When these expectations are violated, for example, when a monosyllabic word carries focus, additional integration effort is required, reflected in longer gaze durations and more regressions. This aligns with previous research showing that mismatches between expected prosody and lexical form can disrupt reading fluency (Breen et al., 2010).

However, we also note the absence of significant length effects on FFD and FPD, we interpret this as reflecting the high contextual predictability and semantic equivalence of the elastic verb pairs. As Staub (2015) argues, predictability effects in reading arise from the graded activation of multiple candidate words rather than the discrete prediction of a single word. This graded activation can facilitate very early stages of lexical or even pre-lexical processing, allowing readers to access and integrate these words efficiently during initial fixations. By contrast, the significant effects on GD, FC and RC indicate that genuine length-related processing demands emerged at later stages, when readers integrated lexical form with discourse-level focus cues.

More generally, the finding that disyllabic words under no focus showed longer gaze durations and more fixations than

monosyllabic words is consistent with prior evidence across languages showing that longer words attract more reading time and fixations (Rayner, 2009; Zang et al., 2018). One possible explanation is that monosyllabic words in non-focus contexts became more predictable and contextually preferred, which could result in their higher skipping rates (Li et al., 2023). This is also aligned with classic skipping effects in alphabetic languages, where monosyllabic words benefit more from parafoveal processing than longer words (Fitzsimmons and Drieghe, 2011).

In follow up analysis, when adding subjective naturalness ratings as a covariate reduced or eliminated some of the main effects for word length and focus in early and late measures. This reduction likely reflects the fact that perceived naturalness was not independent of the experimental manipulations: sentences with monosyllabic vs. disyllabic forms, or with focus vs. non-focus placement, naturally differ in how acceptable or smooth they feel to native speakers. As such, naturalness is partly confounded with the manipulated conditions. Even so, the critical Focus × Length interaction remained robust for key measures such as GD, RPD and SP. This indicates that the added reading difficulty cannot be fully explained by sentence naturalness alone but reflects readers' sensitivity to prosodic expectations during silent reading. The naturalness ratings further showed that monosyllabic forms were generally judged as less natural than disyllabic forms, with the strongest effect in non-focus contexts (Condition D), a pattern not mirrored in the eye-tracking data. We interpret this asymmetry as reflecting a broader stylistic preference for disyllabic forms in written Mandarin, particularly when verbs are backgrounded and

carry little discourse prominence. In focus contexts, monosyllabic verbs (Condition B) were rated more natural than monosyllabic verbs in non-focus contexts (Condition D). This pattern may arise because the naturalness judgments were based on the entire sentence rather than the critical verb alone.

Another potential alternative interpretation to the processing cost for monosyllabic verbs is that it reflects semantic incompleteness rather than a prosodic mismatch. However, our design allows us to rule out this explanation. If semantic incompleteness were the source of difficulty, then Condition A (学习, focus) and Condition B (学, focus) should differ, and Condition D (学, non-focus) should also feel "incomplete," since it contains the same monosyllabic verb. Yet our results show that only Condition B incurred processing costs, while Condition D did not. This pattern demonstrates that the observed effects cannot be reduced to semantic incompleteness. Instead, this descriptive contrast indicates that focus is the key factor: monosyllabic verbs only cause difficulty when they must carry prosodic prominence. Furthermore, the mixed-effects analyses revealed a significant Focus × Length interaction, showing that the effect of word length systematically depends on focus status.

Taken together, the present findings extend the implicit prosody hypothesis (Fodor, 2002) to Mandarin focus processing. The results show that readers project prosodic expectations during silent reading and encounter integration difficulty when these expectations are violated, specifically, when monosyllabic verbs serve as focus carriers. This provides new evidence that prosodic constraints and information-structural requirements jointly shape sentence processing, even in the absence of overt speech.

More broadly, these findings provide new evidence that readers actively integrate prosodic and discourse-level cues during text processing, and they demonstrate that even subtle mismatches between expected prosody and word form can incur measurable processing costs, as reflected in standard eye-tracking measures. Methodologically, this study underscores the value of combining word-level manipulations with discourse context in eye-tracking research to capture how readers adapt their reading strategies in real time. This contributes to current models of the prosody–syntax interface and opens avenues for cross-linguistic and second language research.

These findings also carry pedagogical implications. Elastic verb pairs in Mandarin (e.g., 睡觉 shuìjiào vs. 睡 shuì) illustrate that syllable count can serve as a cue for informational prominence. For L2 learners, selecting the appropriate form is not merely a matter of lexical choice but requires sensitivity to prosodic expectations. The present results indicate that while monosyllabic forms are efficient in non-focus contexts, their use in focus positions may introduce a prosodic mismatch, contributing to perceptions of unnaturalness in learners' speech.

While the present findings are robust, several limitations should be acknowledged. First, all target verbs occurred without overt objects in sentence-final position. This design ensured prosodic comparability across conditions but restricts the generalizability of the results. Future research is needed to determine whether similar focus–length interactions occur when verbs take explicit objects or appear in non-final positions. Second, the stimuli focused exclusively on verbs, and more specifically on single, isolated verbs as the locus of focus. In natural discourse, however, focus often extends beyond a single word to larger units such as verb phrases or subordinate clauses containing multiple verbs. Whether the preference for disyllabic forms generalizes to broader focus domains or other word classes such as nouns and adjectives remains an open question for future research. Third, while we controlled for semantic equivalence, we did not directly measure the cloze probability of the target verbs; future work could use corpus-based cloze measures to quantify contextual predictability more precisely. Last, complementing eye movement measures with real-time physiological indices such as pupillometry or ERPs would further clarify how prosodic mismatches unfold over time.

5 Conclusion

This study provides robust evidence that Mandarin readers integrate prosodic expectations into silent reading, with word length functioning as a prosodic factor and serving as a reliable cue for focus marking. Monosyllabic verbs in focus positions elicited measurable processing costs, such as longer gaze durations, more regressions, and more fixations, indicating that disyllabic forms better match prosodic expectations in focus contexts. These effects remained significant even when controlling for sentence naturalness and internal morphological structure, confirming that the observed processing patterns are not merely artifacts of awkward phrasing or morpho-syntactic factors.

In conclusion, the findings provide strong support for the implicit prosody hypothesis (Fodor, 2002) by showing that prosodic mismatches influence eye movements in real time. They advance our understanding of how readers coordinate lexical, prosodic, and discourse-level information, contribute to models of the prosody–syntax interface, and offer insights into how subtle linguistic cues shape reading fluency. Future research can extend this work to additional word classes, broader focus domains, and cross-linguistic comparisons, particularly in second language contexts.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The studies involving humans were approved by the Swedish Ethical Review Authority (Etikprövningsmyndigheten), Sweden. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

Funding

Acknowledgments

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2025.1668351/full#supplementary-material

SUPPLEMENTARY FILE S1
Full list of elastic verbs (monosyllabic vs. disyllabic pairs) used in the study.

SUPPLEMENTARY FILE S2
Full experimental stimuli sentences (.pdf).

SUPPLEMENTARY FILE S3
(Data): Anonymized raw data files (.tsv), cleaned reading metrics (.csv), word frequency data (.csv), naturalness judgment tables (.csv), and the list of target disyllabic verbs with internal structure coding (.txt).

SUPPLEMENTARY FILE S4
(R Scripts): R scripts and README for data cleaning, modeling, and replication (.zip).

References

Arnhold, A. (2024). No prosody-syntax trade-offs: prosody marks focus in Mandarin cleft constructions. *Lab. Phonol.* 15, 1–49. doi: 10.16995/labphon.11515

Balota, D. A., Cortese, M. J., and Sergent-Marshall, S. D. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 283–316. doi: 10.1037/0096-3445.133.2.283

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Breen, M., and Clifton, Jr., C. (2011). Stress matters: effects of anticipated lexical stress on silent reading. *J. Mem. Lang.* 64, 153–170. doi: 10.1016/j.jml.2010.11.001

Breen, M., Fedorenko, E., Wagner, M., and Gibson, E. (2010). Acoustic correlates of information structure. *Lang. Cogn. Process.* 25, 1044–1098. doi: 10.1080/01690965.2010.504378

Cai, Q., and Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE* 5:e10729. doi: 10.1371/journal.pone.0010729

Chao, Y. R. (1948). *Mandarin Primer: An Intensive Course in Spoken Chinese.* Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674732889

Clifton, Jr., C., Staub, A., and Rayner, K. (2007). "Eye movements in reading words and sentences," in *Eye Movements: A Window on Mind and Brain*, eds. R. P. G. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (Amsterdam: Elsevier), 341–371. doi: 10.1016/B978-008044980-7/50017-3

Dong, Y. (2015). *The prosody and morphology of elastic words in Chinese* (Dissertation). University of Michigan, Ann Arbor, MI, United States. Available online at: https://deepblue.lib.umich.edu/handle/2027.42/116629 (accessed May 30 2025).

Duanmu, S. (2007). *The Phonology of Standard Chinese, 2nd Edn.* Oxford: Oxford University Press. doi: 10.1093/oso/9780199215782.001.0001

Ferrand, L., and New, B. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychol.* 113, 167–183. doi: 10.1016/S0001-6918(03)00031-3

Fitzsimmons, G., and Drieghe, D. (2011). The influence of number of syllables on word skipping during reading. *Psychon. Bull. Rev.* 18, 736–741. doi: 10.3758/s13423-011-0105-x

Fodor, J. D. (2002). "Prosodic disambiguation in silent reading," in *Proceedings of the Northeast Linguistics Society*, Vol. 32, ed. M. Hirotani (Amherst, MA: GLSA, University of Massachusetts), 112–132.

Guo, S. (1938). Zhongguo yuci zhi tanxing zuoyong [The function of elastic word length in Chinese]. *Yen Ching Hsueh Pao [Yenching J.]* 24, 1–34.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *J. Linguist.* 19, 377–417. doi: 10.1017/S0022226700007799

Hauk, O., and Pulvermuller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clin. Neurophysiol.* 115, 1090–1103. doi: 10.1016/j.clinph.2003.12.020

Huang, L., and Duanmu, S. A. (2013). quantitative study of word-length elasticity in Modern Chinese [现代汉语词长弹性的量化研究]. *Yuyan Kexue [Linguistic Sci.]* 12, 8–16.

Karlgren, B. (1918). *Ordet och pennan i Mittens rike.* Stockholm: Svenska Bokförlaget.

Karlgren, B. (1971). *Sound and Symbol in Chinese, Rev Edn.* Hong Kong: Hong Kong University Press.

Kiss, K. É. (2015). "Discourse functions: the case of Hungarian," in *The Oxford Handbook of Information Structure*, eds. C. Féry and S. Ishihara (Oxford: Oxford University Press), 663–685. doi: 10.1093/oxfordhb/9780199642670.013.24

Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *Eur. J. Cogn. Psychol.* 16, 262–284. doi: 10.1080/09541440340000213

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topics, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620607

Lee, P. L., and Sun, Y. (2024). "Focus in Chinese," in *Oxford Research Encyclopedia of Linguistics* (New York, NY: Oxford University Press).

Li, C. N., and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press. doi: 10.1525/9780520352858

Li, Y., Scontras, G., and Futrell, R. (2023). Chinese words shorten in more predictive contexts. *Proc. Annu. Meet. Cogn. Sci. Soc.* 45.

Long, J., Wang, T., and Yu, M. (2022). Sentential position of V-N combination modulates the rhythmic pattern effect during Chinese sentence reading: evidence from eye movements. *Acta Psychol.* 228:103641. doi: 10.1016/j.actpsy.2022.103641

Luo, Y., Zhang, Y., Feng, X., and Zhou, X. (2010). Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience* 169, 654–664. doi: 10.1016/j.neuroscience.2010.05.032

Luo, Y., and Zhou, X. (2010). ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading. *Neuroimage* 49, 2836–2849. doi: 10.1016/j.neuroimage.2009.10.008

Niehorster, D. C., Gullberg, M., and Nyström, M. (2024). Behavioral science labs: how to solve the multi-user problem. *Behav. Res. Methods* 56, 8238–8258. doi: 10.3758/s13428-024-02467-4

Ouyang, I. C., and Kaiser, E. (2015). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Lang. Cogn. Neurosci.* 30, 57–72. doi: 10.1080/01690965.2013.805795

Packard, J. L. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486821

Peking University Center for Chinese Linguistics (2023). *CCL Corpus* (Peking University Corpus of Modern Chinese). Beijing: Peking University. Available online at: http://ccl.pku.edu.cn:8080/ccl_corpus/ (accessed June 30 2025).

Rayner, K. (1978). Eye movements in reading and information processing. *Psychol. Bull.* 85, 618–660. doi: 10.1037/0033-2909.85.3.618

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi: 10.1080/17470210902816461

Robson, S. Y. (2018). The morphology of Chinese disyllabic verbs. *Int. J. Chin. Linguist.* 5, 94–124. doi: 10.1075/ijchl.16020.yon

Roelofs, A. (2002). Syllable structure effects turn out to be word length effects: comment on Santiago et al. (2000). *Lang. Cogn. Process.* 17, 1–13. doi: 10.1080/01690960042000139

Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., and Richlan, F. (2016). Words in context: the effects of length, frequency, and predictability on brain responses during natural reading. *Cereb. Cortex* 26, 3889–3904. doi: 10.1093/cercor/bhw184

Selkirk, E. O. (1985). "Intonation, stress and meaning," in *Proceedings of the 11th Annual Meeting of the Berkeley Linguistics Society*, eds. M. Niepokuj, M. VanClay, V. Nikiforidou, and D. Feder (Berkeley, CA: Berkeley Linguistic Society), 491–504. doi: 10.3765/bls.v11i0.1902

Shi, Y. (2023). "Disyllabification," in *The Evolution of Chinese Grammar* (Cambridge: Cambridge University Press), 78–105. doi: 10.1017/9781108921831.005

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: critical review and theoretical interpretation. *Lang. Linguist. Compass.* 9, 311–327. doi: 10.1111/lnc3.12151

Staub, A., and Rayner, K. (2007). "Eye movements and on-line comprehension processes," in *Oxford Handbook of Psycholinguistics*, ed. G. Gaskell (Oxford: Oxford University Press), 327–342. doi: 10.1093/oxfordhb/9780198568971.013.0019

Stolterfoht, B., Friederici, A. D., and Alter, K. (2007). Processing focus structure and implicit prosody during reading: differential ERP effects. *Cognition* 104, 565–590. doi: 10.1016/j.cognition.2006.08.001

Wolof, R. S. (2024). *The Oxford Guide to the Atlantic Languages of West Africa* (ed. F. Lüpke). Oxford: Oxford University Press, 61–96. doi: 10.1093/oso/9780198736516.003.0004

Xia, Q., Wang, L., and Peng, G. (2016). Nouns and verbs in Chinese are processed differently: evidence from an ERP study on monosyllabic and disyllabic word processing. *J. Neurolinguistics* 40, 66–78. doi: 10.1016/j.jneuroling.2016.06.002

Yan, M., and Calhoun, S. (2019). Priming effects of focus in Mandarin Chinese. *Front. Psychol.* 10:1985. doi: 10.3389/fpsyg.2019.01985

Yang, J., Staub, A., Li, N., Wang, S., and Rayner, K. (2012). Plausibility effects when reading one- and two-character words in Chinese: evidence from eye movements. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1801–1809. doi: 10.1037/a0028478

Zang, C., Fu, Y., Bai, X., Yan, G., and Liversedge, S. P. (2018). Investigating word length effects in Chinese reading. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 1831–1841. doi: 10.1037/xhp0000589