Foveal and Parafoveal Processing of Chinese Three-character Idioms in Reading

Chuanli Zang[1,2], Ying Fu[2], Xuejun Bai[2], Guoli Yan[2] & Simon P. Liversedge[1]

[1] University of Central Lancashire; [2] Tianjin Normal University

Send Correspondence to:

Chuanli Zang,

School of Psychology, University of Central Lancashire,

Preston, Lancashire, PR1 2HE, UK

Email: czang@uclan.ac.uk

Running Head: Preview effects in reading of Chinese idioms

All data sets and analysis scripts are publicly available at: https://osf.io/bp2d8/.

Abstract

Chinese idioms are likely to be represented and processed as Multi-Constituent Units (MCUs, a multi-word unit with a single lexical representation, see Zang, 2019). Chinese idioms with a 1-character verb and 2-character noun structure are processed foveally, but not parafoveally, as a single lexical unit (Yu et al., 2016), probably because the verb only loosely constrains noun identity. By contrast, Chinese idioms with modifier-noun structure are more likely MCU candidates due to significant modifier constraint over the subsequent noun. We investigated whether idioms of this type are parafoveally and foveally processed as MCUs during natural reading. In Experiment 1, we manipulated phrase type (idiom or matched phrase) and preview of the noun (identity, unrelated character or pseudocharacter) using the boundary paradigm (Rayner, 1975). A larger preview effect occurred for idioms on the modifier with shorter fixations for identical than unrelated and pseudocharacter previews. This suggests idioms are parafoveally processed to a greater extent than matched phrases. In Experiment 2, preview of the modifier and noun of idioms and phrases (identity or pseudocharacter) was orthogonally manipulated (c.f., Cutter, Drieghe & Liversedge, 2014). For identity modifiers, a greater noun preview effect occurred for idioms relative to phrases providing further evidence that modifier-noun idioms are lexicalised MCUs and processed parafoveally as single, unified representations.

Keywords: Multi-Constituent Units, preview effects, eye movements, idioms, Chinese reading.

Eye-tracking methodology is now firmly established as an extremely valuable tool in revealing central aspects of cognitive processing underlying on-line sentence reading. Over the past few decades, it has been demonstrated that readers not only process the word under fixation, the foveal word (word *n*), but they also process the upcoming, parafoveal words (e.g., word *n*+1) prior to their fixation (Rayner, 2009; Schotter, Angele, & Rayner, 2012). However, there has been debate regarding the extent to which the upcoming words are lexically processed prior to fixation, with a question of central focus being whether multiple words (e.g., words *n*+1 and *n*+2) are lexically identified serially and sequentially, or in parallel during sentence reading? This debate concerning whether words are lexically identified serially and sequentially, or in parallel, has dominated research on eye movement control during reading over many years. It has even been argued that it might be difficult, or potentially impossible, to conciliate the issue through experimentation using eye-tracking methodology to investigate natural sentence reading (Snell & Grainger, 2019). However, recently a possible (at least, partial) solution to this theoretical impasse has been proposed via the Multi-Constituent Unit (MCU) Hypothesis (Zang, 2019), such that some linguistic units comprised of multiple words such as idioms may be lexicalized and stored as single lexical representations. The constituent words of MCUs may, therefore, be identified simultaneously (in parallel), though lexical processing proceeds sequentially with lexical representations being identified serially. In other words, the lexicality status of linguistic units determines whether sequences of words are processed serially or in parallel. In this context, the purpose of the present study was to use eye-tracking methodology to provide direct evidence in support of this hypothesis.

Traditionally, there are two leading models of eye movement control during reading of alphabetic languages, the E-Z Reader (Reichle, 2011; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003) and SWIFT (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005; Engbert & Kliegl, 2011). Both models assume that lexical processing is the engine that drives the eyes to move during reading. However, according to E-Z Reader, words are lexically identified in a strictly serial order, one at a time, with attention not shifting to upcoming parafoveal words until prior words have been fully identified. In other words, only when identification of the currently fixated word *n* is completed does parafoveal processing of word *n*+1 start, and this occurs before the eyes move to fixate word *n*+1 (Reichle, 2011). Ordinarily, parafoveal processing of word *n*+2 should not start while the eyes are still on word *n*, except when word *n*+1 is very short, and/or highly frequent, and thus very easy to process. Under this circumstance, word *n*+1 is identified quickly in the parafovea, allowing word *n*+2 to be preprocessed simultaneously with the preceding words and prior to a saccade being made directly to *n*+2, that is, skipping word *n*+1. In contrast to these serial processing stipulations, SWIFT (see also the more recently proposed model, OB1, Snell, van Leipsig, Grainger, & Meeter, 2018), adopts a parallel graded attention mechanism, and proposes that multiple words around the point of fixation that fall within the perceptual span can be lexically processed in parallel.

According to this approach, word *n*+2 will be automatically processed at the same time as word *n* and word *n*+1 as long as they all fall within the perceptual span.

The previous empirical evidence from studies investigating reading of alphabetic languages has demonstrated that word *n*+2 preview effects are negligible, or at best, very subtle and only appear when word *n*+1 is high-frequency and/or very short (not longer than three letters) (e.g., Angele & Rayner, 2011; Kliegl, Risse, & Laubrock, 2007; Radach, Inhoff, Glover, & Vorstius, 2013). Effects do not appear to occur when word *n*+1 is longer than three letters (e.g., Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Rayner, Juhasz, & Brown, 2007, see Vasilev & Angele, 2017 for a review). Advocates of parallel lexical identification accounts argue that this is likely because a longer word *n*+1 ensures that word *n*+2 falls out the perceptual span, preventing it from being efficiently and effectively preprocessed during the time that the eyes remain fixating word *n*.

From these considerations, it should be apparent that the controversies that exist between these models center on the critical question of whether lexical processing is operationalized on one, or multiple words simultaneously, and whether there are factors that might modulate such operationalization. Two somewhat self-evident but related questions concern (1) what a “word” is; that is, what are the constituents of a text that are represented lexically and processed as single units during reading, and (2) how do lexical identification processes that occur over time relate to the fixations that are made across those lexical units as they are read. To be clear here, unless it is possible to know without ambiguity the linguistic element, or elements, that are being lexically processed as unitary constituents of the sentence during any particular fixation, then it is simply impossible to directly investigate the question of whether lexical identification occurs serially or in parallel. To be clear, the issue of exactly which elements of the sentence are represented lexically as single units, and consequently, identified via a single cognitive event (i.e., the process of lexical identification), is critical to concerns regarding how processing is distributed spatially (via fixations) over those units.

In most alphabetic languages (e.g., English, German, French, etc.), a lexical unit is often straightforwardly identifiable as being the letter string that is clearly demarcated by spaces on either side. Indeed, this visual (and the corresponding linguistic) characterization of a word in alphabetic languages has fundamentally shaped the theoretical framework within which models of eye movement control have been developed to date. These underlying theoretical assumptions are usually implicit, but have far reaching and constraining implications. Regardless of whether lexical processing is serial and sequential, or parallel and non-sequential, a primary assumption in models of reading is that processing is word based, that is, the “currency” of the process of written sentence comprehension is the word.

Over recent years, there has been increasing interest in how readers process non-alphabetic languages, with Chinese as an example of such a language receiving considerable attention (see Li, Zang, Liversedge, & Pollatsek, 2015; Zang, Liversedge, Bai, & Yan, 2011 for reviews). The increased interest has largely come about due to the characteristics of the written forms of such languages, and how these characteristics offer the opportunity to explore theoretical issues that it is simply impossible to investigate in alphabetic languages. Chinese is a character based, unspaced language. Text is comprised of characters, each with a small amount of equally distributed space on each side. One or more characters form a word, but there are no visual cues such as spaces to demarcate where each word begins and ends. There is often ambiguity regarding where word boundaries lie within a sentence, and readers frequently fail to discriminate words from phrases (e.g., He et al., 2021; Hoosain, 1992; Li, Zang, Liversedge, & Pollatsek, 2015; Liu, Li, Lin, & Li, 2013; Zang, Liversedge, Bai, & Yan, 2011). Given these characteristics of written Chinese and facing the question of whether words are lexically identified serially or in parallel, it becomes essential to consider what units parafoveal preprocessing operates over during reading. Furthermore, it should be apparent that how a Chinese reader lexically represents multi character parafoveal strings, that is, either as a series of discrete and separate words, or instead as a single lexical unit, would be a determinant of whether processing of those characters might be best characterized as involving serial or parallel lexical processing. The key point to note here is that whilst it is very well established that words are critical units over which processing operates during reading, when considering such processing in relation to an unspaced character based language with abundant word boundary ambiguity (e.g., Chinese), it becomes necessary to consider whether lexical processing might sometimes be operationalized over linguistic units that are larger than the word. This might particularly be the case when word boundaries are ambiguous and not visually demarcated.

Recently, a model of Chinese word identification and eye movement control during reading, the Chinese Reading Model (CRM), has been proposed by Li and Pollatsek (2020). The CRM adopts the interactive activation framework (McClelland & Rumelhart, 1981), and assumes that all characters (that might comprise one, two, or more words) within the perceptual span are processed in parallel, and these in turn activate all the possible words comprised of the activated characters (with position specificity maintained). The word units (that may be spatially overlapping) compete with each other until a word wins the competition, at which point it is identified and simultaneously segmented (i.e., word boundary commitments are made). Once it is identified, the eyes saccade forward to the next character beyond the current position, whereupon lexical competition resumes such that all the words in a text are identified. Thus, according to the CRM, lexical identification and word segmentation is part of a unified process, that occurs for each word along the text. The engine that drives the eyes to move forward is lexical identification. However, the CRM, in its current form, stipulates that the units over which visual, linguistic and oculomotor control processes operate are single, individual words. In its current form, the CRM does not offer an account of how readers identify multiple words simultaneously during Chinese reading.

According to the Multi-Constituent Unit (MCU) Hypothesis (see Zang, 2019), some linguistic units that are comprised of multiple words, such as highly familiar phrases, idioms, spaced compounds, and other units, might be lexicalized and represented as a single lexical unit. Under this hypothesis, the mental lexicon is comprised of lexical entries corresponding to individual words, and these would sit alongside entries corresponding to multi-constituent units, that is, lexical representations corresponding to more than one word and each with its own semantic representation. If the multiple individual words are represented as a single lexical entry, then these words might be processed simultaneously in the parafovea and ultimately identified via the activation of that single lexical representation. This would mean that lexical processing progresses sequentially, one constituent at a time. Crucially, the MCU Hypothesis assumes that the lexicality status of multiple word strings determines how they are processed, and in this way, therefore, the MCU Hypothesis offers a way of (at least partially) reconciling the deadlock between accounts in which words are identified serially, and those in which words are identified in parallel.

Cutter, Drieghe and Liversedge (2014) provided evidence suggesting that English spaced compounds operate as MCUs in relation to parafoveal and foveal processing during reading. In their study, they investigated whether a word *n*+2 preview effect could be obtained when word *n*+1 (e.g., *teddy*) was part of a larger, frequently co-occurring, linguistic unit with word *n*+2 (e.g., *teddy bear*). Importantly, Cutter et al. ensured that the words comprising the spaced compounds were longer than 3 letters and not of particularly high frequency (characteristics argued to result in adjacent parafoveal words being identified prior to direct fixation). Cutter et al. orthogonally manipulated the preview of each constituent (*n*+1 and *n*+2) of a spaced compound (e.g., *teddy bear*) such that it was either a nonword or the identity, using the boundary paradigm with the boundary positioned before the first constituent (*teddy*). When the eyes crossed the boundary, each constituent was displayed correctly. Cutter et al. found an interaction between the two previews with a significant *n*+2 preview effect only when word *n*+1 was parafoveally available. Thus, Cutter et al. demonstrated that reliable n+2 preview effects could be obtained when word n+1 was longer than three letters and not of very high frequency. This finding undermines the suggestion that a longer word *n*+1 causes word *n*+2 to be positioned further from fixation in the parafovea making it perceptually more difficult to process. Instead, Cutter et al. argued that because word *n*+1 and *n*+2 were constituents of a spaced compound, processing of word *n*+1 licenses parafoveal processing of word *n*+2 as part of a two constituent MCU which itself is lexically represented. According to this account, processing of the two words is licensed at a linguistic level. Cutter et al.'s study suggests that the multi-word lexicalised unit "*teddy bear*" is being processed in the parafovea in its entirety as a single unit of information, though note that this only occurs when "*bear*" is preceded (and processing is therefore licensed) by "*teddy*".

In respect of Chinese reading, a recent study has shown that frequently used two-constituent phrases are processed as MCUs in reading. Zang, Du, Bai, Yan and Liversedge (2020) directly manipulated the linguistic category of a two-constituent Chinese string: a word, a frequently used two-character phrase that was considered very likely to be a MCU, and a phrase (the target strings were very strictly prescreened to ensure their linguistic categorization). The boundary paradigm (Rayner, 1975) was used to manipulate the preview of the second constituent: identity or pseudocharacter. The boundary was positioned before the two-constituent Chinese string. The results showed a reliable preview effect on the first constituent when it, along with the second constituent together, formed a word or a MCU but not when it formed a phrase. Note that there is considerable parallel between the experiment of Zang et al, and that of Cutter et al., that is, in Zang et al.'s MCU condition, the preview effect associated with the second constituent resembles the word *n*+2 preview effect that Cutter et al. obtained for the second constituent of their spaced compound target string. Thus, Zang et al. argued that their findings further support the MCU Hypothesis, demonstrating that lexical processing can be operationalized over multiple word units during sentence reading.

In another directly related study, Yu et al. (2016) investigated word *n*+2 preview effects in three-character Chinese idioms and matched phrases with a 1-character verb and 2-character noun "1+2" structure. Chinese idioms are fixed phrases that consist of two or more words and for which an abstract meaning is usually retrieved from memory for the whole unit rather than being derived piecemeal from the individual meanings of the constituent words. Idioms, therefore, are very likely to be represented and processed as MCUs. Using the boundary paradigm, Yu et al. manipulated preview of the second constituent of a "1+2" verb-noun idiom (e.g., 揭疮疤, figuratively meaning *expose others' secrets*, and literally meaning *pick a scab*) and the matched phrases (e.g., 留疮疤, meaning *have a scar*) to be correct (identities of the target constituents) or incorrect (unrelated characters). They found reliable effects of phrase type with shorter reading time for idioms than phrases, indicating that these idioms were processed faster than the matched phrases. However, there was no evidence suggesting that the second constituent was processed to a greater extent when it was a part of an idiom than when it was processed as part of a phrase in the parafovea. It was concluded that idioms with verb-noun structure are processed foveally, but not parafoveally, as a single lexical unit.

According to the MCU hypothesis, if a MCU is represented and processed as a single lexical entry, the first constituent should activate such a lexical unit, then this activation will license processing to extend further into the parafovea, resulting in more efficient parafoveal preprocessing of the second constituent. Accordingly, then, it might have been reasonable to expect parafoveal, as well as foveal, effects in the Yu et al. study. In our view, however, these results very likely occurred due to the characteristics of the particular idioms that were adopted in the study. In the Yu et al. stimuli, the verb only loosely constrained the identity of the subsequent noun. According to the Contemporary Chinese dictionary (2008), among Chinese three-character idioms, 40% have a verb-object (VO) structure such as 炒鱿鱼 ('炒' means *fry*, '鱿鱼' means *squid*, and together the constituents 炒鱿鱼figuratively means *fire somebody*), while 57% have a modifier-noun (MN) structure. Idioms with MN structure are either comprised of a 2-character modifier and a 1-character noun such as 乌纱帽 ('乌纱' means *black gauze*, '帽' means *cap*, '乌纱帽' means *an official post*), or a 1-character modifier and a 2-character noun such as 铁饭碗 ('铁' means *metal*, '饭碗' means *rice bowl*, '铁饭碗' means *a secure job*). Idioms with modifier-noun structure particularly for those in "2+1" format, are more likely MCU candidates, and thus, more likely to be lexicalized due to the significant constraint the modifier exerts over potential subsequent nouns.

The previous literature regarding processing of formulaic sequences, and particularly, the lexical representation of idioms (e.g., Carrol & Conklin, 2017; Conklin & Schmitt, 2008, 2012; Siyanova-Chanturia, Conklin & Schmitt, 2011; Swinney & Cutler, 1979) has shown that idioms can be stored and accessed as individual lexical units. For example, the Lexical Representation Hypothesis proposed by Swinney and Cutler (1979) proposed that idioms are represented in the mental lexicon in manner similar to that for morphologically complex long words. Once the first constituent of an idiom is encountered, the meaning of the whole unit and its constituents is activated and retrieved simultaneously. Similarly, "non-lexical" models such as the configuration hypothesis (Cacciari & Tabossi, 1988) and hybrid views (e.g., Libben & Titone, 2008; Titone & Connie, 1999) also assume that idioms can be directly retrieved when their constituent words form a recognizable configuration (i.e., sufficient activation has been accumulated to identify the word sequence as an idiom), or are highly familiar or predictable. These theoretical approaches differ regarding when and how idiomatic meanings are directly retrieved, but they all assume that idioms can be represented and processed as a larger, individual, lexical unit at some point.

In the present study we investigated whether idioms with modifier-noun "2+1" structure are processed as MCUs, both parafoveally as well as foveally, during Chinese reading. Two experiments are presented. In Experiment 1, idioms and matched phrases with modifier-noun structure were selected as target strings. The modifier – the first constituent was identical for both types of string. Using the boundary paradigm (Rayner, 1975), preview of the noun, the second constituent, was manipulated to be the identity, an unrelated character or a pseudocharacter. We used unrelated characters and pseudocharacters to demonstrate that they each are comparable and similar as unrelated baseline stimuli in boundary paradigm experiments (something that has not been demonstrated empirically to date). We predicted that these two conditions would pattern identically throughout the experiment. The boundary was located prior to the target string. In line with Yu et al. (2016), we expect a processing advantage of idioms foveally such that they would be fixated for less time compared to matched phrases. Furthermore, if these idioms are parafoveally processed to a greater extent than phrases, that is, they are processed parafoveally as MCUs, then we should obtain a greater preview benefit effect from the second constituent for idioms than for matched phrases. To extend the findings of Experiment 1, in Experiment 2 we used the target stimuli from Experiment 1, but adopted a similar paradigm to that used by Cutter et al. (2014). We used the boundary paradigm and identity and pseudocharacter stimuli to orthogonally manipulate the preview of the two constituents (i.e., the preview of both the modifier and the noun) of idioms and matched phrases. Again, in line with Cutter et al., we anticipated that if idioms were processed as MCUs, but matched phrases were not, then the preview benefit effect for the second constituent should occur for the identity modifier previews, but not for pseudocharacter modifier previews, and that this effect should be more pronounced for idioms than for matched phrases.

Experiment 1

Method

Participants

One hundred and thirty-two students at Tianjin Normal University (115 females, mean age 20 years) participated in the experiment. The participants were all native Chinese speakers and received monetary compensation for their participation. They had normal or corrected-to-normal vision, and were naive to the purpose of the experiment. They signed an informed consent form before taking part in the experiment.

Apparatus

An SR Research Eyelink 1000 eyetracker recorded participants' eye movements with a sampling rate of 1000Hz. Viewing was binocular, but only the right eye was monitored. Sentences were displayed on a 19-inch DELL CRT monitor with a refresh rate of 150 Hz and a screen resolution of 1024 × 768 pixels. Stimuli were presented in Song font in black on a white background. Participants were seated approximately 65 cm from the monitor. At this viewing distance, each Chinese character corresponded to approximately 1.1 degree of visual angle.

Materials and design

We selected 84 idioms with a two-character modifier (Constituent 1) and a one-character noun (Constituent 2, the third character) structure from the Chinese Idiom Dictionary (2009) and the Modern Chinese Idiom Standard Dictionary (2001). Fifteen participants who did not take part in the formal eye-tracking study were required to provide the figurative meaning of the idioms (which were all figurative) and rate their familiarity on a 5-point scale ("1" = "very unfamiliar", "5" = "very familiar"). Idioms were selected if they were clearly and correctly defined by over 60% (M = 91%, SD=12%) of the participants, and their mean familiarity was 4.2 (SD = 0.4). We also constructed a set of 84 matched phrases with an identical syntactic structure to the idioms. Each idiom and its counterpart matched phrase shared the same first constituent (i.e., the two-character modifier), and differed only in their second constituents (i.e., the one-character noun) with these being controlled for stroke complexity, word frequency and character frequency (*Fs* <1) based on the database developed by Cai and Brysbaert (2010), see Table 1.

Table 1 Statistical properties for the idioms and phrases in Experiment 1

Preview	Property	Idioms	Phrases
Identity of the second constituent (the third character)	Number of strokes	8.6 (3.3)	8.5 (2.9)
	Word frequency	65.6 (86.8)	66.1 (82.7)
	Character frequency	298.5 (694.1)	285.0 (617.0)
Unrelated character	Number of strokes	8.5 (2.7)	8.5 (2.7)
	Word frequency	65.9 (80.2)	65.9 (80.2)
	Character frequency	200.6 (223.9)	200.6 (223.9)
Pseudocharacter	Number of strokes	8.6 (2.6)	8.6 (2.6)

Standard deviations are provided in parentheses. The unit of frequency is per million.

Eighty-four sentence frames were constructed, with each set of targets embedded in the middle of each sentence. The preceding context was identical and neutral for both idioms and phrases (see Figure 1). Thirty participants (15 in each of the target strings) were required to rate the sentence naturalness on a 5 point scale, and the mean was 3.9 (SD = 0.4, 5 = very natural), with no difference between the two target strings (F = 1). In addition, a group of 54 participants conducted a sentence completion task to assess the predictability of the target strings given the preceding context (17 participants), as well as of the second constituent given the preceding context including the first constituent (17 participants), and of the first constituent given the second and the preceding context (20 participants). Both idioms and phrases were unpredictable (M = 0%, SD = 2%), whereas the second constituents of idioms given the first (M = 68%, SD = 31%) were more predictable than those of phrases (M = 0.4%, SD = 1.7%, F = 405), and the first constituents of idioms given the second (M = 10%, SD = 17%) were more predictable than those of phrases (M = 0.8%, SD = 0.4%, F = 30). According to the Center for Chinese Linguistics corpus (http://ccl.pku.edu.cn:8080/ccl_corpus/), the transitional probability of the two constituents (the likelihood of occurrence for the constituents forming the idiom or phrase) for idioms (M = 28.4%, SD = 30.6%) was higher than that for phrases (M = 0.7%, SD = 2.4%, F = 68) [1].

The boundary paradigm (Rayner, 1975) was used to manipulate the preview of the second constituent (the third character) of idioms and phrases with the invisible boundary placed prior to the target string. Once the eyes crossed the boundary, an

[1] A further set of LMM analyses were conducted in which either predictability of the second constituent on the basis of the first, predictability of the first constituent on the basis of the second (should readers lexically identify words in parallel), or transitional probability, was included as a centered continuous covariate, to examine the possibility that they might contribute to the main effects that we were interested in. The LMMs with predictability of the second constituent given the first as a covariate showed that the effect of phrase type became less robust across all measures except for total fixation durations in the first constituent and whole target string analyses. However, all the preview effects and the interactions between preview and phrase types maintained. When including predictability of the first constituent given the second, as well as transitional probability as a covariate, for all measures across all regions, the pattern of results was identical to the results we report here. Our analyses demonstrate clearly that the covariates did not cause our effects. Finally, we also calculated variance inflation factors (VIF) for phrase type and predictability of the second constituent given the first, or predictability of the first constituent given the second. In the first analysis capturing predictability of the second constituent given the first, all the VIFs across all measures in all regions were less than 4.79; in the second analysis capturing the predictability of the first constituent given the second, all VIFs were less than 1.30. These results indicate a moderate or no correlation between phrase type and predictability and that it is appropriate to capture variance associated with predictability and transitional probability by including them in our LMMs as a covariate.

identity, an unrelated character, or a pseudocharacter preview was replaced by the target character (see Figure 1). The unrelated characters were real characters with their stroke complexity, word frequency and character frequency matched with the target characters in the idioms and phrases (*F*s < 1). Pseudocharacters were created using Windows True Font software and looked like real characters but were meaningless. The unrelated, matched character and pseudocharacter previews did not contain semantic or phonetic radicals of the target characters but had a similar number of strokes (*F*s < 1).

Phrase type	Preview type	Sentence
	Identity	小辉充分利用各种|**垫脚石**当上了区域经理。
Idiom	Unrelated character	小辉充分利用各种|**垫脚平**当上了区域经理。
	Pseudocharacter	小辉充分利用各种|**垫脚[illegible]**当上了区域经理。
	Identity	小辉充分利用各种|**垫脚布**让家里保持干净。
Phrase	Unrelated character	小辉充分利用各种|**垫脚平**让家里保持干净。
	Pseudocharacter	小辉充分利用各种|**垫脚[illegible]**让家里保持干净。

Figure 1 An example of the Chinese sentences used in Experiment 1. The first constituent was identical across conditions and the preview of the second constituent was manipulated. The vertical line represents the position of the invisible boundary. When the eyes crossed the boundary, the preview was replaced by the target character (the target strings are in bold, but were presented normally in the experiment). The English translation for the sentence is "Xiaohui made full use of various **stepping stones** to become a regional manager/ Xiaohui made full use of various **floor cloths** to keep house clean".

The design of Experiment 1 was a 2 (Phrase Type: idiom or phrase) × 3 (Preview of the Second Constituent: identity, unrelated character or pseudocharacter) within-participant repeated measures design. Six files were thus constructed, with each file containing 84 sentences (14 sentences in each condition). The experimental conditions across these files were rotated according to a Latin Square, and each participant read 84 experimental sentences (as well as 42 filler sentences without a display change) presented randomly from one of the files. There were eight practice sentences presented prior to the formal experimental sentences. On one third of the trials participants were required to answer comprehension questions with a yes/no response.

Procedure

Before the start of the experiment, each participant was presented with an information sheet and a written consent form. They were instructed that they would read sentences carefully and they were required to try to understand them to the best of their ability. After they finished reading a sentence, they pressed response keys on a button box to terminate the display and answer a yes/no comprehension question that would appear occasionally after a sentence. They were then required to sit in front of the eye tracker and complete a 3-point horizontal calibration procedure, until the calibration resulted in an average error below 0.20 degrees. Once the calibration was completed successfully, the sentences were presented in turn. On each trial, a drift correction dot was presented at the same position as the first character of the upcoming sentence. Participants were required to fixate the dot to trigger the onset of the sentence. After the experiment, the participants were required to report whether they had noticed any changes in relation to the characters or text while they were reading. The whole experiment lasted approximately 30 min.

Results and Discussion

Data from two participants were discarded from analyses due to the low comprehension accuracy (below 85%). The mean comprehension accuracy for all the valid participants was 96% indicating that they fully understood the sentences. With respect to display change awareness, 10 participants were unaware of the display change and the remainder were aware of the display change but were unable to report which characters had changed. We excluded fixation durations that were shorter than 80ms or longer than 1200ms from the analyses, and removed trials if a) track loss occurred or fewer than three fixations were made in total (0.3% of the data); b) blinks occurred during display changes or during a fixation on the target word (3%); and c) the display changes triggered early or late (16%). Finally, we removed any observations for each measure that were above or below three standard deviations from each participant's mean prior to conducting the analyses prior to conducting the analyses (1.5%).

Analyses were carried out for the pre-target word (*n*), the first constituent (*n*+1), the second constituent (*n*+2), the whole target string. For each target region the following eye movement measures were computed: *first fixation duration* (FFD, the duration of the first fixation on a region during first pass reading), *single fixation duration* (SFD, the duration of fixations when only one fixation was made during first pass reading), *gaze duration* (GD, the sum of all fixations on a region from first entering the region until leaving it during first pass reading), *total fixation duration* (TFD, the sum of all fixations on a region, including both forward and regressive fixations), *go-past time* (the sum of all fixations from the first fixation in the region until a fixation to the right of that region – this measure includes any fixations made after any regressions to earlier regions of the sentence prior to the eyes moving to the right of the region to fixate upcoming text), and *skipping probability* (SP, the probability that a region is not fixated during first pass reading). Means and standard deviations for the eye movement measures across all the regions are shown in Table 2.

Linear mixed models (LMM) were conducted using the lme4 package (version 1.1-15) in R (version 3.3.3, R Development Core Team, 2014) to analyze the data. Phrase Type, Preview and their interaction were treated as fixed factors. For the preview condition, successive contrasts were carried out, with comparisons of identity vs unrelated character previews, and unrelated character vs pseudocharacter previews. Participants and items were treated as crossed random factors. For each eye movement measure, we first ran a model with the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) consisting of slopes for all the fixed effects across subjects and items, but trimmed this down for those models that failed to converge. The model was trimmed starting with items and then participants, with the removal of the correlations between factors, interactions, then random slopes until the model converged. Fixation time analyses were carried out using log-transformed data to increase the normality. Note though, analyses for untransformed and log-transformed durations produced very similar pattern of effects. Skipping data were analyzed using logistic GLMMs given the binary nature of the variable. Fixed effect estimations for the eye movement measures across all regions are shown in Table 3. All data sets and analysis scripts are publicly available at the following website: https://osf.io/bp2d8/.

Table 2 Eye movement measures for all regions across the six experimental conditions in Experiment 1.

Analysis region	Phrase	Preview	FFD	SFD	GD	TFD	Go-past	SP
The Pretarget Word (*n*)	Idiom	Identity	227(40)	225(41)	243(54)	335(92)	322(121)	0.39(0.18)
		Unrelated character	226(43)	226(46)	244(58)	327(93)	313(99)	0.39(0.17)
		Pseudocharacter	229(41)	228(41)	249(55)	343(97)	310(92)	0.42(0.19)
	Phrase	Identity	222(41)	221(42)	241(53)	339(101)	308(100)	0.42(0.18)
		Unrelated character	226(41)	226(42)	243(55)	342(100)	310(112)	0.41(0.17)
		Pseudocharacter	224(41)	224(43)	241(53)	351(107)	300(107)	0.40(0.18)
The 1st Constituent (*n+1*)	Idiom	Identity	234(40)	233(42)	257(58)	351(100)	324(108)	0.23(0.19)
		Unrelated character	253(44)	252(46)	288(66)	399(120)	357(118)	0.19(0.18)
		Pseudocharacter	255(47)	254(49)	289(70)	407(136)	350(104)	0.18(0.17)
	Phrase	Identity	245(44)	246(45)	276(66)	445(165)	360(142)	0.20(0.19)
		Unrelated character	256(46)	255(48)	293(72)	484(179)	364(115)	0.19(0.16)
		Pseudocharacter	259(48)	255(49)	303(76)	483(178)	365(124)	0.19(0.20)
The 2nd Constituent (*n+2*)	Idiom	Identity	232(47)	232(47)	235(48)	280(74)	289(110)	0.56(0.16)
		Unrelated character	253(55)	254(56)	260(56)	311(84)	353(139)	0.52(0.19)
		Pseudocharacter	259(58)	258(58)	266(62)	314(83)	342(137)	0.51(0.19)
	Phrase	Identity	256(58)	256(62)	267(65)	347(95)	371(137)	0.49(0.18)
		Unrelated character	276(55)	276(59)	292(62)	383(107)	420(161)	0.46(0.21)
		Pseudocharacter	279(62)	278(61)	293(64)	371(104)	417(154)	0.44(0.21)
The Whole Region	Idiom	Identity	234(39)	231(42)	327(93)	479(161)	417(144)	0.06(0.10)
		Unrelated character	257(44)	260(60)	399(125)	569(195)	493(166)	0.06(0.11)
		Pseudocharacter	259(47)	255(56)	405(123)	578(210)	491(168)	0.05(0.09)
	Phrase	Identity	248(43)	242(49)	388(134)	672(270)	511(199)	0.06(0.10)
		Unrelated character	259(43)	258(59)	452(168)	748(279)	568(214)	0.06(0.09)
		Pseudocharacter	263(46)	260(61)	455(152)	742(280)	570(220)	0.06(0.10)

Note. Standard deviations are provided in parentheses. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration; Go-past = go-past time; SP = skipping probability.

Table 3 LMM analyses for all measures across all regions in Experiment 1.

		Measures																	
		FFD			SFD			GD			TFD			Go-past			SP		
Region	Effect	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
The Pretarget Word (*n*)	Phrase Type	-0.01	0.01	-1.49	-0.01	0.01	-1.54	-0.01	0.01	-1.27	0.02	0.01	1.41	-0.02	0.01	-1.75	0.06	0.04	1.29
	Unrelated character vs. Identity	0.01	0.01	1.07	0.02	0.01	1.50	0.01	0.01	0.98	-0.00	0.01	-0.26	0.00	0.02	0.06	-0.01	0.05	-0.10
	Pseudocharacter vs. Unrelated	-0.00	0.01	-0.09	-0.00	0.01	-0.04	0.00	0.01	0.24	0.03	0.01	1.82	-0.01	0.01	-0.99	0.05	0.05	0.98
	Phrase Type × Unrelated character vs. Identity	0.02	0.02	1.26	0.02	0.02	1.22	0.01	0.02	0.52	0.03	0.03	0.92	0.02	0.03	0.52	-0.05	0.11	-0.42
	Phrase Type × Pseudocharacter vs. Unrelated	-0.03	0.02	-1.42	-0.03	0.02	-1.38	-0.03	0.02	-1.52	-0.03	0.03	-1.03	-0.03	0.03	-1.12	-0.19	0.11	-1.80
The 1st Constituent (*n+1*)	Phrase Type	0.02	0.01	**2.89**	0.02	0.01	**2.37**	0.04	0.01	**3.59**	0.17	0.02	**8.04**	0.04	0.01	**3.60**	-.05	0.06	-0.94
	Unrelated character vs. Identity	0.06	0.01	**7.35**	0.06	0.01	**6.31**	0.08	0.01	**7.51**	0.10	0.01	**7.49**	0.07	0.01	**4.45**	-.20	0.07	**-2.85**
	Pseudocharacter vs. Unrelated	0.01	0.01	1.24	0.01	0.01	0.56	0.02	0.01	1.55	0.01	0.01	0.71	-.00	0.01	-0.25	-.03	0.07	-0.47
	Phrase Type × Unrelated character vs. Identity	-0.03	0.02	-1.77	-0.04	0.02	**-2.18**	-0.05	0.02	**-2.39**	-.05	0.03	-1.76	-.07	0.03	**-2.29**	0.20	0.14	1.43
	Phrase Type × Pseudocharacter vs. Unrelated	-.00	0.02	-0.02	-.01	0.02	-0.54	0.02	0.02	0.92	0.00	0.03	0.17	-.00	0.03	-0.08	0.09	0.14	0.64
The 2nd Constituent (*n+2*)	Phrase Type	0.07	0.01	**5.97**	0.08	0.01	**5.87**	0.10	0.01	**7.38**	0.17	0.02	**7.78**	0.18	0.02	**7.56**	-0.26	0.04	**-5.78**
	Unrelated character vs. Identity	0.08	0.01	**5.68**	0.08	0.01	**5.69**	0.09	0.01	**6.25**	0.10	0.02	**5.86**	0.15	0.02	**7.04**	-0.15	0.05	**-2.78**
	Pseudocharacter vs. Unrelated	-0.00	0.01	-0.10	0.00	0.01	0.09	-0.00	0.01	-0.31	-0.02	0.02	-1.57	-0.02	0.02	-0.76	-0.06	0.05	-1.05

	Phrase Type × Unrelated character vs. Identity	0.01	0.03	0.41	0.01	0.03	0.26	0.02	0.03	0.71	0.01	0.03	0.28	-0.03	0.04	-0.75	0.03	0.11	0.25
	Phrase Type × Pseudocharacter vs. Unrelated	-0.04	0.03	-1.65	-0.03	0.03	-1.06	-0.04	0.03	-1.51	-0.04	0.03	-1.21	-0.01	0.04	-0.19	-0.03	0.11	-0.26
The whole region	Phrase Type	0.03	0.01	**3.15**	0.01	0.01	1.33	0.10	0.01	**6.88**	0.27	0.03	**10.57**	0.13	0.02	**7.60**	-0.00	0.10	-0.04
	Unrelated character vs. Identity	0.07	0.01	**8.13**	0.07	0.01	**6.62**	0.17	0.02	**10.91**	0.16	0.02	**9.89**	0.16	0.02	**9.25**	-0.06	0.12	-0.49
	Pseudocharacter vs. Unrelated	0.01	0.01	1.29	0.01	0.01	0.75	0.02	0.01	1.69	0.01	0.02	0.76	0.00	0.02	0.17	-0.09	0.12	-0.74
	Phrase Type × Unrelated character vs. Identity	-0.04	0.02	**-2.50**	-0.06	0.02	**-2.70**	-0.05	0.03	-1.90	-.04	0.03	-1.60	-0.08	0.03	**-2.56**	0.03	0.24	0.11
	Phrase Type × Pseudocharacter vs. Unrelated	0.00	0.02	0.09	0.02	0.02	0.66	0.01	0.03	0.39	-.00	0.03	-0.15	0.02	0.03	0.66	0.35	0.24	1.42

Note. Significant terms featured in bold, and marginal terms are underlined. *b* = regression coefficient.

The Pretarget Word (*n*)

There were no reliable effects on the pretarget word, providing no evidence for parafoveal-on-foveal effects. The lack of effects at this point in the sentence is not unexpected due to the distance between the fixated pretarget word and the manipulation of the preview of the second constituent (and hence the third character) of the target strings.

The First Constituent (*n* + 1)

For the first constituent analyses, there was a significant effect of phrase type in all fixation time measures such that readers spent less time processing idioms than phrases (all $t > 2.36$), demonstrating a foveal processing advantage for Chinese idioms over matched phrases during reading (Yu et al., 2016). Relative to the identity preview, readers spent more time and skipped the first constituent less often when they were presented with unrelated character or psedocharacter previews (all $t > 4.44$, $|z| > 2.84$). These effects replicate standard preview effects (Liversedge & Findlay, 2000; Rayner, 2009). Note, there were no differences between the latter two previews, indicating that both incorrect previews provided a comparable lack of preview benefit for the first constituent. This result suggests that both unrelated character and pseudocharacter previews provide appropriate baseline conditions in Chinese boundary paradigm experiments involving the manipulation of preview benefit.

Very importantly in relation to the MCU Hypothesis, preview interacted with phrase type across the fixation time measures (all $|t| > 2.17$, though the interactions were marginal in FFD and TFD, $|t| > 1.75$, $p < .08$). The planned contrasts showed that the preview effect was more robust for idioms (FFD: b = 0.08, SE = 0.01, t = 6.19; SFD: b = 0.08, SE = 0.01, t = 5.73; GD: b =0.11, SE = 0.02, t = 7.05; TFD: b = 0.12, SE = 0.02, t = 6.02; Go-past: b = 0.10, SE = 0.02. t = 4.56) than phrases (FFD: b = 0.05, SE = 0.01, t = 3.97; SFD: b = 0.04, SE = 0.01, t = 2.98; GD: b =0.06, SE = 0.02, t = 3.68; TFD: b = 0.08, SE = 0.02, t = 3.82; Go-past: b = 0.03, SE = 0.02. t = 1.27, see Figure 2). Clearly, when the two constituents formed an idiom compared with a phrase, readers preprocessed the second constituent in the parafovea to a greater extent. In line with findings of English spaced compounds (Cutter et al., 2014), presumably, the first constituent of the idiom licensed processing of the entire MCU, leading to early and pronounced preprocessing the second constituent prior to fixation. Note, the preview effect from the second constituent of phrases was also reliable, though the effect was quite reduced. This effect is likely due to the fact that there are no visual word boundaries in Chinese text, and phrases with a two-character modifier and a one-character noun structure are also often segmented off-line and categorised as long single words (Liu et al., 2013). Overall, these results provide evidence that the idioms with a modifier-noun structure were lexicalized and processed parafoveally as a single lexical representation during Chinese reading.

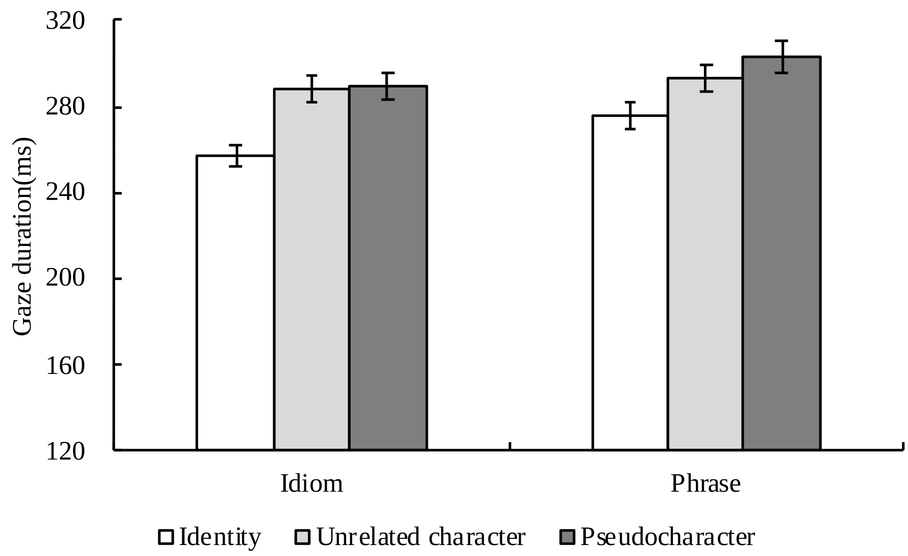

Figure 2. Preview effects for the first constituents of idioms and phrases for GD in Experiment 1 (Error bars represent standard errors of the mean).

The Second Constituent (*n* + 2)

There were reliable effects of phrase type and preview of an unrelated character relative to an identity preview in all eye movement measures with shorter fixations and increased skipping for idioms than phrases (all $|t|$ or $|z| > 5.77$), and for identity than unrelated character previews (all $|t|$ or $|z| > 2.77$). Again, these results suggest that constituents are easier to process when they form idioms compared with phrases, once more demonstrating a processing advantage for Chinese idioms relative to matched phrases during reading. Furthermore, as to be expected, standard preview effects were obtained with constituents being easier to process when readers had identical previews compared to unrelated character or pseudocharacter previews. The interactions between preview and phrase type were not reliable in the second constituent region, presumably because the identity of the first and second constituents are revealed immediately after the boundary is crossed. Preview effects are immediate and dissipate rapidly. More fixations after the eyes cross the boundary land on the first constituent and carry most of the preview effect, meaning that effects on the second constituent are weaker than those on the first. We note that this pattern of effects appears to be consistent with the two constituents being treated as a single unit.

The Whole Target Region (*n* + 1 and *n* + 2)

The whole target region includes both the first and the second constituents. Again, the effect of phrase type was reliable in FFD, GD, TFD and Go-past time such that readers spent less time fixating idioms than phrases (all $|t| > 3.14$). There was also a reliable effect of preview across all fixation time measures such that readers fixated the target string for less time when they received an identical preview compared with an unrelated character preview (all $|t| > 6.61$). Furthermore, similar to the effects observed for the first constituent, there were interactive effects of preview and phrase type that were significant in the FFD, SFD and Go-past time measures (all $|t| > 2.49$), and marginal in GD ($t = -1.90$, $p = .058$). Again, the planned contrasts showed more robust preview effects for idioms (FFD: $b = 0.09$, $SE = 0.01$, $t = 7.55$; SFD; $b = 0.11$, $SE = 0.02$, $t = 6.60$; GD: $b = 0.19$, $SE = 0.02$, $t = 9.78$; Go-past; $b = 0.20$, $SE = 0.02$, $t = 9.44$) than phrases (FFD: $b = 0.05$, $SE = 0.01$, $t = 4.15$; SFD: $b = 0.05$, $SE = 0.02$, $t = 2.75$; GD; $b = 0.14$, $SE = 0.02$, $t =6.54$; Go-past: $b = 0.12$, $SE = 0.02$, $t = 4.82$). These results replicate the findings from the first constituent analyses indicating that readers preprocess the constituents of the idioms to a greater extent than those of the phrases, consistent with the suggestion that idioms are processed as MCUs and accessed via a single lexical entry during Chinese reading.

In summary, the results of Experiment 1 were straightforward. Consistent with Yu et al. (2016), a processing advantage for idioms was obtained foveally with shorter fixations compared with phrases. More importantly, a larger preview effect from the second constituent occurred for idioms than phrases when fixations were made on the first constituent (as well as the whole target region) with shorter fixations for identical than unrelated and pseudocharacter previews. This suggests that idioms are parafoveally processed to a greater extent than phrases, a result consistent with the suggestion that idioms are parafoveally (and foveally) processed as MCUs [2].

In order to extend the findings from Experiment 1, we undertook a second experiment in which the preview of each constituent of the idioms and phrases used in Experiment 1, was orthogonally manipulated as per Cutter et al. (2014). Specifically, before the eyes crossed the boundary to the right of the pretarget word *n*, participants received (a) identical previews for both constituents n+1 and n+2, (b) an identical

---

[2] If the preview benefit effects in Experiment 1 were caused by the predictability of the second constituent given the first, then the magnitude of the preview benefit effect should be positively correlated with the degree of such predictability. To test this for all the items (idioms and matched phrases) we correlated the magnitude of the preview benefit effect with the predictability of the second constituent based on the first. Our results showed no significant correlations (all $r < 0.15$, $p > .05$) across all measures in the analysis of all regions, except for GD in the first constituent analysis ($r = 0.16$, $p = .04$) and FFD in the whole target string analysis ($r = 0.17$, $p = .03$). Given that the predictability values for the matched phrases were substantially reduced relative to those for the idioms, it is possible that the inclusion of the matched phrase data in these correlations artificially deflated the magnitude of effects. To ensure that this was not the case, we undertook a further set of analyses based solely on the idiom stimuli in which we again correlated preview benefit effects with the predictability of the second constituent given the first. As with the first set of analyses, we obtained no reliable correlations (all $r < 0.18$, all $p > .10$). Both sets of these analyses are completely consistent with our claim that the effects we report are due to the MCU status of our stimuli rather than predictability relations that may exist between the MCU constituents. We consider that these results along with Experiment 1 LMMs that included predictability as a covariate demonstrate convincingly that the MCU status of the target string was determinant in producing the effects we observed.

preview for *n*+1 and a pseudocharacter preview for *n*+2, (c) a pseudocharacter preview for *n*+1 and an identical preview for *n*+2, or (d) pseudocharacter previews for both *n*+1 and *n*+2. We predict a standard *n*+1 preview effect and a possible parafoveal-on-foveal effect when the n+1 preview was a pseudocharacter compared with an identity. However, importantly, the comparisons between conditions (a) and (b), and between conditions (c) and (d) across idioms and matched phrases allow us to evaluate the extent to which *n*+2 preview effects occurred in each type of target string when *n*+1 was, or was not, available. If idioms are processed as MCUs, being identified via a single lexical unit parafoveally and foveally, a *n*+2 preview effect would be expected only when the first constituent n+1 is available and such an effect should be more pronounced for idioms than phrases. These predictions follow directly from the MCU Hypothesis, and would directly align with the findings of Cutter et al., and be consistent with the results of Experiment 1.

Experiment 2

Method

Participants

One hundred and thirty-six students at Tianjin Normal University (121 females, mean age 20 years) who did not take part in Experiment 1 were recruited to participate in Experiment 2. They were all native Chinese speakers with normal or corrected-to-normal vision, and naïve to the purpose of the experiment.

Apparatus

The same apparatus was used as in Experiment 1.

Materials and design

Using the boundary paradigm the previews of the first (n + 1) and the second constituent (n + 2) were orthogonally manipulated to be either an identity or a pseudocharacter. Hence, in Experiment 2, we employed a 2 (Phrase Type: idiom or phrase) × 2 (n + 1 preview: identity or pseudocharacter) × 2 (n + 2 preview: identity or pseudocharacter) within participant design. Eighty sets of sentences with 80 pairs of target strings from Experiment 1 were used in Experiment 2. Note that 4 sets of sentences with 4 pairs of target strings from Experiment 1 were removed to ensure the same number of items for each condition in Experiment 2. Among the stimuli selected for the current experiment, the stroke complexity of two constituents, word frequency and character frequency of the second constituents were matched carefully for idioms and phrases. Control details are shown in Table 4, and an example of the experimental sentences is shown in Figure 3.

Table 4 Statistical properties for the idioms and phrases in Experiment 2

Preview	Property	Idioms	Phrases
Identity	Stroke number of n+1	14.6 (4.0)	14.6 (4.0)
	Stroke number of n+2	8.7 (3.3)	8.7 (2.8)
	Word frequency of n+2	61.4 (80.8)	61.6 (75.4)
	Character frequency n+2	296.9 (710.6)	282.2 (629.5)
Pseudocharacter	Stroke number of n+1	14.6 (4.0)	14.6 (4.0)
	Stroke number of n+2	8.7 (2.6)	8.7 (2.6)

Standard deviations are provided in parentheses. The unit of frequency is per million.

<table>
<tr><th>Phrase type</th><th>n + 1 preview</th><th>n + 2 preview</th><th>Sentence</th></tr>
<tr><td rowspan="4">Idiom</td><td rowspan="2">Identity</td><td>Identity</td><td>小辉充分利用各种|垫脚石当上了区域经理。</td></tr>
<tr><td>Pseudocharacter</td><td>小辉充分利用各种|垫脚汇当上了区域经理。</td></tr>
<tr><td rowspan="2">Pseudocharacter</td><td>Identity</td><td>小辉充分利用各种|吴崖石当上了区域经理。</td></tr>
<tr><td>Pseudocharacter</td><td>小辉充分利用各种|吴崖汇当上了区域经理。</td></tr>
<tr><td rowspan="4">Phrase</td><td rowspan="2">Identity</td><td>Identity</td><td>小辉充分利用各种|垫脚布让家里保持干净。</td></tr>
<tr><td>Pseudocharacter</td><td>小辉充分利用各种|垫脚汇让家里保持干净。</td></tr>
<tr><td rowspan="2">Pseudocharacter</td><td>Identity</td><td>小辉充分利用各种|吴崖布让家里保持干净。</td></tr>
<tr><td>Pseudocharacter</td><td>小辉充分利用各种|吴崖汇让家里保持干净。</td></tr>
</table>

Figure 3 An example of the Chinese sentences used in Experiment 2. The preview of the first and the second constituent was manipulated. The vertical line represents the position of the invisible boundary. When the eyes crossed the boundary, the preview was replaced by the target character (The target strings are in bold, but were presented normally in the experiment).

The mean sentence naturalness score in Experiment 2 was 3.9 (SD = 0.4), with no difference between idioms and phrases ($F < 1$). Both idioms and phrases were unpredictable from the preceding sentential context (M = 0%, SD = 2%). Again, as in Experiment 1, the second constituents of idioms given the first (M = 69%, SD = 31%) were more predictable than those of phrases (M = 0.4%, SD = 1.7%), the first constituents of idioms given the second (M = 10%, SD = 18%) were more predictable than those of phrases (M = 0.9%, SD = 0.5%), and the transitional probability of the two constituents for idioms (M = 28%, SD = 30%) was higher than that for phrases (M = 0.7%, SD = 2.4%)[3]. We constructed eight files with each file containing 80

[3] As in Experiment 1, we undertook a series of analyses in which we included the predictability of the second constituent given the first, the predictability of the first constituent given the second, or the transitional probability of the two constituents as covariates in the LMMs. In these analyses, for all measures across all regions, the pattern of results was identical to the results we report here. Once again, these analyses demonstrate clearly that the covariates did not cause our effects. Also, similar to

experimental sentences (10 sentences in each condition). Conditions across these files were rotated according to a Latin Square. Each participant read these sentences (mixed with 40 filler sentences without display change) presented randomly from one of these files. Eight practice sentences were presented prior to the experimental sentences, and on one third of the trials participants were shown a yes/no comprehension question.

Procedure

The procedure was identical to Experiment 1.

Results and Discussion

The mean comprehension accuracy for all participants was 96% indicating that they could understand sentences well. With respect to display change awareness, 10 participants were unaware of the display change and the remainder were aware of the display change but unable to report which characters had changed. We used the same data exclusion criteria as in Experiment 1. Fixations shorter than 80ms or longer than 1200ms were excluded. Trials were removed due to 1) a track loss occurring or fewer than three fixations being made (0.2% of the data), 2) participants blinking during display changes or during a fixation on the target region (3% of the data), 3) the display change triggered early or late (16% of the data). Finally, we also removed any

---

Experiment1, we calculated variance inflation factors (VIF) for phrase type and predictability of the second constituent given the first, or predictability of the first constituent given the second. All the VIFs across all measures in all regions were less than 4.76, though a little higher for TFDs in the first constituent analysis (5.25) and the whole region analysis (5.30). For phrase type and predictability of the first constituent given the second, all the VIFs across all measures in all regions were less than 1.30, indicating a moderate or no correlation between phrase type and two types of predictability. Again, these results indicate that it is appropriate to capture variance associated with predictability and transitional probability by including them in our LMMs as covariates.

observations for each measure that were above or below three standard deviations from each participant's mean prior to conducting the analyses prior to conducting the analyses (1.1%). Analyses were carried out for the same eye movement measures on the same target regions as in Experiment 1, the pre-target word (*n*), the first constituent (*n*+1) and the second constituent (*n*+2). The means and standard deviations across all the regions are shown in Table 5.

Linear mixed-effects models (LMM) using the lme4 package (Version 1.1-21, Bates) in R (Version 3.6.0) were conducted to analyze the data. In order to allow for quite direct comparison of our results with those reported by Cutter et al. (2014), each preview of *n*+1 and *n*+2, as well as their interaction was treated as a fixed factor for idioms and phrases, respectively. Two contrasts were set up to examine the n+2 preview effect at each level of the n+1 preview (i.e., for the n+1 identity preview and for when the preview was a pseudocharacter). Participants and items were treated as crossed random factors. The random effects structure of the model and the trimming procedure was the same as in Experiment 1. Fixed effect estimations for the eye movement measures across all regions are shown in Table 6.

Table 5 Eye movement measures for all regions across the eight experimental conditions in Experiment 2.

Analysis Region	Phrase	*n* + 1 Preview	*n* + 2 Preview	FFD	SFD	GD	TFD	Go-past	SP
The Pretarget Word (*n*)	Idiom	Identity	Identity	220(48)	220(48)	223(49)	264(65)	280(136)	.61(.18)
			Pseudocharacter	219(47)	219(47)	224(54)	271(76)	272(116)	.61(.17)
		Pseudocharacter	Identity	222(47)	222(47)	225(49)	282(76)	281(122)	.59(.19)
			Pseudocharacter	221(51)	221(52)	227(57)	285(82)	281(125)	.60(.21)
	Phrase	Identity	Identity	220(46)	220(46)	222(47)	276(87)	277(125)	.60(.19)
			Pseudocharacter	220(43)	220(43)	221(43)	281(80)	264(99)	.60(.18)
		Pseudocharacter	Identity	221(46)	220(45)	222(47)	306(98)	268(101)	.60(.17)
			Pseudocharacter	224(55)	225(62)	227(59)	291(81)	267(110)	.60(.19)
The 1st Constituent (*n+1*)	Idiom	Identity	Identity	230(40)	229(40)	247(51)	340(118)	312(101)	.25(.21)
			Pseudocharacter	251(48)	251(51)	277(63)	371(122)	335(123)	.22(.22)
		Pseudocharacter	Identity	290(56)	307(73)	361(91)	465(171)	468(176)	.13(.17)
			Pseudocharacter	297(57)	311(74)	367(100)	471(163)	453(142)	.13(.18)
	Phrase	Identity	Identity	237(46)	236(47)	264(63)	431(163)	336(124)	.23(.20)
			Pseudocharacter	249(51)	251(54)	281(72)	456(169)	337(135)	.21(.20)
		Pseudocharacter	Identity	291(56)	304(78)	369(107)	545(188)	492(191)	.15(.18)
			Pseudocharacter	288(58)	302(80)	361(93)	535(168)	474(176)	.15(.17)
The 2nd Constituent (*n+2*)	Idiom	Identity	Identity	231(72)	231(72)	232(75)	265(81)	284(144)	.59(.19)
			Pseudocharacter	247(62)	247(65)	256(71)	295(90)	327(131)	.52(.19)
		Pseudocharacter	Identity	231(63)	231(64)	233(63)	285(86)	343(273)	.55(.20)
			Pseudocharacter	241(60)	243(62)	246(64)	290(100)	330(196)	.57(.21)
	Phrase	Identity	Identity	259(60)	258(61)	270(71)	334(99)	370(156)	.54(.21)
			Pseudocharacter	268(66)	268(70)	283(71)	370(106)	422(179)	.49(.22)
		Pseudocharacter	Identity	255(56)	254(55)	265(60)	352(98)	394(163)	.47(.21)
			Pseudocharacter	254(58)	252(59)	267(66)	350(107)	418(207)	.50(.21)
The Whole Region	Idiom	Identity	Identity	230(37)	229(41)	300(81)	452(168)	384(136)	.08(.13)
			Pseudocharacter	253(47)	260(67)	372(112)	527(200)	456(161)	.08(.13)
		Pseudocharacter	Identity	286(54)	309(99)	436(117)	605(214)	577(195)	.04(.08)
			Pseudocharacter	294(57)	320(91)	448(129)	623(236)	577(205)	.05(.10)
	Phrase	Identity	Identity	239(40)	237(47)	366(137)	625(250)	467(182)	.07(.12)
			Pseudocharacter	251(49)	255(66)	419(143)	701(266)	519(193)	.07(.13)

Pseudocharacter	Identity	288(51)	301(85)	477(142)	782(282)	647(228)	.04(.10)
	Pseudocharacter	286(55)	296(87)	469(134)	758(249)	659(245)	.04(.10)

Note. Standard deviations are provided in parentheses. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration; Go-past = go-past time; SP = skipping probability.

Table 6 LMM analyses across all regions in Experiment 2.

		FFD			SFD			GD			TFD			Go-past			SP		
Region	Effect	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
										Idiom									
The Pretarget Word (*n*)	*n*+1 preview	0.01	0.01	0.77	0.01	0.01	0.44	0.01	0.01	0.86	0.05	0.02	**3.22**	0.01	0.02	0.60	-0.04	0.06	-0.77
	n+2 preview	0.00	0.01	0.08	0.00	0.01	0.04	0.01	0.01	0.52	0.02	0.02	1.10	0.01	0.02	0.53	0.02	0.06	0.28
	n+1 × *n*+2	0.01	0.03	0.23	0.01	0.02	0.26	0.01	0.03	0.49	-0.00	0.03	-0.00	0.04	0.04	1.20	0.07	0.11	0.63
										Phrase									
	n+1 preview	0.01	0.01	0.66	0.00	0.01	0.08	0.01	0.01	0.46	0.06	0.02	**3.59**	0.00	0.02	0.23	-0.00	0.06	-0.06
	n+2 preview	0.00	0.01	0.34	0.01	0.01	0.52	0.01	0.01	0.71	-0.02	0.02	-1.01	-0.02	0.02	-0.91	-0.00	0.06	-0.02
	n+1 × *n*+2	0.03	0.02	1.03	0.02	0.02	1.00	0.03	0.02	1.23	-0.05	0.03	-1.63	0.02	0.04	0.63	0.05	0.12	0.41
										Idiom									
The 1st Constituent (*n+1*)	*n*+1 preview	0.19	0.01	**15.35**	0.24	0.02	**15.80**	0.32	0.02	**19.94**	0.29	0.02	**17.20**	0.36	0.02	**17.93**	-0.84	0.09	**-9.63**
	n+2 preview	0.05	0.01	**5.00**	0.05	0.01	**4.00**	0.06	0.01	**4.78**	0.05	0.01	**3.71**	0.03	0.01	1.75	-0.20	0.11	-1.82
	n+1 × *n*+2	-0.07	0.02	**-3.20**	-0.09	0.02	**-3.75**	-0.11	0.02	**-4.86**	-0.10	0.03	**-3.44**	-.12	0.03	**-3.87**	0.17	0.17	1.00
										Phrase									
	n+1 preview	0.17	0.01	**15.54**	0.19	0.02	**12.06**	0.28	0.02	**17.25**	0.23	0.02	**14.59**	0.38	0.02	**15.48**	-0.56	0.08	**-6.77**
	n+2 preview	0.02	0.01	**1.99**	0.03	0.01	**2.16**	0.02	0.01	1.63	0.03	0.02	1.79	0.01	0.02	0.41	-0.04	0.08	-0.48
	n+1 × *n*+2	-0.04	0.02	**-1.97**	-0.04	0.02	-1.83	-0.06	0.03	**-2.34**	-0.07	0.03	**-2.19**	-0.05	0.04	-1.49	0.15	0.17	0.88
										Idiom									
The 2nd Constituent (*n+2*)	*n*+1 preview	-0.01	0.01	-0.62	-0.00	0.01	-0.06	-0.01	0.02	-0.62	0.02	0.02	0.91	0.03	0.02	1.46	-0.01	0.06	-0.13
	n+2 preview	0.05	0.02	**3.16**	0.06	0.02	**3.42**	0.07	0.02	**4.03**	0.06	0.02	**2.81**	0.07	0.02	**3.34**	-0.10	0.07	-1.51
	n+1 × *n*+2	-0.01	0.03	-0.26	-0.00	0.03	-0.11	-0.02	0.03	-0.76	-0.05	0.04	-1.55	-0.11	0.04	**-2.43**	0.43	0.13	**3.39**
										Phrase									
	n+1 preview	-0.03	0.01	**-2.29**	-0.04	0.01	**-2.44**	-0.04	0.02	**-2.53**	-0.01	0.02	-0.32	-0.02	0.02	-0.65	-0.13	0.06	**-2.16**

	n+2 preview	0.01	0.02	0.97	0.02	0.02	1.09	0.03	0.02	1.60	0.03	0.02	1.81	0.08	0.02	**3.54**	0.09	0.06	-1.50
	n+1 × *n*+2	-0.04	0.03	-1.29	-0.04	0.03	-1.30	-0.04	0.03	-1.18	-0.13	0.04	**-3.51**	-0.06	0.05	-1.25	0.29	0.12	**2.33**
										Idiom									
	n+1 preview	0.18	0.01	**14.71**	0.22	0.02	**12.32**	0.29	0.01	**20.23**	0.26	0.02	**16.13**	0.35	0.02	**20.95**	-0.72	0.14	**-5.20**
	n+2 preview	0.06	0.01	**5.77**	0.07	0.01	**4.63**	0.11	0.01	**7.53**	0.09	0.01	**6.47**	0.09	0.01	**5.90**	0.13	0.14	0.95
The Whole Region	*n*+1 × *n*+2	-0.06	0.02	**-3.27**	-0.07	0.03	**-2.72**	-0.19	0.03	**-7.23**	-0.17	0.03	**-5.94**	-0.20	0.03	**-6.88**	0.24	0.27	0.89
										Phrase									
	n+1 preview	0.15	0.01	**12.00**	0.17	0.02	**9.31**	0.22	0.02	**14.55**	0.19	0.02	**11.30**	0.33	0.02	**16.74**	-0.62	0.15	**-4.25**
	n+2 preview	0.02	0.01	**2.07**	0.03	0.01	1.77	0.06	0.02	**4.18**	0.07	0.01	**4.57**	0.07	0.02	**4.20**	-0.03	0.15	-0.18
	n+1 × *n*+2	-0.04	0.02	**-2.16**	-0.05	0.03	-1.76	-0.15	0.03	**-4.82**	-0.17	0.03	**-5.53**	-0.11	0.03	**-3.31**	0.19	0.29	0.67

Note. Significant terms are marked in bold, and marginal terms are underlined. b = regression coefficient.

The Pretarget Word (*n*)

The *n*+1 preview effect was reliable in TFD for both idioms and phrases such that readers spent longer processing the pretarget word *n* when the *n*+1 preview was a pseudocharacter compared with an identity. This pattern of effects suggests that there was an influence of the n+1 preview which was only available prior to the eyes fixating *n*+1. Note, though, that this effect was late, occurring in total time, but not in first pass measures. Clearly, there was some sensitivity to the *n*+1 preview prior to its fixation, but this effect reflects an orthographic parafoveal-on-foveal effect, and that influence was relatively late in processing (being carried cumulatively by both first and second pass fixations). There were no other reliable effects in this region.

The First Constituent (*n* + 1)

For the first constituent analyses, there was a significant effect of *n*+1 preview in all eye movement measures for both idioms and phrases, such that readers spent longer fixating *n*+1, and skipped it less often when they received a pseudocharacter preview compared with an identity preview (all t or $z > 6.76$), which replicated preview effects reported widely in the previous literature (see Rayner, 2009). The *n*+2 preview effect was also reliable at *n*+1 for idioms in FFD, SFD, GD and TFD, and for phrases in FFD and SFD, with longer fixations when the *n*+2 preview was a pseudocharacter rather than an identity (all $t > 1.98$). Importantly, in relation to the MCU Hypothesis, there was a reliable interaction between *n*+1 and *n*+2 previews for idioms across all fixation time measures (all $|t| > 3.19$). The planned contrasts showed that this was due to a reliable *n*+2 preview effect when *n*+1 was an identity preview (all $t > 3.33$) rather than a pseudocharacter (all $t < 1.39$). This result replicates Cutter et al.'s (2014) findings perfectly, and we similarly conclude that the presence of *n*+1 in the parafovea licenses preprocessing of *n*+2 to a significant degree and thus results in reliable n+2 preview effects. For phrases, the interaction between the two previews was reliable, but much less robust in FFD, GD and TFD (all $|t| > 1.96$, see Figure 4). The planned contrasts for phrases showed a similar pattern, though as we predicted, the effects were smaller for phrases (FFD: 12ms; GD: 17ms; TFD: 25ms) than for idioms (FFD: 21ms; GD: 30ms; TFD: 31ms; *n*+2 preview effect when *n*+1 was an identity: all $t > 2.60$; *n*+2 preview effect when n+1 was a pseudocharacter: all $|t| < 0.55$) [4]. These results are consistent with those from Experiment 1. As suggested earlier, presumably the preview effects for the phrases here arise due to the unspaced and relatively dense nature of Chinese orthography, as well as the constraint imposed by the two-character modifier over the subsequent one-character noun. Regardless, the results from Experiment 2 align almost perfectly with the findings from Experiment 1, and indicate that the second constituent of our target strings was processed prior to the fixation (and note that this occurred when it was positioned three characters away from fixation) to a greater extent when the first constituent alongside the second formed an idiom compared with when it formed a matched phrase. In our view, these data provide further strong support for the MCU hypothesis.

---

[4] We also directly tested the *n*+1 × *n*+2 preview interaction with phrase type. The three-way interaction was only reliable on Go-past time in the whole target region analysis ($b = 0.10$, $SE = 0.04$, $t = 2.18$). Further analysis showed that the *n*+1 × *n*+2 preview interaction was larger for idioms ($b = -0.20$, $SE = 0.03$, $t = -6.03$) than for phrases ($b = -0.11$, $SE = 0.04$, $t = -2.92$). The other three-way interactions were not reliable (all $|t|$ or $|z| < 1.39$, $p > 0.16$).

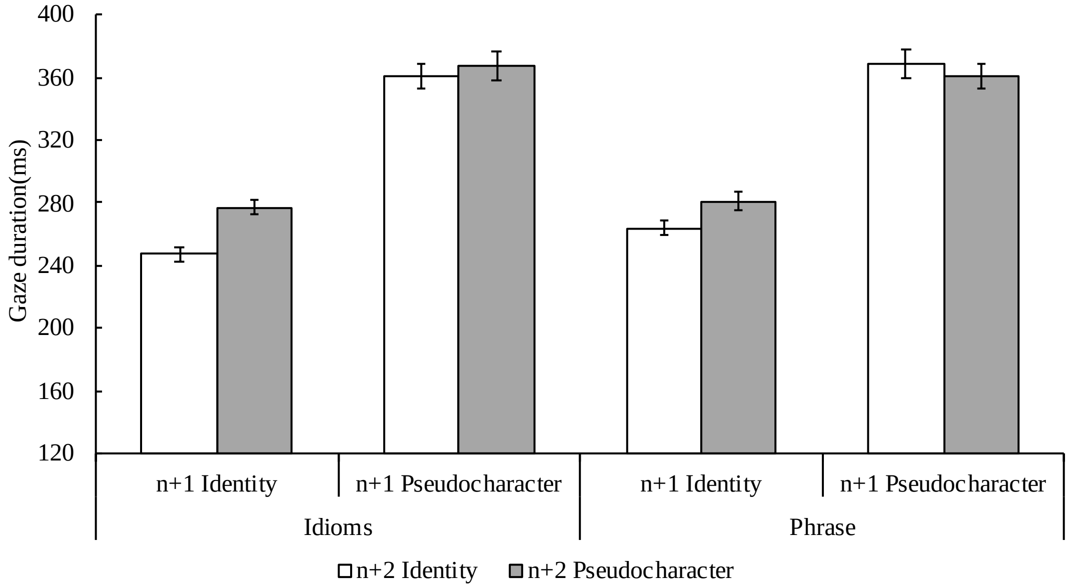

Figure 4. Preview effects for the first constituents of idioms and phrases for GD in Experiment 2 (Error bars represent standard errors of the mean).

The Second Constituent (*n* + 2)

For the second constituent analyses, the *n*+2 preview effect was reliable for idioms in all fixation time measures, with longer fixations for pseudocharacter previews than identical previews (all $t > 2.80$). There was also a reliable interaction between the two previews in skipping probability ($z = 3.39$) and go-past times ($|t| = 2.43$). The planned contrasts showed a reliable *n*+2 preview effect with lower skipping rate and longer go-past times for pseudocharacter than identical previews and this effect was only reliable when *n*+1 was an identity ($|t|$ or $|z| > 3.55$) rather than a pseudocharacter ($|t|$ or $|z| < 1.26$), which is consistent with the findings from the first constituent analyses.

In contrast, for phrases, the *n*+1 preview effect was reliable in FFD, SFD, GD and skipping probability with longer fixations and more skips when *n*+1 was an identity compared with a pseudocharacter. The reversed pattern of the *n*+1 preview effect occurred on the second constituent in first pass reading time measures. This effect likely occurred because when the target was an idiom, the first constituent *n*+1 licensed immediate processing of the second constituent resulting in it being processed as a single unit (i.e., a lexicalized idiom). However, when the second constituent of a phrase was not licensed to the same degree, a longer period is required to process it during first pass reading. There was also an interaction between the two previews in skipping probability (z = 2.33) and TFD ($|t|$ = 3.51). The planned contrasts showed a reliable *n*+2 preview effect with lower skipping rate and longer total fixation duration for pseudocharacter than identical previews and this effect was only reliable when *n*+1 was an identity ($|t|$ or $|z| > 2.71$) rather than a pseudocharacter preview ($|t|$ or $|z| < 1.20$).

The Whole Target Region (*n* + 1 and *n* + 2)

As with Experiment 1, the whole target region includes both the first and the second constituents. For the whole target region, the pattern of results was the same as that for the first constituent analyses, and effects became even stronger as we combined the data from the two individual constituents together. Specifically, there was a significant effect of *n*+1 preview in all eye movement measures for both idioms and phrases, with longer fixations and lower skipping rate for pseudocharacter previews than identity previews (all t or $z > 4.24$). The *n*+2 preview effect was also reliable for idioms and phrases in all fixation time measures (all $t > 2.06$, though marginal for phrases in SFD, $t = 1.77$, $p = .08$), with longer fixations when the *n*+2 preview was a pseudocharacter rather than an identity. In line with findings from the first constituent analyses, there was a reliable interaction between *n*+1 and *n*+2 previews for idioms and phrases across fixation time measures (all $|t| > 2.15$, though marginal for phrases in SFD,

$t = 1.76$, $p = .08$). The n+2 preview effect was only reliable when the first constituent was parafoveally available (all $t > 2.93$) but marginal or non-significant when it was unavailable (all $t < 1.78$). Finally, all of these effects were more pronounced for idioms than phrases. In summary, these results were extremely consistent with those observed in Experiment 1 and offer strong support for the MCU Hypothesis [5].

General Discussion

Using the boundary paradigm, the present study set out to investigate whether Chinese idioms with a modifier-noun structure are processed as MCUs, and whether idioms of this type show a processing advantage compared with matched phrases, both parafoveally as well as foveally, during Chinese reading. In both the three-character idioms and the matched phrases, the first constituent (n+1) was a two-character modifier and the second constituent (n+2) was a one-character noun. In Experiment 1, the first constituent of each idiom and matched phrase set was identical, and the preview of the second constituent was manipulated to be an identity, an unrelated character or a pseudocharacter with the invisible boundary located before the target string. The results showed that idioms were fixated for less time than phrases – a processing advantage of

[5] As in Experiment 1, to test whether predictability may have contributed to the magnitude of the preview effects in Experiment 2, we correlated the magnitude of the preview benefit effect with the predictability of the second constituent based on the first. Our results showed no significant correlations (all $r < 0.15$, $p > .05$) across all measures in the analysis of all regions. Similar to Experiment 1, we also undertook analyses based solely on the idiom stimuli to preclude the possibility that any correlations might be diminished by the low predictability values obtained for the matched phrase stimuli. Again, these analyses produced no robust effects (all $r < 0.20$, all $p > .05$). These analyses, along with the counterpart analyses from Experiment 1, and the LMM analyses including predictability and transitional probability as a covariate, all provide strong evidence in support of the MCU hypothesis and little support for accounts based on predictability or transitional probability.

idioms over phrases foveally (see Yu et al., 2016). Importantly, a larger parafoveal preview effect occurred for idioms on the first constituent (as well as on the whole target region) with shorter fixations for an identical than for an unrelated or pseudocharacter preview, suggesting that idioms are parafoveally processed in their entirety to a greater extent than phrases. To reinforce the effects from Experiment 1, In Experiment 2, the preview of each constituent (i.e., *n*+1 and *n*+2) of idioms and matched phrases were orthogonally manipulated to be an identity or a pseudocharacter as per Cutter et al., (2014). The results showed that readers obtained the *n*+2 preview effect only when the *n*+1 preview was an identity, and this effect was more pronounced for idioms than for phrases, again, suggesting that idioms in the present study are parafoveally processed as MCUs.

The findings with respect to a foveal processing advantage with idioms over phrases in Chinese replicated Yu et al. (2016), in which they examined preview effects using idioms with a verb and noun structure, and observed that idioms were foveally processed more quickly than matched phrases. Yu et al.'s results suggested that idioms are stored and accessed as individual lexical units, a suggestion that is in line with previous literature regarding the lexical representation of idioms (e.g., Carrol & Conklin, 2017; Conklin & Schmitt, 2008, 2012; Siyanova-Chanturia, Conklin & Schmitt, 2011; Swinney & Cutler, 1979). In one important respect, however, the pattern of results in the present study is inconsistent with the findings of Yu et al. (2016). In relation to parafoveal processing of idioms, Yu et al. did not find any evidence to suggest preprocessing of word *n*+2 in idioms occurred to a greater extent than it did in matched phrases in the parafovea. Yu et al. argued that the idioms in their study were only processed foveally as a single unit, and that lexical representations of those idioms might not have become sufficiently activated prior to readers crossing the invisible boundary, to produce preview benefit effects suggesting the idiom was processed parafoveally as a MCU. Recall, though, that as we noted earlier, the idioms in the Yu et al. study were comprised of an initial single-character verb followed by a two-character noun. It is likely that in such idioms, the verb only very loosely constrains the identity of the noun. In contrast, in the present study (as evidenced by our pre-screen completion data) the initial two-character modifier tightly constrains the potential identify of the upcoming single-character noun. In such a situation, when the reader encounters the first constituent, the modifier, of an idiom in the parafovea, it is more likely that readers rapidly activate the lexical representation corresponding to the entire unit (the lexicalized idiom), including the successive constituent noun, before it is directly fixated (for a similar argument, but one that is divorced from issues of parafoveal and foveal processing in relation to fixations made during natural reading, see the Lexical Representation Hypothesis, Swinney & Cutler, 1979). However, it remains an empirical argument as to whether the precise structure of Chinese idioms exerts a modulatory influence on how their constituents are processed parafoveally and foveally, and clearly, further evidence is required to elucidate this issue.

The present results extend and support the findings of Zang et al. (2021) who showed highly frequently occurring Chinese two-character phrases operate as MCUs, being lexicalized and processed in the parafovea as a single unit during reading. In the present study, despite the preview (of the third character of the target string) being one character further from the point of fixation, we still found that it was processed to a greater extent when it formed part of a larger lexical unit compared with when it was the second word of a two word string. These results demonstrate that the determinant of the emergence of this effect is linguistic rather than perceptual in nature. In this regard, our findings are also entirely consistent with those of Cutter et al. (2014) who showed that English spaced compounds (e.g., *teddy bear*) operate as MCUs parafoveally. In Cutter et al.'s experiment, they examined preview effects for each constituent of a spaced compound with the previews for each constituent being either the identity or a nonword. They found a reliable preview effect for word *n*+2 only when word *n*+1 was an identity, and this occurred even though *n*+1 was quite long (on average 5.65 characters) and comparatively low in frequency, that is, under sub-optimal conditions for such an effect.

The *n*+2 preview effect has theoretical implications for current models of eye movement control during reading. As reviewed in the Introduction, this effect is generally not predicted by models in which words are lexically identified serially such as E-Z Reader unless *n*+1 is short and/or of high frequency. Short and high frequency *n*+1 words can be recognized very rapidly and this allows for processing of word *n*+2 such that its influences might be apparent (e.g., Reichle, 2011; though see also Schotter, Reichle, & Rayner, 2014 for computational modeling simulations of *n*+2 parafoveal influences). In contrast, the *n*+2 effect is predicted by parallel processing models such as SWIFT, on the condition that word n+2 falls within the perceptual span (Engbert &

Kliegl, 2011; see also Risse, Hohenstein, Kliegl, & Engbert, 2014). However, empirically, the prior research has shown only very subtle, limited, preview effects of word *n*+2 in reading both in alphabetic languages such as English and non-alphabetic languages such as Chinese. Specifically, the *n*+2 preview effect has only been reported when word *n*+1 was high frequency, and/or very short word (e.g., no more than three letters long) but not when it was longer (see Vasilev & Angele, 2017 for a review). For Chinese, a *n*+2 preview effect has only been documented when word *n*+1 was a function word or a very high frequency single-character word, but not when it was a low frequency word (Yan, Kliegl, Shu, Pan, & Zhou, 2010; Yang, Wang, Xu, & Rayner, 2009; Yang, Rayner, Li, & Wang, 2012).

In a Bayesian meta-analysis based on 11 studies investigating *n*+2 preview effect, Vasilev and Angele (2017) found when both alphabetic and Chinese studies were analyzed, this effect was mainly constrained to the first fixation duration measure, being associated with a 5ms effect size, but there was a high probability (87% on word *n*+2, and 85% on word *n*+1) that this effect was bigger than 1ms. When only the alphabetic studies were analyzed, the estimated effect was smaller, however, there remained a high probability (70% on word *n*+2, and 85% on word *n*+1) that this effect was bigger than 1ms. Vasilev and Angele concluded that the *n*+2 effect was small but did exist, with bigger such effects for Chinese than for alphabetic studies due to the densely packed nature of Chinese written text. They also noted that the length of word *n*+1 determined the extent to which word *n*+2 was preprocessed with larger effects when word *n*+1 was short (though the influence of this variable could not be examined in detail due to an insufficient number of studies).

The present results may pose difficulties both for models in which words are lexically identified serially, and those in which words are lexically identified in parallel. First, the *n*+1 constituents in the present study were longer and of lower frequency than would ordinarily be necessary for *n*+2 parafoveal effects to occur. The effects we reported for the idioms were driven by a character presented three characters from fixation and word *n*+1 was not of particularly high frequency. A second, and we believe more problematic aspect of our results is that whilst these effects occurred for the idioms, they did not occur (or these effects were smaller) for otherwise comparable phrases matched for length and frequency. Thus, it is not apparent how frequency based "ease of processing" (of word *n*+1) arguments, or accounts based on limitations of the spatial extent of processing can explain the differential effects. A critical question concerns why the *n*+2 preview effects that we obtained occurred for idioms, but to a far lesser degree for matched phrases.

An obvious, potential, reason for these effects may have been that they were driven by the predictability of the second constituent of the idiom, given the first, or if words are lexically identified out of sequence, the predictability of the first constituent of the idiom, given the second. However, the findings from the present study are not due, or at least not entirely due, to the predictability of the second constituent given the first, the predictability of the first constituent given the second, or the transitional probability of the two constituents of idioms relative to phrases. In our pre-screen analyses we quantified these three variables and then introduced them as covariates in different versions of the analyses we undertook. In all these analyses, despite the inclusion of the covariates, the *n*+2 effects that we observed remained statistically robust, indicating that the processing advantage for idioms in relation to foveal and parafoveal processing during Chinese reading extends beyond the factors of predictability and transitional probability.

It is also worth briefly considering how processing would have to unfold if predictability were to be driving the effects that we observed. Recall that our stimuli were identical up to *n*+2, and that the sentences were completely neutral with respect to target predictability up to the target string. Recall also that the boundary in both Experiments was positioned prior to *n*+1. This means that in order for predictability to be responsible for the effects observed at *n*+2, then prior to crossing the boundary, participants must have parafoveally identified a two character word to the right of fixation (that was not predictable on the basis of sentential context), integrated it with sentential context, and on that basis, predicted *n*+2. We note that the n+1 constituents were not very high frequency words such as determiners, prepositions or "de" particles (Zang et al., 2018), and they were relatively long (97% of Chinese words are one or two characters long meaning that a Chinese two character word is relatively long). Under these circumstances, based on a significant body of empirical evidence, it seems unlikely (to us at least) that readers were processing words *n*+1 and *n*+2 to the degree that would be necessary for such effects to occur.

Before leaving this issue, it may be worth considering an important and neglected theoretical perspective in relation to the predictability account. As described, the predictability account rests upon word by word processing. That is, the predictability relation can only hold and exert an influence between constituents. However, the MCU hypothesis stipulates that these multi-constituent units have their own, individual, lexical representations and that (parafoveal and foveal) processing is operationalised across the extent of the unit to the degree that acuity permits. To be clear, the unit is treated as a whole. If this suggestion is correct, then an MCU would be processed somewhat similarly to a (longer) individual word. To this extent, the idea of there being a predictability dependency between the (separate) constituents of the unit does not really make too much sense. For example, does "choco" predict "late" in relation to the identification of "chocolate"? Or does "choc" predict "olate", or "cho" predict "colate"? Whilst it is possible that predictability relations between constituents might influence MCU identification, any such influence would be much more comparable to the orthographic constraints that early letters in a word impose on the likelihood of later letters appearing in a word (c.f., orthographic uniqueness points, Miller, Juhasz & Rayner, 2006), rather than reflecting a predictability relation between two different words that are represented lexically individually. Of course, in experiments investigating lexical processing during reading it is possible to quantify dependency relations between letter strings that appear early relative to those late in a word, but this would be considered one amongst many potential influences on the identification process. Similarly, any predictability relation between early and late constituents in a

MCU would likely be one influence amongst others in respect of the MCU's identification in the lexicon (again, we make these arguments under MCU processing assumptions). Critically, however, the characteristics of the MCU as a whole would be a primary determinant of ease of identification. Furthermore, in much the same way that all the letters of "chocolate" contribute to the word's identification (both in and of themselves, as well as in relation to each other in their respective positions), so too would the constituent letters and words of the MCU (e.g., "teddy bear"). Thus, according to the MCU hypothesis, the letters and constituents (or in this case characters in Chinese) of an MCU will be processed in parallel (contingent on acuity limitations) and those letters (or characters) represent the intrinsic characteristics of the MCU as a whole. The MCU hypothesis does not preclude influences of predictability (in the same way that theories of lexical processing do not preclude orthographic influences over lexical identification). Instead, the MCU hypothesis specifies that any such influences will operate in respect of the identification of the unit as a whole.

From our perspective, the present findings provide evidence to support the MCU hypothesis (Zang, 2019). Chinese idioms, especially those with a 2-character modifier, 1-character noun structure, appear to be represented lexically as MCUs and processed parafoveally and foveally as single lexical representations (despite being comprised of two words). According to the MCU Hypothesis, when two or more words comprise a lexical unit (e.g., in the case of an idiom), visual and linguistic processing is operationalized over both words simultaneously in the parafoveal region. Conversely, when they do not comprise a lexical unit, that is, when they form a phrase, visual and linguistic processing is operationalized sequentially over each successive word (since each of the words has its own lexical representation). Therefore, the critical issue is not one of serialism versus parallelism in processing. Instead, the issue is one of whether one, or multiple, upcoming words are represented and processed lexically as a single unit, that is, whether an upcoming word string is a lexicalised MCU with its own individual lexical representation. The next, fundamental, issues are what determines that multiple words will be processed as a MCU, and how readers segment MCUs from a series of consecutive Chinese character/word sequences during on-line sentence reading.

Previous theories of language use and processing have provided insight into the question of lexicalization. For instance, the Usage Based theory (Bybee, 2006) assumes that all linguistic units (words and multiple word sequences) are represented and processed in a similar way, and they are influenced by the frequency of occurrence. For example, using a phrasal-decision task, Arnon and Snider (2010) found frequency effects for four-word phrases (*don't have to worry* vs. *don't have to wait*) even when the frequency of their constituents was controlled, demonstrating that frequent multi-word phrases can be represented in a way similar to smaller units like words. It is possible that every time a word sequence is encountered and used, it activates nodes in the mental lexicon, and over time, these specific patterns of node activation gradually come to be recognized and processed as a single unit. Thus, frequently encountered word sequences could be lexically represented and processed as single units (see also Zang et al., 2020). The Exemplar Based theory (Bod, 2006) posits that linguistic experience, rather than abstract linguistic rules, determines language acquisition and processing. Our results are compatible with these views; it appears that to a significant extent, the frequency of exposure is a key determinant of the formation of a MCU within the language. With increased frequency of exposure, the degree to which the elements of the unit are represented and licensed to be processed as a whole increases gradually. However, more research is required to investigate the precise nature of the mechanism underlying how a MCU comes to be represented and processed lexically. Finally, as we discussed earlier, a recent model of eye movement control during Chinese reading, the CRM (Li & Pollatsek, 2020) has been put forward and this might have the potential to account for the present findings if it were modified such that MCUs, as well as words, could be lexically represented and identified as single units. That is to say, the CRM has the potential to extend the operationalization of lexical processing beyond individual words to frequently occurring MCUs in Chinese reading.

To summarize, the present study investigated whether Chinese idioms with modifier-noun structure that are very likely to be MCU candidates due to significant modifier constraint over the subsequent noun, are foveally and parafoveally processed as MCUs during natural reading. In Experiment 1, we observed that a larger preview effect occurred for idioms on the modifier with shorter fixations for identical than unrelated and pseudocharacter previews, suggesting that idioms are parafoveally processed to a greater extent than phrases. In Experiment 2, we observed for identity modifiers, a greater preview effect from nouns occurred for idioms relative to phrases, suggesting modifier-noun idioms are MCUs over which parafoveal and foveal processes are operationalized, and for which there are single unified lexical representations.

References

Angele, B., & Rayner, K. (2011). Parafoveal processing of word n + 2 during reading: Do the preceding words matter? *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1210-1220.

Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating n + 1 and n + 2 previews simultaneously. *Visual Cognition*, *16*, 697-707.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62(1)*, 67-82.

Barr, D., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5*, e10729.

Carrol, G., & Conklin, K. (2017). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea “has legs’. Bilingualism: Language and Cognition, 20, 299-317.

Chinese Idiom Dictionary [Hanyu Guanyongyu Cidian]. (2009). Beijing: The Commercial Press.

Contemporary Chinese dictionary [XianDai Hanyu Cidian]. (2008). Beijing: The Commercial Press.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72-89.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32,* 45-61.

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1778-1786.

Engbert, R. & Kliegl, R. (2011). Parallel graded attention models of reading. In Liversedge, S.P., Gilchrist, I.D. & Everling, S. (Eds.), *The Oxford Handbook of Eye Movements* (pp.787-800). New York, NY: Oxford University Press.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research. 42*, 621-636.

Engbert, R., Nuthmann, A., Richter, E.M., and Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777-813.

He, L., Song, Z., Chang, M., Zang, C., Yan, G., & Liversedge, S.P. (2021). Contrasting off-line word segmentation with on-line word segmentation during reading. *British Journal of Psychology*, in press.

Hoosain, R. (1992). Psychological reality of the word in Chinese. In H. C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 111-130). Amsterdam, the Netherlands: North-Holland.

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n + 2. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1250-1255.

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

Li, X., Zang, C., Liversedge, S.P., & Pollatsek, A. (2015). The role of words in Chinese reading. In Pollatsek, A., & Treiman, R. (Eds). *The Oxford Handbook of Reading* (pp. 232-244). New York, NY: Oxford University Press.

Liu, P., Li, W., Lin, N., Li, X. (2013). Do Chinese readers follow the national standard rules for word segmentation during reading? *PLoS ONE, 8*, e55440.

Liversedge, S.P. & Findlay, J.M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science*, *4*, 6-14.

Miller, B., Juhasz, B. J., & Rayner, K. (2006). The orthographic uniqueness point and eye movements during reading. *British Journal of Psychology, 97(2)*, 191–216.

Modern Chinese Idiom Standard Dictionary [XianDai Guanyongyu Guifan CiDian]. (2001). Changchun: Changchun Publishing House.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Radach, R., Inhoff, A. W., Glover, L., & Vorstius, C. (2013). Contextual constraint and n + 2 preview effects in reading. *The Quarterly Journal of Experimental Psychology, 66*, 619-633.

Rayner, K. (1975). The perceptual span and peripheral cues during reading. *Cognitive Psychology, 7,* 65-81.

Rayner, K. (2009). The thirty-fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1506.

Reichle, E. D., (2011). *Serial-Attention Models of Reading*. In Liversedge, S.P., Gilchrist, I.D. & Everling, S. (Eds.), The Oxford Handbook of Eye Movements (pp.767-786). New York, NY: Oxford University Press.

Rayner, K., Juhasz, B. J., & Brown, S. J. (2007). Do readers obtain preview benefit from word n + 2? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 230-245.

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13,* 115-119.

Reichle, E.D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445-476.

Risse, S., Hohenstein, S., Kliegl, R., & Engbert, R. (2014). A theoretical analysis of the perceptual span based on SWIFT simulations of the *n*+2 boundary paradigm. *Visual Cognition, 22*(3-4), 283-308.

Schotter, E.R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74*, 5-35.

Schotter, E.R., Reichle, E.D., & Rayner, K. (2014). Re-thinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and N+2 preview effects. *Visual Cognition, 22*, 309-333.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 776-784.

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences. 23*, 537-546.

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-Reader: A model of word recognition and eye movements in text reading. *Psychological Review, 125(6),* 969-984.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior, 18*, 523-534.

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31(12),* 1655-1674.

Vasilev, M.R., & Angele, B. (2017). Parafoveal preview effects from word N+1 and word N+2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin and Review, 24,* 666-689.

Yan, M., Kliegl, R., Shu, H., Pan, J., & Zhou, X. (2010). Parafoveal load of word n + 1 modulates preprocessing effectiveness of word n + 2 in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1669-1676.

Yang, J., Rayner, K., Li, N., & Wang, S. (2012). Is preview benefit from word n + 2 a common effect in reading Chinese? Evidence from eye movements. *Reading and Writing, 25*, 1079-1091.

Yang, J., Wang, S., Xu, Y., & Rayner, K. (2009). Do Chinese readers obtain preview benefit from word n + 2? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1192-1204.

Yu, L., Cutter, M.G., Yan, G., Bai, X., Fu, Y., Drieghe, D., & Liversedge, S.P. (2016). Word n+2 preview effects in three-character Chinese idioms and phrases. *Language, Cognition and Neuroscience, 31*, 1130-1149.

Zang, C. (2019). New perspectives on serialism and parallelism in oculomotor control during reading: The multi-constituent unit hypothesis. *Vision, 3, 50*, 1-13.

Zang, C., Du, H., Bai, X., Yan, G., & Liversedge (2021). Preview effects in reading of Chinese two-constituent words and phrases, *submitted*.

Zang, C., Liversedge, S.P., Bai, X., & Yan, G. (2011). Eye movements during Chinese reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling. (Eds). *The*

Oxford Handbook on Eye Movements (pp. 961–978). New York, NY: Oxford University Press.

Zang, C., Zhang, M., Bai, X., Yan, G., Angele, B., & Liversedge, S.P. (2018). Skipping of the very high frequency structural particle de in Chinese reading, *The Quarterly Journal of Experimental Psychology, 71(1),* 152-160.

Acknowledgments

We acknowledge support from ESRC Grant (ES/R003386/1).