RESEARCH ARTICLE

Orthographic processing was facilitated by phonologically but not semantically similar distractors: evidence from the visual structure judgment of Mandarin Chinese characters using the Flanker paradigm

Xingyu Yang[1,2,3] · Likan Zhan[1,2,3]

Abstract

Phonology is essential in visual word recognition, yet its exact role remains debated. This is partly because the paradigms used in the literature were typically the priming paradigm and its variations, with tasks requiring participants to fully process the test items. Given these two limitations, the observed effects in the literature may be confounded by the specific paradigm and/or task employed. Accordingly, the current study extends previous findings using a different paradigm and task. Specifically, the flanker paradigm, in which the target character was presented simultaneously with two distractor characters for a brief period (50 ms), was utilized. Participants were asked to judge the visual structure (left-right structure versus top-down structure) of the target character. With these two modifications, we found that the judgment was facilitated when the target and distractors were phonologically similar (compared to phonologically dissimilar trials, experiment 1), while no facilitation was observed when the items were semantically similar (compared to semantically dissimilar trials, experiment 2). These data suggest rapid phonological activation, including from parafoveal input, but delayed semantic engagement in Mandarin Chinese under 50-ms flanker displays.

Keywords Flanker paradigm · Phonology · Orthography · Chinese characters · Backward influence

Introduction

Word recognition begins with a word's physical form and culminates in its meaning (semantics). In the auditory domain, the relevant form is phonology (sound); in the visual domain, it is orthography. In auditory domain,

✉ Likan Zhan
zhanlikan@hotmail.com

1 Cognitive Science and Allied Health School, Beijing Language and Culture University, Beijing, China

2 Institute of Life and Health Sciences, Beijing Language and Culture University, Beijing, China

3 Key Laboratory of Language and Cognitive Science (Ministry of Education), Beijing Language and Culture University, Beijing, China

meaning is accessed via phonology. In visual domain, however, meaning may be accessed directly from orthography or mediated by phonology (Carreiras et al. 2014; Coltheart et al. 2001; Duan and Cai 2024; Harm and Seidenberg 2004; Li et al. 2022; Perfetti et al. 1992, 2005; Yeatman and White 2021). Nevertheless, the role of phonology in visual word recognition remains unresolved, particularly in logographic systems such as Mandarin Chinese. In Mandarin, mappings between orthography and semantics are often more transparent, and phonology is thought to be less essential than in alphabetic systems like English.

Even so, phonology is generally considered to be activated relatively early during visual word recognition, including in Mandarin Chinese (Tan & Perfetti, 1998). Evidence for early phonological activation in Mandarin primarily comes from priming paradigms and their variants. These studies show that participants' responses to target items were affected by forward primes (Perfetti and Tan 1998; Perfetti and Zhang 1991) and by backward primes (Tan et al. 1995, 1996), when they were required to write the target character (Perfetti and Zhang 1991; Tan et al. 1995, 1996), to name

it (Perfetti and Tan 1998), to judge lexical status for two-character words (Zhou et al. 1999), to judge whether it fit a prior definition (Chua 1999), or to judge whether two characters had similar meanings (Chua 1999; Perfetti and Zhang 1995; Xu et al. 1999; Zhang et al. 1999) or similar pronunciations (Chua 1999; Perfetti and Zhang 1995; Zhang et al. 1999). Priming effects are observed when prime and target are graphically, phonologically, or semantically similar (Tan & Perfetti, 1998). Taken together, the evidence suggests that phonology is activated once the orthographic representation is sufficiently specified to distinguish the character from competitors (Perfetti and Tan 1998).

As summarized, however, paradigms used in literature were normally the priming paradigm and its variations, and the tasks used in literature typically required the full processing of the characters. The effects observed in the literature might be confounded by the specific paradigm and/or the specific task being used. Accordingly, the current study intends to extend the previous findings with a different paradigm and a different task.

First, we will use the flanker task with brief presentation time. In the flanker task (Eriksen and Eriksen 1974), participants are asked to respond to a target item that is flanked (surrounded) by distractor items. When introducing the paradigm into visual word recognition, the target item is typically presented in the fovea area, and the distractor items (normally two) are symmetrically presented in the parafovea areas. The presentation time is normally short, preventing participants from moving their attention from the target item to the distractor items. Participants' task is to determine whether the target item is a word or not (Dare and Shillcock 2013; Grainger et al. 2014; Snell and Grainger 2018; Snell et al. 2018, 2019; Snell, Meeter, et al., 2017; Snell, Vitu, et al., 2017), whether the object is a natural or artificial object (Meade et al. 2021; Snell, Declerck, et al., 2017), whether the word is a noun or a verb (Snell 2024), or whether they know the Chinese character or not (Yu et al. 2023). The general observation is that reaction times were shortest in the no-flanker condition, longer in the congruent condition, and longest in the incongruent condition. Accuracy was highest in the no-flanker condition, lower in the congruent condition, and lowest in the incongruent condition. With the help of the flanker paradigm, it is believed that the target item presented at the fovea area and the distractor items presented at the parafovea area can be processed in parallel at the same time (Schotter and Payne 2019; Snell and Grainger 2019a, b; White et al. 2019). Priming and flanker paradigms differ markedly in their temporal and spatial properties. Temporally, priming normally presents the prime and target sequentially with a stimulus-onset asynchrony (SOA) > 0, whereas the flanker paradigm presents the target and flankers simultaneously, which can be viewed as SOA = 0. Spatially, the target typically appears at fixation in both paradigms; the prime in priming is usually centrally presented, whereas flankers are placed adjacent to the target in the parafoveal region. Given these differences, the flanker paradigm is essential for the present study. The prime-to-target temporal structure in priming forces participants to process the prime and target sequentially—prime information is activated first and then influences the processing of the target. By removing this temporal structure and presenting both the target and flankers for an extremely brief duration (e.g., 50 ms which is unlikely to support strategic sequential processing), the flanker paradigm encourages parallel processing of distractors and targets, thereby minimizing pre-activation or expectancy effects while allowing us to probe how orthographic, phonological, and semantic information interact when processed concurrently. In this sense, the flanker paradigm provides a complementary and necessary test of priming-based findings.

Second, we will ask participants to report the orthographical property of the target character (visual structure, to be specific), rather than its phonological or semantic properties. As a logographic writing system, Chinese characters exhibit different visual structures based on the arrangement of their radicals. For instance, the compound character "娘" ("mother" in English) has a left-right structured character, with the semantic radical "女" on the left and the phonetic radical "良" on the right. Similarly, "草" ("grass" in English) is a top-bottom structured character, with the semantic radical "艹" on the top and the phonetic radical "早" at the bottom. A character can also be simple, consisting of only one radical, such as "月". To correctly judge the visual structure of a compound character does not necessarily need to activate the phonology and the semantics of a character. It does not even necessarily need to fully access the orthographical structure of the character. As an orthographic feature, the visual structure of a character can be recognized even by a novice who is unfamiliar with its meaning or pronunciation. This task can be used to determine whether the effects being observed in the literature are automatic or task-specific.

Together with the two modifications, we conducted two experiments to explore whether phonological congruency (Experiment 1, phonological flankers) and/or semantic congruency (Experiment 2, semantic flankers) has an effect on the orthographic judgments of the target character. Given these experimental designs, we can get the following predictions: (1) If the phonological effects reported in literature is not automatic but is task specific (Qu et al. 2022), no flanker effect should be observed in the two experiments; (2) If the phonological effects observed in literature results from a kind of feedback when the character is fully processed, then the flanker effect should both be observed in the two experiments; (3) If the flanker effect is observed in

experiment 1 but not in experiment 2, we can get the following two conclusions: First, the phonology is automatically activated in visual word recognition; Second, the access of the phonological property is faster than that of the semantic property.

Experiment 1

Method

Participant

Sixty native Chinese students (37 females, 23 males, age: 18–30) were recruited from Beijing Language and Culture University and received 15 yuan for their participation (Initially, it was 40 subjects; then, after a post hoc power analysis recommended by reviewers, 20 more subjects were added). All participants had normal or corrected-to-normal vision. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of Beijing Language and Culture University. Informed consent for participation was obtained from all subjects involved in the study.

Stimuli

We selected 120 high-frequency target characters with clear top-bottom (60) or left-right (60) structures, each paired with a single-component homophone as a flanker character and no target–flanker pair shared a radical. Our selection of left-right and top-down structured characters rather than other structures was guided by two considerations: (1) These configurations occur more frequently in standard written Chinese due to their higher prevalence in everyday usage, and (2) they provide more transparent spatial-directional relationships compared to semi-enclosed or holistic structures, which exhibit greater positional complexity.

Each target and flanker character appeared twice during the experiment: once in the homophonic flanker condition

and once in the non-homophonic flanker condition. The experiment consisted of two blocks. Half of the target characters were presented with their homophonic flankers in one block, whereas in another block, the same target characters were presented with non-homophonic flankers (which were homophones of other target characters). An additional set of 60 high-frequency characters with clear top-bottom (30) or left-right (30) structures was used as a no-flanker baseline in both blocks.

Within each block, mean frequencies of target characters in each condition were 3.414±0.796, 3.345±0.721, and 3.338±0.942 (no flanker condition), based on the logCHR indices from the SUBTLEX-CH list (Cai and Brysbaert 2010). The mean strokes of the target character in each condition were 9.217±3.13, 9.067±2.73, and 9.2±2.576 (no flanker condition). Examples of stimuli used in Experiment 1 are shown in the left panel of Table 1. For the visual similarity between flankers and target characters, three indices (Structural Similarity Index Measure [SSIM], Mean Squared Error [MSE], and Peak Signal-to-Noise Ratio [PSNR]) commonly used in image similarity analysis were calculated. These indices are widely used metrics for evaluating the degree of similarity between images. SSIM evaluates the resemblance between two images in terms of luminance, contrast, and structural patterns, yielding values closer to 1 when the images are more similar. MSE quantifies the average squared pixel-wise difference; lower values indicate higher similarity. PSNR transforms the MSE into a logarithmic scale, with higher values reflecting better similarity between images. All characters were rendered as grayscale images using the same font SimHei (the same font shown to participants during the experiment) at 48 pt. Each character was drawn with a black foreground on a white background onto a fixed 64×64-pixel canvas and centered using its bounding box. This normalization ensured a common font, nominal size, and canvas across items for fair comparison. We then computed three image-based similarity measures for each target–flanker pair. In Experiment 1, the indices for Block 1 were as follows: SSIM (congruent vs. incongruent)=0.401±0.051 vs. 0.396±0.054, MSE (congruent vs. incongruent)=13.049±2.308 vs. 13.336±2.343, and PSNR (congruent vs. incongruent)=37.047±0.814 vs. 36.954±0.828. In Block 2, the indices were: SSIM (congruent vs. incongruent)=0.401±0.045 vs. 0.415±0.079, MSE (congruent vs. incongruent)=13.111±1.865 vs. 13.122±2.203, and PSNR (congruent vs. incongruent)=37.001±0.655 vs. 37.013±0.747.

Flankers in Experiment 1 were single-component characters. In Experiment 2, flankers' structures were identical to their paired target characters.

Table 1 Examples of stimuli in both experiments

Target Structure	Experiment 1 (phonological flanker)		Experiment 2 (semantic flanker)	
	Left-right	Top-bottom	Left-right	Top-bottom
Target character	酒 tɕiou(214)	宫 kuŋ(55)	竹 tʂu(35)	凳 təŋ(51)
	Wine	Palace	Bamboo	Stool
Congruent flanker	九 tɕiou(214)	弓 kuŋ(55)	猫 mɑu(55)	票 pʻiɑu(51)
	Nine	Bow	Cat	Ticket
Incongruent flanker	又 tʂʻʌ(55)	止 tʂɿ(214)	钉 tiŋ(55)	蛋 tan(51)
	Fork	Stop	Needle	Egg

Procedure

Each character was presented in SimHei font, subtending approximately 1.10° of visual angle. The entire stimulus string in flanker conditions (including two flankers and one target character) subtended approximately 4.44° of visual angle horizontally, with equal spacing between adjacent characters. Trials in both blocks were presented randomly.

Vertical fixation bars were used to keep participants' fixation at the center of the screen throughout the experiment. Each trial began with a fixation display for 500 ms, followed by the presentation of the stimulus characters for 50 ms, which is the shortest presentation time used in previous studies to avoid potential sequential processing (Snell 2024; Yu et al. 2023). Afterward, a post-mask consisting of six hashmarks ('#') was presented for 2,000 ms. Participants were instructed to focus on the target characters only and to make a judgment about whether the current character was a left-right structure (press "J") or a top-bottom structure (press "F") as quickly as possible. Green-dot feedback was given for correct responses, while red-dot feedback was given for incorrect responses or reaction time exceeding 2,000 ms. Twenty practice trials were conducted before the formal experiment to familiarize participants with the procedure, and they were given a brief break between blocks. The entire procedure lasted approximately 15 min. Figure 1 shows the full procedure.

Data analyses

All analyses were conducted in *R* (R Core Team 2025) using the *lme4* package (Bates et al. 2015), and p-values were obtained from the *lmerTest* package (Kuznetsova et al. 2017). Trials without responses and those with reaction time exceeding three standard deviations from the mean were excluded from both accuracy and reaction time analyses. Additionally, for the reaction time analysis, trials with incorrect responses were further excluded. The structure of the target character (top-bottom and left-right) and flanker
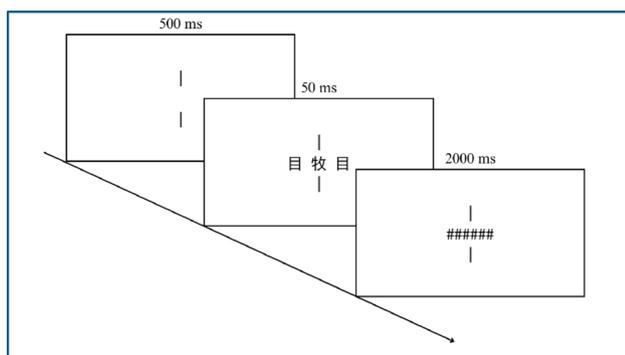

Fig. 1 An example of the procedure

type (no flanker, phonologically congruent, and phonologically incongruent), as well as their interaction, were treated as fixed factors in both the linear mixed-effects model (LMM) for reaction time and the generalized linear mixed-effects model (GLMM) for accuracy.

Both linear mixed-effects models (LMMs) and generalized linear mixed-effects models (GLMMs) incorporated random effects comprising intercepts for participants, items, and block orders. Additionally, random slopes were specified for congruency and structure within participants, as well as congruency and structure within items, thereby constructing the maximal theoretically justified model structure. When convergence issues arose during initial model fitting, we systematically simplified the random effects structure by first removing correlation parameters between random effects, followed by iteratively eliminating random slopes for interaction terms and subsequently for main effects until model convergence was achieved. All post-hoc pairwise comparisons were performed using the *emmeans* package (Lenth 2025). Since our analysis employed treatment contrasts coding, Type III ANOVA was applied to both LMMs and GLMMs via the *car* package (Fox and Weisberg 2019) to ensure unbiased estimation of interaction effects independent of term order.

Results

Two participants were excluded from analysis because of low accuracy (51.94% and 65%). The rest 58 participants all achieved accuracy above 70%. Prior to analysis, 59 trials (~0.3%) without responses and 420 trials (~2.02%) with reaction time exceeding three standard deviations from the mean were excluded. Additionally, 2259 trials (~11.04%) with incorrect responses were removed from the reaction time analysis.

For the phonologically congruent and incongruent conditions, the mean accuracy ($M \pm SD$) was 0.883 ± 0.321, 0.881 ± 0.324, respectively, and for the no-flanker baseline condition, the mean accuracy was 0.905 ± 0.293. For the top-bottom structure and left-right structure characters, the mean accuracy was 0.874 ± 0.332 and 0.905 ± 0.293. The results for accuracy and reaction time are illustrated in Figs. 2 and 3, respectively.

The GLMM didn't reveal any significant effect caused by flanker type or the interaction between congruency and target structure. Accuracy wasn't significantly different between the phonologically congruent and incongruent conditions ($\beta = -0.036$, $SE = 0.082$, $z = -0.441$, $p = .659$) or between phonologically congruent and no-flanker conditions ($\beta = 0.165$, $SE = 0.119$, $z = 1.393$, $p = .164$). However, the structure of the target character significantly influenced
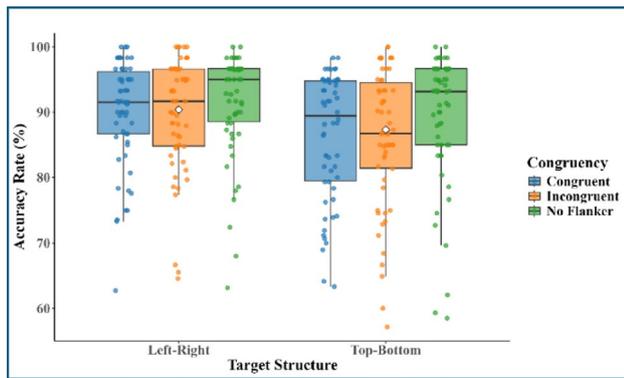
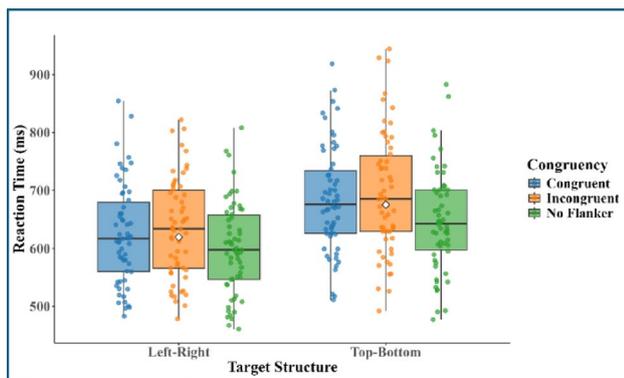**Fig. 2** Median accuracy in all conditions (Experiment 1)

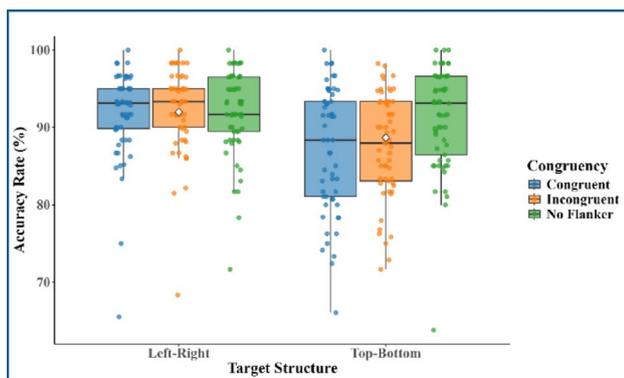**Fig. 3** Median reaction time in all conditions (Experiment 1)

Fig. 4 Mean accuracy in all conditions (Experiment 2)

accuracy ($\beta = -0.372$, $SE=0.103$, $z = -3.606$, $p < .001$). Specifically, the left-right target structure was associated with an increase in accuracy compared to the top-bottom reference level.

The LMM for reaction time indicated significant differences among the flanker type conditions. Specifically, responses in the congruent flanker condition were significantly faster than those in the incongruent condition ($\beta=0.011$, $SE=0.004$, $t=2.807$, $p < .01$). Moreover, responses in both congruent ($\beta = -0.025$, $SE=0.006$, $t=-4.44$, $p < .0001$) and incongruent conditions ($\beta=-0.035$,
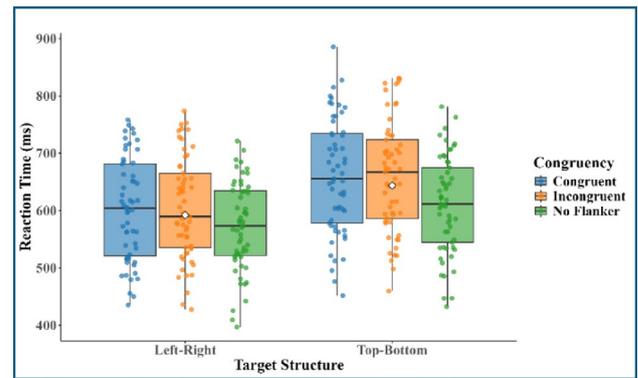

Fig. 5 Mean reaction time in all conditions (Experiment 2)

$SE=0.006$, $t=-6.334$, $p < .0001$) were significantly slower compared to the no-flanker condition. No significant interaction effect was observed in LMM. The target structure had a significant effect on reaction time ($\beta=0.061$, $SE=0.005$, $t=11.9$, $p < .0001$), indicating that the left-right target structure was associated with faster reaction time compared to the top-bottom structure.

The central box represents the interquartile range (IQR) of reaction times (25th to 75th percentile). The middle horizontal line (within the box) indicates the median reaction time for that condition. Vertical lines extending from the box show the range of "typical" data points ($\approx 1.5 \times IQR$ from the box edges). Individual dots represent single participants' mean reaction times for specific condition combinations. White-filled diamonds show group means for each condition.

Experiment 2

Method

Participant

Sixty native Chinese students (47 females 13 males, age: 18–32) were recruited from Beijing Language and Culture University and received 15 yuan for their participation. All participants had normal or corrected-to-normal vision. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Beijing Language and Culture University. Informed consent for participation was obtained from all subjects involved in the study.

Stimuli

A total of 120 high-frequency Chinese characters with clear top-bottom (60) or left-right (60) structures were

systematically paired with flanker characters sharing both orthographic configuration and semantic category (artificial/natural), forming 60 artificial and 60 natural pairs. The semantic relationship between target and flanker characters was quantified using Word2Vec-derived cosine similarity (Li et al. 2018), with mean similarities of 0.193 ± 0.071 for congruent pairs and 0.199 ± 0.01 for incongruent pairs. Statistical comparison revealed no significant difference between conditions ($t = -0.567$, $p = .571$). Only 5% of target–flanker pairs shared one radical in two blocks.

Across two experimental blocks, each target character appeared twice: once with a semantic-incongruent flanker and once with its semantic-congruent flanker. The target character and its flanker maintained phonological dissimilarity in both conditions. An additional 60 high-frequency characters (30 natural/30 artificial, balanced across orthographic types) served as no-flanker baseline stimuli in both blocks.

Within each block, mean target character frequencies were 2.949 ± 0.65, 3.179 ± 0.767, and 2.869 ± 0.8 (baseline condition), with stroke counts of 11.033 ± 3.329, 10.7 ± 3.529, and 10.75 ± 3.062 (baseline condition), respectively. Semantic similarity quantification through Word2Vec embeddings revealed significantly higher cosine similarity in congruent pairs ($M = 0.256$, $SD = 0.097$) compared to incongruent pairs ($M = 0.191$, $SD = 0.059$), $t = 6.235$, $p < .001$. Examples of stimuli used in Experiment 2 are shown in the right panel of Table 1. For the visual similarity between flankers and target characters in Experiment 2, the indices for Block 1 were as follows: SSIM (congruent vs. incongruent) $= 0.413 \pm 0.076$ vs. 0.395 ± 0.062, MSE (congruent vs. incongruent) $= 16.095 \pm 2.412$ vs. 16.463 ± 2.408, and PSNR (congruent vs. incongruent) $= 36.116 \pm 0.692$ vs. 36.014 ± 0.666. In Block 2, the indices were: SSIM (congruent vs. incongruent) $= 0.394 \pm 0.077$ vs. 0.388 ± 0.066, MSE (congruent vs. incongruent) $= 16.576 \pm 2.207$ vs. 16.286 ± 2.068, and PSNR (congruent vs. incongruent) $= 35.978 \pm 0.628$ vs. 36.049 ± 0.571.

Procedure and data analyses

The procedure and data analysis closely followed Experiment 1, except that all analyses incorporated the semantic category of target characters (natural vs. artificial [manmade]) and its interaction with structure type and congruency.

Results

Three participants were excluded from the analysis because of their low accuracy (34.72%, 53.06%, and 62.22%). Prior to analysis, 41 trials ($\sim 0.2\%$) without responses and 407 trials ($\sim 1.99\%$) with reaction time exceeding three standard deviations from the mean were excluded. Additionally, 1943 incorrectly responded trials ($\sim 9.68\%$) were removed from the reaction time analysis.

Mean accuracy ($\pm SD$) across conditions was as follows: semantically congruent (0.895 ± 0.307), incongruent (0.900 ± 0.300), and no-flanker baseline (0.915 ± 0.279); characters with top-bottom structure (0.886 ± 0.317) and left-right structure (0.920 ± 0.272); and natural-category (0.900 ± 0.300) versus artificial-category targets (0.906 ± 0.292).

The binomial GLMM for accuracy initially showed significant main effects of congruency ($\beta = 0.281$, $SE = 0.125$, $z = 2.252$, $p = .024$) and semantic category ($\beta = 0.347$, $SE = 0.174$, $z = 1.995$, $p = .046$). However, since significant interaction effects were detected, these main effects cannot be interpreted as overall patterns and are conditional on other variables being at their baseline levels (treatment coding: semantic category = natural, structure = left-right). To evaluate the overall contributions of fixed effects, we performed Type III ANOVA for the GLMM using the *car* package. The results showed no significant main effects of target structure ($\chi^2 = 2.716$, $p = .099$) or flanker type ($\chi^2 = 5.382$, $p = .068$), but a significant main effect of semantic category ($\chi^2 = 3.981$, $p = .046$), suggesting a higher accuracy for semantic category of artificial objects compared with natural objects. A significant flanker type $\times$ structure interaction ($\chi^2 = 8.441$, $p = .015$) was observed. To unpack the significant interaction, we conducted simple effects analyses using the *emmeans* package, comparing flanker type levels within each structure. P-values were adjusted via the Tukey method for multiple comparisons. For left-right structure, no significant differences between flanker type levels were observed (all $ps > 0.05$). While for the top-bottom structure, both incongruent vs. no-flanker ($\beta = -0.424$, $SE = 0.131$, $z = -3.24$, $p < .01$) and congruent vs. no-flanker ($\beta = -0.39$, $SE = 0.131$, $z = -2.971$, $p < .01$) contrasts were significant, indicating that flankers reduced accuracy only when characters were arranged vertically (top-bottom).

The linear mixed-effects model (LMM) for reaction time revealed significant effects of character structure ($\beta = 0.051$, $SE = 0.007$, $t = 7.285$, $p < .0001$), with left-right structured characters showing faster responses than top-bottom arrangements. Mirroring Experiment 1, both congruent ($\beta = -0.031$, $SE = 0.008$, $t = -4.09$, $p < .0001$) and incongruent flanker conditions ($\beta = -0.029$, $SE = 0.008$, $t = -3.75$, $p < .001$) exhibited significantly slower reaction

times compared to the no-flanker baseline. However, these main effects were context-dependent, as evidenced by the interaction effects (full interaction statistics were omitted and made available in OSF supplementary materials). Type III ANOVA confirmed a robust main effect of structure ($\chi^2 = 53.066$, $p < .001$) and flanker type ($\chi^2 = 17.447$, $p < .001$), while only marginal significance emerged for the three-way structure × flanker type× semantic category interaction ($\chi^2 = 5.986$, $p = .0501$). Post-hoc simple effects analysis (Tukey-adjusted) demonstrated significant slowing in both incongruent ($\beta = 0.04$, $SE = 0.004$, $z = 10.529$, $p < .0001$) and congruent ($\beta = 0.039$, $SE = 0.004$, $z = 10.259$, $p < .0001$) conditions relative to no-flanker trials, whereas the congruent-incongruent contrast was not significant ($\beta = -0.001$, $SE = 0.003$, $z = -0.4$, $p = .916$). This pattern diverges from Experiment 1, where congruent-incongruent differences reached significance.

Discussion

The influential Lexical Constituency Model (LCM) for Chinese word recognition proposes that orthography, phonology, and semantics are partially separable yet tightly coordinated constituents, with phonological activation arising rapidly from orthographic input and contributing to lexical access (Perfetti et al. 2005). Although often contrasted with cascade-style activation in alphabetic scripts, LCM characterizes character identification as occurring once orthographic evidence reaches a functional identification criterion, at which point phonological and semantic constituents become strongly engaged (Li et al. 2022). Much of the empirical support for LCM has come from paradigms that isolate a single word, including foveal priming, naming, and lexical decision, which emphasize sequential activation and foveal processing. However, during natural reading multiple words fall within the perceptual span, and orthographic, phonological, and semantic information from foveal and parafoveal regions can be processed at least partially in parallel, as demonstrated in previous studies (Pan et al. 2024; Snell & Grainger, 2019). Under the logic of constituency activation, parafoveal orthographic or phonological evidence could raise the activation of a candidate sufficiently to facilitate foveal identification, and concurrent signals from neighboring words may jointly contribute to lexical activation or sharpen competition. Extending constituency-based accounts to parallel multi-word input is therefore essential for evaluating their ecological validity.

In the current study, to summarize, when the target character was simultaneously presented with two distractors for a brief time (50ms), the judgement of the target's visual structure (left-right structure versus top-down structure)

was facilitated when the target and the distractors were phonologically similar (compared to phonologically dissimilar trials, experiment 1), but no facilitation was observed when the items were semantically similar (compared to semantically dissimilar trials, experiment 2). Since 50 ms is too short to support sequential processing, these patterns are best attributed to parallel processing across concurrently presented items. Under this interpretation, the findings refine constituency-based accounts in two ways. First, phonological information can be integrated rapidly enough across multiple items to influence a task that does not explicitly require phonological access, suggesting that early phonological engagement operates under parallel input rather than being restricted to serial, single-item processing. Second, no detectable semantic compatibility effect emerged under the same brief, simultaneous conditions when meaning was not task-diagnostic, indicating that early semantic involvement may be more sensitive to factors such as exposure duration, evidence strength, or task relevance. Besides this, our results had several implications.

First, our task asked participants to report the structure of the target character, which neither needs to access its phonology nor needs to access its semantics. Nevertheless, the effect of phonology was observed. This suggests that the effects of phonology observed in previous literature were not task-specific and were automatically activated once the visual form of the character was encountered. Furthermore, in Mandarin Chinese, the left-right arrangement and the top-down arrangement are two common structures, so accessing the structure of a character is not informative enough to differentiate the character from other characters. Regardless of the limitations of the structure judgment, the effect of the phonology is still observed. This suggests that activation of the phonological property does not need to wait for the orthographical property to be completely processed (Perfetti and Tan 1998).

Second, the dissociation between the phonology and semantics was incompatible with the feedforward account (Carreiras et al. 2014) of visual word recognition. According to this account, the lexical representation and the semantics are directly accessed from the orthography of the target character, and the access of the lexical representation is then used to activate the character's phonology. If this account is on the right track, the effect of the phonology should be accompanied by the effect of the semantics. The phonology of the distractors, but not their semantics, was found to affect the judgment of the target's visual structure, implying that the feedforward account is inappropriate to explain the current results.

Third, the early access of the phonology but not the semantics cannot be used to determine whether the phonology mediates the orthography and the semantics. On one

hand, it is compatible with the account that the semantics of a visual word is mediated by its phonology, which is activated by its orthography (Li et al. 2022). According to this account, if the exposure time is too brief (50ms in the current study), the processing might only arrive at the phonology node, but does not arrive at the semantics node. This will result in the phenomenon that the effect of the phonology was observed, but not the effect of the semantics. On the other hand, it is also compatible with the account that the phonology and the semantics are simultaneously processed, but the semantics might need more time to process (Li et al. 2022). This will also result in the phenomenon that the effect of the phonology was observed, but not the effect of the semantics. Further experiments are needed. A first line is to lengthen the presentation time in the current flanker paradigm (and/or manipulate SOA) to test whether a semantic congruency effect would appear. Although a similar syntactic congruency effect has been observed in English with a 50-ms presentation, this may be too short for semantic activation in Mandarin Chinese; the semantic congruency effect is therefore expected to emerge with a longer presentation time. A second line is to use a semantic judgment task and phonology–semantic congruent pairs to differentiate the two possible accounts. When comparing performance (or brain-imaging indices) between phonology–semantic congruent pairs and semantic-only or phonology-only congruent pairs, the advantage of phonology–semantic congruent pairs may be larger under the mediated account but minimal under a direct/parallel account in a semantic-judgment task.

Besides the implications above, the phonological effect observed in the current study mirrored previous findings using priming paradigms. However, due to the substantial differences between the two paradigms, we propose that the congruency effects in priming and flanker paradigms may not be conceptually equivalent. Although both paradigms show similar congruency effects and influences from non-target characters on target processing, their underlying mechanisms differ. Effects in priming paradigms are typically explained by spreading activation or expectancy, owing to the sequential presentation of primes and targets (Collins & Loftus, 1975; Neely, 1977). In contrast, effects in flanker paradigms are usually interpreted as evidence for parafoveal integration and parallel processing of both foveal and parafoveal information. We therefore interpret our effect as converging and complementary, rather than redundant, with priming-based evidence. Nevertheless, the exact mechanisms underlying these effects remain unclear. Moreover, since previous priming studies did not employ structure judgment tasks, a direct comparison with the present study is not feasible.

Limitation Although the no-flanker condition informs single-character processing without flanker-induced interference, the present design lacks a neutral baseline. As a result, it remains ambiguous whether the congruent condition reflects facilitation or the incongruent condition reflects interference. Future work would benefit from including a neutral-flanker baseline to adjudicate between these alternatives.

Author contributions Xingyu Yang: Conceptualization, Investigation, Methodology, Resources, Software, WritingLikan Zhan: Conceptualization, Methodology, Resources, Funding acquisition, Project administration, Supervision, Writing;

References

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Usinglme4. J Stat Softw 67(1). https://doi.org/10.18637/jss.v067.i01

Cai Q, Brysbaert M (2010) SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. PLoS ONE 5(6):e10729. https://doi.org/10.1371/journal.pone.0010729

Carreiras M, Armstrong BC, Perea M, Frost R (2014) The what, when, where, and how of visual word recognition. Trends Cogn Sci 18(2):90–98. https://doi.org/10.1016/j.tics.2013.11.005

Chua FK (1999) Phonological recoding in Chinese logograph recognition. J Experimental Psychology: Learn Memory Cognition 25(4):876–891. https://doi.org/10.1037/0278-7393.25.4.876

Collins AM, Loftus EF (1975) A spreading-activation theory of semantic processing. Psychol Rev 82(6):407–428. https://doi.org/10.1037/0033-295X.82.6.407

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J (2001) DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychol Rev 108(1):204–256. https://doi.org/10.1037/0033-295x.108.1.204

Dare N, Shillcock R (2013) Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. Q J Experimental Psychol 66(3):487–504. https://doi.org/10.1080/17470218.2012.703212

Duan X, Cai ZG (2024) Chinese character processing. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.1012

Eriksen BA, Eriksen CW (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 16(1):143–149. https://doi.org/10.3758/bf03203267

Fox J, Weisberg S (2019) An R Companion to Applied Regression. Sage

Grainger J, Mathot S, Vitu F (2014) Tests of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. Acta Psychol 146:35–40. https://doi.org/10.1016/j.actpsy.2013.11.014

Harm MW, Seidenberg MS (2004) Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. Psychol Rev 111(3):662–720. https://doi.org/10.1037/0033-295X.111.3.662

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Softw 82(13). https://doi.org/10.18637/jss.v082.i13

Lenth RV (2025) *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://rvlenth.github.io/emmeans/

Li S, Zhao Z, Hu R, Li W, Liu T, Du X (2018), July Analogical Reasoning on Chinese Morphological and Semantic Relations. In I. Gurevych & Y. Miyao, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* Melbourne, Australia

Li X, Huang L, Yao P, Hyönä J (2022) Universal and specific reading mechanisms across different writing systems. Nat Reviews Psychol 1(3):133–144. https://doi.org/10.1038/s44159-022-00022-6

Meade G, Declerck M, Holcomb PJ, Grainger J (2021) Parallel semantic processing in the flankers task: Evidence from the N400. Brain Lang 219:104965. https://doi.org/10.1016/j.bandl.2021.104965

Neely JH (1977) Semantic priming and retrieval from lexical memory: Roles of inhibition less spreading activation and limited-capacity attention. J Exp Psychol Gen 106(3):226–254. https://doi.org/10.1037/0096-3445.106.3.226

Pan Y, Frisson S, Federmeier KD, Jensen O (2024) Early parafoveal semantic integration in natural reading. eLife 12:RP91327. https://doi.org/10.7554/eLife.91327

Perfetti CA, Liu Y, Tan LH (2005) The lexical constituency model: Some implications of research on Chinese for general theories of reading. Psychol Rev 112(1):43–59. https://doi.org/10.1037/0033-295X.112.1.43

Perfetti CA, Tan LH (1998) The time course of graphic, phonological, and semantic activation in Chinese character identification. J Experimental Psychology: Learn Memory Cognition 24(1):101–118. https://doi.org/10.1037/0278-7393.24.1.101

Perfetti CA, Zhang S (1991) Phonological processes in reading Chinese characters. J Experimental Psychology: Learn Memory Cognition 17(4):633–643. https://doi.org/10.1037/0278-7393.17.4.633

Perfetti CA, Zhang S (1995) Very early phonological activation in Chinese reading. J Experimental Psychology: Learn Memory Cognition 21(1):24–33. https://doi.org/10.1037/0278-7393.21.1.24

Perfetti CA, Zhang S, Berent I (1992) Reading in English and Chinese: Evidence universal phonological principle. In: Frost R, Katz L (eds) Orthography, Phonology, Morphology, and Meaning. Elsevier

Qu J, Pang Y, Liu X, Cao Y, Huang C, Mei L (2022) Task modulates the orthographic and phonological representations in the bilateral ventral Occipitotemporal cortex. Brain Imaging Behav 16(4):1695–1707. https://doi.org/10.1007/s11682-022-00641-w

R Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing

Schotter ER, Payne BR (2019) Eye movements and comprehension are important to reading. Trends Cogn Sci 23(10):811–812. https://doi.org/10.1016/j.tics.2019.06.005

Snell J (2024) The reading brain extracts syntactic information from multiple words within 50 milliseconds. Cognition 242:105664. https://doi.org/10.1016/j.cognition.2023.105664

Snell J, Declerck M, Grainger J (2017a) Parallel semantic processing in reading revisited: Effects of translation equivalents in bilingual readers. Lang Cognition Neurosci 33(5):563–574. https://doi.org/10.1080/23273798.2017.1392583

Snell J, Grainger J (2018) Parallel word processing in the flanker paradigm has a rightward bias. Atten Percept Psychophys 80(6):1512–1519. https://doi.org/10.3758/s13414-018-1547-2

Snell J, Grainger J (2019a) Consciousness is not key in the serial-versus-parallel debate. Trends Cogn Sci 23(10):814–815. https://doi.org/10.1016/j.tics.2019.07.010

Snell J, Grainger J (2019b) Readers are parallel processors. Trends Cogn Sci 23(7):537–546. https://doi.org/10.1016/j.tics.2019.04.006

Snell J, Mathot S, Mirault J, Grainger J (2018) Parallel graded attention in reading: A pupillometric study. Sci Rep 8(1):3743. https://doi.org/10.1038/s41598-018-22138-7

Snell J, Meade G, Meeter M, Holcomb P, Grainger J (2019) An electrophysiological investigation of orthographic spatial integration in reading. Neuropsychologia 129:276–283. https://doi.org/10.1016/j.neuropsychologia.2019.04.009

Snell J, Meeter M, Grainger J (2017b) Evidence for simultaneous syntactic processing of multiple words during reading. PLoS ONE 12(3):e0173720. https://doi.org/10.1371/journal.pone.0173720

Snell J, Vitu F, Grainger J (2017c) Integration of parafoveal orthographic information during foveal word reading: beyond the sublexical level? Q J Experimental Psychol 70(10):1984–1996. https://doi.org/10.1080/17470218.2016.1217247

Tan LH, Hoosain R, Peng D-l (1995) Role of early presemantic phonological code in Chinese character identification. J Experimental Psychology: Learn Memory Cognition 21(1):43–54. https://doi.org/10.1037/0278-7393.21.1.43

Tan LH, Hoosain R, Siok WWT (1996) Activation of phonological codes before access to character meaning in written Chinese. J Experimental Psychology: Learn Memory Cognition 22(4):865–882. https://doi.org/10.1037/0278-7393.22.4.865

Tan LH, Perfetti CA (1998) Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. Read Writ 10(3/5):165–200. https://doi.org/10.1023/a:1008086231343

White AL, Boynton GM, Yeatman JD (2019) You can't recognize two words simultaneously. Trends Cogn Sci 23(10):812–814. https://doi.org/10.1016/j.tics.2019.07.001

Xu Y, Pollatsek A, Potter MC (1999) The activation of phonology during silent Chinese word reading. J Experimental Psychology: Learn Memory Cognition 25(4):838–857. https://doi.org/10.1037/0278-7393.25.4.838

Yeatman JD, White AL (2021) Reading: The confluence of vision and language. Annual Rev Vis Sci 7:487–517. https://doi.org/10.1146/annurev-vision-093019-113509

Yu R, Wu Y, Gu F (2023) Parallel phonological processing of Chinese characters revealed by flankers tasks. Front Psychol 14:1239256. https://doi.org/10.3389/fpsyg.2023.1239256

Zhang S, Perfetti CA, Yang H (1999) Whole word, frequency-general phonology in semantic processing of Chinese characters. J Experimental Psychology: Learn Memory Cognition 25(4):858–875. https://doi.org/10.1037/0278-7393.25.4.858

Zhou X, Marslen-Wilson W, Taft M, Shu H (1999) Morphology, orthography, and phonology reading chinese compound words. Lang Cogn Processes 14(5–6):525–565. https://doi.org/10.1080/016909699386185