Journal of Memory and Language 136 (2024) 104508

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Parafoveal processing of Chinese four-character idioms and phrases in reading: Evidence for multi-constituent unit hypothesis

Chuanli Zang [a,b,*], Shuangshuang Wang [b], Xuejun Bai [b], Guoli Yan [b], Simon P. Liversedge [a]

[a] School of Psychology and Humanities, University of Central Lancashire, United Kingdom
[b] Key Research Base of Humanities and Social Sciences of the Ministry of Education, Faculty of Psychology, Tianjin Normal University, China

ARTICLE INFO

Keywords:
Multi-Constituent Unit (MCUs)
Idioms
Preview effects

ABSTRACT

The perceptual span in Chinese reading extends one character to the left and three to the right of the point of fixation. Thus, four-character idioms and phrases often extend rightward beyond these limits during reading. We investigated whether such idioms, frequent phrases and equibiased strings are processed parafoveally as Multi-Constituent Units (MCUs). Using the boundary paradigm in Experiments 1 and 2, we separately manipulated preview (identities or pseudocharacters) of the first two and the last two characters of idioms and frequently used phrases. In Experiment 3, we examined processing of strings judged to be a single lexical unit, equi-biased ambiguous strings and matched unambiguous multi-word strings. Experiments 1 and 2 produced greater preview benefit for the final two characters when the first two characters were presented after identity rather than pseudocharacter previews. In Experiment 3, preview effects were largest for single units, reduced for equi-biased strings and smallest for multi-word strings. Together the results demonstrate that four-character idioms and frequently used phrases are processed as MCUs.

Introduction

Over the past few decades, a great deal has been learned about eye movement control during reading. One suggestion that has been widely accepted is that the decision about when readers move their eyes during reading is primarily determined by word identification, and that the decision about where readers fixate next is also word based (e.g., Cutter, Drieghe & Liversedge, 2017, 2018; Liversedge & Findlay, 2000; Rayner, 1998, 2009). This word-based processing approach has dominated much of the research investigating eye movements and reading over many years. And this is understandable as the vast majority of research that was conducted early in this field was based on alphabetic writing systems such as English, whereby each word is visually separated from other words by spaces and is, thus, a perceptually salient and unambiguous unit. However, in non-alphabetic languages like Chinese, no spaces or other forms of visual or lexical cues are available to indicate where a word starts and ends, and there is often ambiguity regarding which character strings actually comprise a word (Bai et al., 2008; He et al., 2021; Hoosain, 1992; Liu et al., 2013). This orthographic characteristic raises the question as to what are the linguistic units over which processes operate during natural Chinese reading? And, beyond this, how do Chinese readers compute where words begin and end? Put differently, is the traditional word-based approach appropriate for unspaced languages like Chinese? And is there viability in adopting a less rigid view as to what constitutes the unit over which visual, linguistic and oculomotor control processes are operationalized during reading (or perhaps during any particular fixation during reading)?

The Multi-Constituent Unit (MCU) Hypothesis that has recently been proposed (Zang, 2019; Zang et al., 2021, 2023) might offer potential to elucidate these questions in respect of eye movement control during reading. This hypothesis rests on a simple idea such that readers may have a mental lexicon formed of individual word entries, but also entries corresponding to frequently occurring MCUs such as “teddy bear”, “kick the bucket” and “fish and chips”, etc. Each MCU is comprised of more than a single word and each has its own lexical entry and semantic representation. We consider MCUs to refer to frequently used multiple constituent, often multiple word, strings that might be represented lexically as single units. A MCU must be recognizable as a single unit and will likely have meaning beyond the meaning of its constituent elements (in the sense that “fish and chips” in the UK does not mean “fish” alongside “chips”, but instead, a particular meal often served in the UK). It is important to be clear, here, that in putting forward the MCU

* Corresponding author at: School of Psychology and Humanities, University of Central Lancashire, Preston, Lancashire PR1 2HE, United Kingdom.
E-mail address: czang@uclan.ac.uk (C. Zang).

https://doi.org/10.1016/j.jml.2024.104508
Received 26 February 2023; Received in revised form 30 January 2024; Accepted 31 January 2024
Available online 17 February 2024

Hypothesis, we are not only suggesting that MCUs are lexicalised (and we fully acknowledge that this idea has been suggested by several other researchers in the past, e.g., Conklin & Schmitt, 2008, 2012; Shaoul & Westbury, 2011; Siyanova-Chanturia et al., 2011; Titone & Connine, 1999; Wood, 2015; Wray, 2002; Wulff & Titone, 2014), but beyond this, we are suggesting that during natural reading, parafoveal and foveal processing, and the computation of oculomotor metrics (on-line decisions of where and when to move the eyes) are very likely operationalized over these units. That is to say, we are making claims with respect to process as well as representation. From this perspective, the elements that are parafoveally, and foveally, processed will vary from fixation to fixation (contingent on the degree to which those upcoming constituents are lexicalised and recognisable as single units represented in the lexicon). Adopting this perspective provides an opportunity to move away from the idea of "word-based processing" at a general level (regardless of any suggestions as to whether such word-based processing might occur serially or in parallel with respect to lexical identification) to a position where processing occurs sequentially, but in respect of lexicalised units that themselves may be words or MCUs. If the MCU Hypothesis is supported, then we feel that it might allow for flexible and dynamic visual and cognitive processing during reading, rather than reading processes being tied to units of orthography that are often constant and fixed.

It is important to be as clear as possible regarding the candidate strings that are likely to be MCUs, and we next focus on Chinese MCUs since the current experiments investigate Chinese reading. In our view, MCUs in Chinese are likely to be comprised of frequently used multiple word strings that might be represented lexically as single units. For example, 绿草, in which 绿 means *green* and 草 means *grass*, and in which each of the two constituent characters is a free morpheme and a single-character word. Each constituent can be substituted (e.g., 枯草 means *dry grass*, and 绿墙 means *green wall*), thus together 绿草 is a phrase. Because it is frequently used, we assume that it is represented as a unit in lexicon in the same way that individual words like 绿茶 (which means *green tea*, and which is a word because neither of the constituent characters of 绿茶can be substituted in order to maintain its special meaning, that is, the particular tea) are represented (Zang et al., 2023). Perhaps more interesting and compelling MCU candidates in relation to the current theoretical claims are more freestanding units such as frequently used Chinese idioms (e.g., 铁饭碗, 铁 meaning *metal*, 饭碗 meaning *rice bowl*, whilst the full idiom 铁饭碗 means *a secure job,* Zang et al., 2023), familiar phrases such as 森林公园 (meaning *forest park*, Li, Zang, Liversedge, & Pollatsek, 2015) and perhaps modern internet phrases such as 新冠 (meaning *new coronavirus variant*), etc. These MCUs may be lexically identified as a single element during reading, and if this is the case, then visual, linguistic processing and oculomotor commitments may take place accordingly in respect of those units in the same way that they do with respect to a word.

Some may consider that this definition of a MCU may be unclear in that under our suggestion, opaque Chinese compound words might be categorised as MCUs rather than words.[1] To illustrate, consider the example "马上" which means "immediately" and comprises the words "马" (itself meaning "horse") and "上" (meaning "up" or "get on"). According to our MCU definition, "马上" might be considered to meet the criteria for categorisation as a MCU, and this seems implausible given that almost all Chinese readers would very likely agree that "马上" is a word. On the face of it, this might appear a strong argument to suggest that the definition of a MCU is ambiguous, and therefore, unhelpful. However, we do not consider that this is the case for two reasons. First, when children and those learning Chinese as a second language are taught to read Chinese, the word "马上" is explicitly taught to those individuals to be a word. They do not establish this lexical entry on the basis of encountering it multiple times and deriving its meaning as being a meaning that is distinct from the meaning of its compositional characters. Instead, they are instructed to represent the character string as a word meaning "immediately". It is for this reason that there is absolutely no ambiguity amongst Chinese readers as to this being a word (and that it certainly is not an MCU). Given this, it is important that we also stipulate within our definition of a MCU that character strings that are explicitly taught to be words fall outside of the MCU category.

Beyond issues of direct instruction, there is a second reason why we maintain our view that the MCU definition is helpful. Note, again, that MCUs are phrases (again, not words) that are comprised of multiple words that are each free morphemes. Thus, whilst the constituent characters of "马上" are free morphemes, if either is replaced in the two character string by an alternative character that produces a legal two character string, then the meaning of the character string fundamentally changes to be different from "immediately". As can be seen from our earlier example, for the frequently used two-character string "绿草", a well recognised unitary phrase, each character can be substituted and even with substitutions, the phrase retains its basic meaning. In contrast, if we replace either of the characters in "马上" with an alternative character that produces a legal two character string, then the meaning of the string fundamentally changes to be different from "immediately" (recall that "马" means "horse" and "上" means "up", but in "马虎", "马" means horse, "虎" means "tiger", but together the two character word means "careless", a meaning completely unrelated to the meaning of either of the constituent characters, nor the meaning "immediately"). Thus, on this view, a word is the smallest, indivisible (or non-compositional) construct that if further divided, breaks down into meaningless syllables or other words. With this stipulation in mind, we feel that the MCU definition remains useful. It is also important to note that idioms are widely regarded as being distinct from words within the Chinese linguistic literature. For example, Huang and Liao (2007, P266) state: "Idioms are commonly used standardized *fixed phrases*, and they are a special type of lexical unit" and "Idioms are a type of *fixed phrase* that has been passed down traditionally and used now, with rich and written language-style meanings". Furthermore, the Chinese Academy of Social Sciences, Institute of Linguistics, (2016), Modern Chinese Dictionary (7th Edition) states: "Idioms are concise, fixed *word combinations* (or *word groups*) or *short sentences* that people have been using for a long time" (our italics). Finally, our MCU definition might also be criticised because many idioms or set phrases are listed as words in Chinese dictionaries (including very influential volumes such as the "Modern Chinese Dictionary", 2016, 7th Edition and "Lexicon of Common Words in Contemporary Chinese" by Li and Su, 2021, 2nd Edition). However, this idea is often referred to as "listedness" in the linguistic literature and it has long been argued that listedness should not be adopted as a defining criterion for words for two reasons: (1) often strings that are not words are listed; (2) often strings that are words are not listed (Packard, 2003; Di Sciullo & Williams, 1987). Thus, the presence or absence of character strings in "word" listings seems an inappropriate basis on which to define a word (or criticise our definition of a MCU).

As we noted previously, the MCU hypothesis has been significantly influenced by work on formulaic language including multi-word sequences, collocations, idioms, prefabricated chunks, lexical bundles and so on (e.g., Conklin & Schmitt, 2008, 2012; Shaoul & Westbury, 2011; Siyanova-Chanturia et al., 2011; Titone, & Connine, 1999; Wood, 2015; Wray, 2002; Wulff & Titone, 2014). These studies have demonstrated that multi-constituent units or multi-word units that occur often together can be represented lexically. Here we adopt the term MCU. However, as we have noted, the novelty of the present approach, is that such lexicalized elements will be directly associated with the operationalization of foveal, parafoveal processing and oculomotor commitments that occur on a moment-by-moment basis during natural sentence reading. Furthermore, it is our view that such a characterization of processing might allow for new perspectives in respect of a longstanding theoretical controversy in the field of eye movement

[1] The authors are grateful to Xingshan Li for making them aware of this point.

control during reading, namely, whether words are lexically identified serially and sequentially, or instead in parallel.

The debate regarding serialism or parallelism of lexical identification in reading has played out largely in the context of formal computational models of eye movement control. For instance, the E-Z Reader model (e.g., Reichle et al., 1998; Reichle, 2011) operates according to a serial and sequential framework, such that attention, acting something like a spotlight, is allocated to only one word at a time, in a strictly serial order. Therefore, words are lexically processed sequentially, with lexical processing of the parafoveal word $n + 1$ only occurring after lexical processing of the current word n has been completed. According to the E-Z Reader model, lexical processing of the parafoveal word $n + 2$ should not start while readers are fixating word n (unless word $n + 1$ is highly frequent and/ or very short, and thus recognized very rapidly, Reichle & Schotter, 2020). In contrast, the SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert & Kliegl, 2011) operates according to a parallel graded framework, and posits that attention is distributed over multiple words within the perceptual span (McConkie & Rayner, 1975). Therefore, multiple words can be lexically identified simultaneously, with the parafoveal word $n + 1$ or even $n + 2$ potentially being lexically identified before they receive a fixation and before word n is identified. It is important to note, though, that the existing literature on reading of both alphabetic and non-alphabetic languages has demonstrated that $n + 2$ preview effects are negligible, or are at best subtle, and do not generally occur when $n + 1$ is of low frequency (e.g., Yan, Kliegl, Shu, Pan, & Zhou, 2010; Yang, Wang, Xu, & Rayner, 2009; Yang, Rayner, Li, & Wang, 2012; see also Zang, Liversedge, Bai & Yan, 2011 for a review), and/or more than three letters long (e.g., Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Rayner, Juhasz, & Brown, 2007; see also Vasilev & Angele, 2017 for a review). In response to these findings, advocates of SWIFT have argued that a word $n + 1$ that is of increased length could result in word $n + 2$ being pushed further away from fixation, potentially out of the perceptual span, thereby preventing effective preprocessing of $n + 2$ while the eyes are still on word n. Furthermore, SWIFT has been criticized with respect to how readers interpret words incrementally for sentence comprehension if those words are recognized out of sequential order. In an attempt to deal with this issue, the OB1-Reader model (Snell, van Leipsig, Grainger, & Meeter, 2018) adopts a similar parallel processing approach to that in the SWIFT model, but introduces a spatial mapping mechanism whereby a spatiotopic sentence-level representation in working memory provides a reference frame for representing the location of each word in a sentence (usually for words n-2 to $n + 2$ around the point of fixation). As such, word identification is a process of mapping activated representations onto possible spatial locations in the spatiotopic representation, and this is done on the basis of their approximate lengths as determined by visual cues, spaces, and expected syntactic and/or semantic information from the visual input. One issue for this account is that it is not obvious as to how the spatial mapping mechanism might operate when applied to unspaced alphabetic and non-alphabetic languages where word boundary information is not clearly marked.

Recently, a new Chinese Reading Model (CRM) of word segmentation, identification and eye movement control during Chinese reading has been proposed (Li & Pollatsek, 2020). This model is, arguably, the most relevant to the present work as it solely seeks to account for eye movement control in Chinese reading and it incorporates procedures by which word segmentation during reading is achieved. At the core of the model is an interactive activation lexical identification system (McClelland & Rumelhart, 1981) with representations within this system becoming activated on the basis of orthographic information obtained from foveal and parafoveal vision. Visual input feeds directly into the lexical processing system activating candidate representations that are consistent with that input. Through the antagonistic processes of activation and mutual inhibition, the activated representations compete against each other until a single lexical entry is activated beyond others at which point that word is lexically identified and simultaneously segmented from the upcoming character string. In this way, the CRM does not take a clear perspective with respect to whether lexical identification occurs serially or in parallel.

Arguably, it is the case that the theoretical debate with respect to serialism and parallelism in lexical identification and oculomotor control during reading has reached something of an impasse. Evidence has been presented to support the serial position and evidence has been presented to support the parallel position with those favoring each position championing the particular evidence in their support. Nonetheless, the debate is still alive, with some even arguing that it might be impossible to resolve the issue in natural sentence reading using eye-tracking methodology (Snell & Grainger, 2019, though see Schotter & Payne, 2019). Furthermore, as discussed earlier, regardless of whether lexical processing might be serial or parallel, all the current eye movement control models have considered processing to be word based by default. However, a word-based processing perspective has become increasingly challenging to consider in respect of reading unspaced text like Chinese, given that the length of any particular word is not demarcated, and sometimes the concept of a word is ambiguous, with significant disagreement from individual to individual concerning the particular characters in a sentence that form a particular word. The MCU hypothesis, to us, takes first steps to attempt to reconcile (at least to some extent) the serial/parallel debate. Further, it allows us to consider a position beyond word-based processing. To be specific, this hypothesis allows us to explain how readers might operationalize processing at a lexical level serially and sequentially, but process more than one word at a time. Also, elements that are represented lexically will be determinant of how many words are processed during any particular fixation. To this extent, the MCU hypothesis has the potential to offer theoretical progression. We also note that the MCU hypothesis may offer theoretical value in the consideration of agglutinate languages like Finnish, Turkish and German that are word spaced but frequently have long words that are themselves comprised of multiple lexical units joined together without spaces. We will return to its implications in relation to processing during reading in agglutinate languages in the General Discussion.

There is increasing empirical evidence in favour of the MCU hypothesis. Initial research showed that frequently co-occurring two-constituent spaced compounds (e.g., “teddy bear”) operate as MCUs during English sentence reading (Cutter, Drieghe, & Liversedge, 2014). Using the boundary paradigm (Rayner, 1975), Cutter et al. manipulated the preview of each constituent (i.e., $n + 1$ and $n + 2$) to be presented in full as an identity, or as a nonsense letter string. The boundary was placed prior to the first constituent of the spaced compound, and before the eyes crossed the boundary, a preview was presented at the target location. Once the eyes crossed the boundary, the preview was replaced with the whole spaced compound. A reliable interaction between the previews of each constituent was obtained, with a robust $n + 2$ preview effect only when the $n + 1$ identity preview was available in the parafovea. In other words, when the first constituent (e.g., “teddy”) was present in the parafovea before the eyes crossed the boundary, compared with a nonsense string, readers pre-processed the second constituent (e.g., “bear”) to a far greater extent, even though the first constituent was five or six letters long. Thus, the parafoveal presentation of the first constituent appeared to license processing beyond that word to the second constituent forming the base compound, and Cutter et al. argued that this happens because the two constituents are represented as a single lexical entry and processed as a MCU.

In line with Cutter et al.’s study, Zang et al. (2021, 2023) provide further evidence to suggest that frequently used Chinese two-character phrases, and three-character idioms with a modifier-noun structure can be represented lexically as single units and processed foveally (see also Yu et al., 2016) and parafoveally as MCUs during Chinese reading. The boundary paradigm was used in their experiments, with the boundary located before the target string and the preview being two to three characters away from the point of fixation. Zang et al (2023)

included three types of two-constituent Chinese strings (words, frequently occurring word pairs forming recognizable and familiar MCUs, or simple matched phrases) that shared identical first constituents. In their experiments, Zang et al., manipulated the preview of the second constituent to be an identity or a pseudocharacter and obtained a reliable preview effect when the second constituent was part of a word or a MCU, but not a phrase, indicating that frequently used MCUs, like words, were lexicalized and processed as single units. Similarly, Zang et al. (2021, 2023) examined whether three-character idioms with a modifier-noun (MN) structure are processed as MCUs. These idioms are comprised of either a two-character modifier and a one-character noun (e.g., '乌纱' means *black gauze*, '帽' means *cap*, together the idiom '乌纱帽' means *an official post*) or a one-character modifier and a two-character noun (e.g., '铁' means *metal*, '饭碗' means *rice bowl*, together the idiom '铁饭碗' means *a secure job*). The counterpart matched phrases were included with identical first constituents to the corresponding idioms, and the preview of their second constituents was manipulated to be an identity or a pseudocharacter using the boundary paradigm. Consistently, the results showed more pronounced preview benefit for idioms than matched phrases, suggesting that relative to phrases, these three-character idioms are processed parafoveally as unified MCUs and are very likely lexicalized.

Highly familiar idioms that are four characters long in Chinese might also have the potential to be MCUs and to be processed as single lexical representations parafoveally during reading, though it is important to note that four character strings are quite long. Approximately 80% of Chinese words are one or two characters in length (Lexicon of Common Words in Contemporary Chinese Research Team, 2008), and thus, four-character idioms are character strings of length that might ordinarily be formed from two or three words. Furthermore, the perceptual span in Chinese, that is, the region to the right of fixation from which useful information is obtained during a fixation, is estimated to be about 3 characters (Inhoff & Liu, 1998). This means that a four-character idiom that sits directly to the right of the point of fixation extends slightly beyond the right-extent of the perceptual span. Consequently, idioms of this length, positioned in this way with respect to fixation may not be processed as single units due to perceptual constraints. However, if these long idioms were processed as single representations parafoveally, then this would provide very strong evidence for the MCU hypothesis. That is to say, the four-character idioms represent quite a strong test of the MCU hypothesis. Given that these four-character idioms are used frequently in Chinese and have a relatively fixed structure (e.g., 咬文嚼字, '咬文' means *bite phrases* and '嚼字' means *chew characters*, together it means *choose words with great care*), it is not surprising that some researchers have considered them as long words by default (Gu et al., 2023; Li, Rayner, & Cave, 2009). For example, in a series of naming and character detection tasks, Li et al (2009) asked participants to report as many characters as possible after they were presented briefly with four-character strings on a screen in which the four characters formed a single four-character word (actually, an idiom, e.g., 不知所措meaning *be at a loss*), two two-character related words (e.g., 美满婚姻, meaning *happy marriage*), two two-character unrelated words (e.g., 急速切实 meaning *quick, feasible*), and a four-character nonword (e.g., 艾抵积促). Characters were reported most accurately in the single word condition, less in the related words condition, and least in the unrelated words and nonword conditions. These findings indicate that four-character strings can be recalled more effectively in the foveal region when they form a meaningful unit such as an idiom, or co-occur frequently together as units of a familiar phrase. Apparently, with reading experience, multiple words that frequently co-occur become bound together and are processed (and arguably lexically represented) as a single unit. Note, though, the stimuli in the Li et al. study were selected from a corpus in which word units were identified via automatic computational segmentation algorithms (word segmentation programs). It is known that human word segmentation is often ambiguous, leading to various segmentations regarding units as words, idioms and phrases among different readers.

In He et al.'s (2021) study, they focused on this issue and conducted a large scale pre-screen segmentation judgement study to examine the linkage between off-line segmentation preferences and on-line processing commitments. Robust effects were observed in their second experiment, in which three sets of four-character strings were selected, with the first set of strings being predominantly categorized as being a single four-character word (e.g., one-word strings), the second set unambiguously categorized as two separate two-character words (two-word strings), and the third set categorized approximately equally often as a single four-character word and as two separate two-character words (ambiguous strings). Furthermore, two groups of participants were identified with each group consistently segmenting four-character strings as single four-character words (1-word segmenters), or two two-character words (2-word segmenters) respectively. These participants were required to read sentences containing target strings from the three groups. He et al. found that participants spent less time reading one-word strings than ambiguous and two-word strings, regardless of whether participants were categorized as one-word or two-word segmenters in the off-line segmentation task. In relation to the present study, He et al.'s findings suggest that one-word strings that are comprised of two two-character words but widely categorized as single words may have been lexicalized, represented in the mental lexicon and processed during reading as single units widely across participants.

It has been demonstrated that meanings of idioms are often retrieved directly from memory (e.g., Libben & Titone, 2008; Titone & Connine, 1999; Swinney & Cutler, 1979), therefore, idioms and perhaps also familiar phrases are good candidates to be processed as a whole parafoveally. In Experiment 1, these four-character idioms were selected as target strings and the preview of each constituent (comprised of the first and the second two characters, $n + 1$ and $n + 2$) using the boundary paradigm, with the boundary placed prior to the four-character idioms. If these idioms are processed as MCUs, then the first constituent should activate the whole lexical unit, then this activation will license preprocessing of the second constituent to a greater extent, and thus more parafoveal processing of the second constituent should be obtained. Experiment 2 aimed to extend the findings from Experiment 1 and examined whether four character phrases that are highly familiar to Chinese readers and are often judged as single four-character words as per He et al. (2021), are also processed as MCUs. If both experiments show consistent preview effects associated with the second constituent, then this is in line with our claim that the lexical status of these multi-character strings will determine whether they are, or are not, processed as MCUs.

In Experiment 3, all the target strings were selected from Experiment 2 of He et al. (2021) and were always comprised of two two-character words. However, as described earlier, in an off-line prescreen, these strings were categorized as either being one-word strings (i.e., the same as the phrases used in Experiment 2), two separate word strings, or ambiguous strings. Each set of four-character strings shared the initial two-character word (i.e., the first constituent, $n + 1$), and the second two-character words (i.e., the second constituent, $n + 2$) were matched across conditions for frequency and stroke complexity (see details in the Method section of Experiment 3). Again, the boundary paradigm was used to manipulate the preview of the second constituent of the four-character string with the boundary placed prior to the entire four-character target string. We predicted greatest parafoveal preview benefit for the second constituent for the one-word strings, an intermediate level of benefit for the ambiguous strings and least for the two-word strings.

Table 1
The stroke number of each constituent character of the identity and its counterpart pseudocharacter preview in Experiment 1.

	C/P 1	C/P 2	C/P 3	C/P 4	C/P 1 + 2	C/P 3 + 4
Identity	5.9 (3.0)	6.8 (2.9)	7.3 (3.1)	8.3 (3.6)	12.7 (4.0)	15.6 (4.8)
Pseudocharacter	6.0 (2.9)	6.8 (2.9)	7.3 (3.1)	8.3 (3.7)	12.8 (3.8)	15.7 (4.8)

Standard deviations are provided in parentheses. C = Character; P = Pseudocharacter.

Experiment 1

Method

Participants

One hundred native Chinese-speaking students (82 females, mean age 21 years) from Tianjin Normal University participated in the experiment. They had normal or corrected-to-normal vision and were naïve as to the purpose of the experiment.

Apparatus

Participants' eye movements were monitored using an SR Research Eyelink 1000 with a sampling rate of 1000 Hz. Sentences were presented in Song font, on a single line on a 24-inch DELL CRT monitor with a refresh rate of 144 Hz and a screen resolution of 1024 × 768 pixels. At a viewing distance of 70 cm, each Chinese character corresponded to approximately 1.1 degree of visual angle.

Materials and design

Eighty four-character idioms were selected from an on-line Idiom Dictionary (https://cy.5156edu.com). To confirm that Chinese readers were familiar with these idioms, 16 readers who did not participate in the main eye-tracking study were required to provide the meaning of each idiom and rate its familiarity. The prescreen assessment showed that these idioms were correctly defined by over 85 % of participants (M = 97 %, SD = 4 %), and their familiarity was 4.4 (SD = 0.3) on a 5-point scale ("1″ = "very unfamiliar", "5" = "very familiar").

Eighty Chinese sentences were constructed with each target idiom embedded in roughly the middle of each sentence. The sentences ranged between 15 and 25 characters in length and were rated on a 5-point scale ("1″ = very natural, "5" = very unnatural) for their naturalness by 15 participants who did not participate in the eye-tracking study. The mean naturalness was 1.5 (SD = 0.3). In addition, a group of 15 participants were presented with the sentence frame up to, but not including the target, and were asked to complete the sentence. The idioms were unpredictable (M = 0.7 %, SD = 2.3 %) given the preceding sentential context. A separate group of 19 participants were required to assess the predictability of the second constituent given the prior context including the first constituent, and the second constituent was generated 47 % (SD = 25 %) of the time, though further analyses showed that its predictability did not affect any of the findings of the study (for details, see Table A1 in the Appendix). Furthermore, according to Chinese Lexical Database (CLD, Sun, Hendrix, Ma, & Baayen, 2018), the conditional probability of the second constituent of idioms given the first, computed by the frequency of the idiom/the sum of frequencies of all idioms with the same first constituents, was relatively low (19 %, SD = 14 %). In other words, the whole idioms were generated less than 20 % of the time on the basis of the first constituent. Further analyses showed that this variable did not affect the results (see Table A2 in the Appendix).

Using the gaze-contingent boundary paradigm (Rayner, 1975), the preview of each constituent (n + 1, the first two characters; n + 2, the last two characters) of an idiom was orthogonally manipulated to be either an identity or a pseudocharacter preview (as per Cutter et al., 2014). Hence, the current experiment adopted a 2 (n + 1 preview: Identity or Pseudocharacter) × 2 (n + 2 preview: Identity or Pseudocharacter) within-participant repeated measures design. Pseudocharacters are very rarely used Chinese characters but which have been categorized as pseudocharacters in a prescreen test by 30 university students (Zang et al., 2020, 2021, 2023). As shown in Table 1, each constituent character of the identity (idiom) and its counterpart pseudocharacter preview was controlled for the number of strokes, all t < 1.90. An example set of sentences under different preview conditions is illustrated in Fig. 1.

Procedure

Each participant filled out a written consent form upon their arrival. They were then seated at the eye tracker with their head stabilized by a head and chin rest. The eye tracker was calibrated using a three-point horizontal calibration procedure with an average error below 0.20 degrees of visual angle. Once the calibration was completed successfully, sentences were presented in turn. Following each sentence, the calibration was checked using a drift check procedure and participants were recalibrated whenever necessary. After completing reading a sentence, they pressed response keys to terminate the sentence display. There were six practice sentences to familiarize the participant with the procedure, and 80 experimental sentences interspersed with 40 filler sentences without display changes. Both experimental and filler sentences were presented in a random order for each participant and experimental conditions were rotated across four files with each participant reading sentences only from one of the files. Participants were instructed to read each sentence silently for comprehension. On one third of the trials, they were presented with a comprehension question. After the experiment, participants were asked to complete a display change awareness questionnaire to report whether they had noticed any changes to the characters or text while they were reading. The whole experiment took

n+1 Preview	*n*+2 Preview	Sentence
Identity	Identity	真正的友谊是在|**风雨同舟**的旅程中缔造的。
	Pseudocharacter	真正的友谊是在|**风雨丞兗**的旅程中缔造的。
Pseudocharacter	Identity	真正的友谊是在|**闫眚同舟**的旅程中缔造的。
	Pseudocharacter	真正的友谊是在|**闫眚丞兗**的旅程中缔造的。

Fig. 1. An example of the stimuli under different preview conditions used in Experiment 1. The preview of the first and the second constituent of the target string was manipulated with an invisible boundary (represented by the vertical line) positioned prior to the target. Before the eyes crossed the boundary, the preview appeared at the position of the target string. Once the eyes crossed the boundary, the preview changed to the correct target (both the target and the preview are in bold for illustrative purposes, but were presented normally in the experiment). The idiom "风雨同舟" literally means in the same boat under wind and rain and figuratively means face challenges and adversities. The English translation for the sentence is "True friendships are built on the journey of facing challenges and adversities together".

Table 2
Eye movement measures for all regions in Experiment 1.

Region	$n+1$ preview	$n+2$ preview	FFD	SFD	GD	Go-past	TFD	SP
The pretarget region (n)	Identity	Identity	221(4)	219(4)	240(6)	310(9)	341(10)	0.28(0.02)
		Pseudocharacter	220(4)	217(4)	241(6)	306(10)	354(12)	0.29(0.02)
	Pseudocharacter	Identity	223(4)	221(4)	247(6)	317(10)	393(14)	0.28(0.02)
		Pseudocharacter	217(4)	216(4)	239(5)	298(9)	379(13)	0.27(0.02)
Constituent 1 ($n+1$)	Identity	Identity	233(4)	233(5)	256(6)	318(9)	343(11)	0.30(0.02)
		Pseudocharacter	241(5)	239(5)	266(6)	334(11)	359(11)	0.26(0.02)
	Pseudocharacter	Identity	280(7)	290(7)	336(10)	455(20)	456(18)	0.15(0.02)
		Pseudocharacter	277(6)	283(7)	329(9)	442(16)	439(16)	0.17(0.02)
Constituent 2 ($n+2$)	Identity	Identity	227(4)	225(4)	245(6)	305(10)	326(10)	0.30(0.02)
		Pseudocharacter	237(4)	235(4)	261(6)	345(13)	349(11)	0.25(0.02)
	Pseudocharacter	Identity	228(5)	225(5)	249(6)	334(12)	348(13)	0.26(0.02)
		Pseudocharacter	231(5)	229(5)	252(6)	343(12)	345(14)	0.25(0.02)
The whole region	Identity	Identity	233(4)	235(5)	370(12)	472(18)	559(22)	0.05(0.01)
		Pseudocharacter	247(5)	261(6)	418(14)	540(21)	621(24)	0.03(0.01)
	Pseudocharacter	Identity	277(6)	292(8)	473(16)	673(30)	726(31)	0.02(0.01)
		Pseudocharacter	275(6)	286(7)	472(17)	661(26)	716(31)	0.01(0.00)

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

approximately 30 min.

Results and discussion

Participants answered the comprehension questions correctly (96 % accuracy) and were unaware of any display changes and/or unable to report which characters had changed. The eye movement data were preprocessed with fixations longer than 80 ms and shorter than 1200 ms entering the analyses. Furthermore, trials were removed if fewer than three fixations were made (0.9 %); or if the display change occurred too early (3.2 %) or late (7.0 %); or the display change was triggered by hooks, incorrect saccades or blinks (5.7 %). Finally, any observations for each measure that were above or below three standard deviations from each participant's mean were excluded prior to conducting the analyses (1.5 %, the average number of observations removed across all measures and regions).

Analyses were conducted for the pretarget region (n, the preboundary two characters), the first constituent ($n+1$), the second constituent ($n+2$), and the whole target string whereby $n+1$ and $n+2$ comprised a single region. For each region we computed eye movement measures including: *first fixation duration* (FFD, the duration of the first fixation on a region regardless of whether it is fixated more than once during first-pass reading), *single fixation duration* (SFD, the duration of a fixation when it is the only one made on a region during first-pass reading), *gaze duration* (GD, the sum of all first-pass fixations on a region before making a saccade to another region), *go-past time* (the sum of all fixations from first entering a region until moving to the right of the region including those fixations made after regressions to any earlier regions), *total fixation duration* (TFD, the sum of all fixations on a region), and *skipping probability* (SP, the probability of a region being not fixated during first-pass reading). The means and standard errors for all these measures across all the regions are shown in Table 2.

To analyze data, Linear mixed models (LMMs) were conducted using the lme4 package (version 1.1–27) in R (version 4.0.5, R Development Core Team, 2021). Each preview and their interaction were treated as fixed factors. Participants and items were included as crossed random factors. Models were run starting with the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), then trimmed down if they failed to converge (details of the model trimming procedure, R scripts for statistical analyses and data are available at https://osf.io/6wuhs/). The fixation time analyses were analyzed on log-transformed data to increase normality whilst the skipping data were conducted using logistic GLMMs given the binary nature of the variable. When there was an interaction between both previews, a separate analysis was run, and two contrasts were set up to test for simple effects and examine the $n+2$ preview effect when the $n+1$ was an identity (contrast 1) or a pseudocharacter preview (contrast 2). Fixed effect estimations for all eye movement measures across all regions are displayed in Table 3.

The pretarget region (n)

The $n+1$ preview effect was significant in TFD with longer total fixations on the pretarget region when the $n+1$ preview was a pseudocharacter than when it was an identity. This indicates a relatively late influence of the orthographic properties of the $n+1$ preview at word n, and this is an effect that is consistent with the effect reported by Zang et al., (2021,2023). The interaction between $n+1$ and $n+2$ previews was also reliable in TFD, however the planned contrasts did not show any reliable $n+2$ preview effect when the $n+1$ preview was an identity or a pseudocharacter. Unexpectedly, there was a reversed $n+2$ preview effect in Go-past time with longer time on the pretarget region when the $n+2$ preview was an identity compared with a pseudocharacter. Note that, Go-past time includes the time spent re-reading earlier regions as well as the pretarget region itself, and therefore it is likely that this $n+2$ effect reflects processing associated with integration of the pre-target word with sentential context that occurred after it was initially processed. Thus, it appears that integration of word n in relation to constituent n + 2 was more difficult when constituent $n+1$ was unavailable compared with when it was available. None of the other effects were reliable.

The first constituent (n + 1)

The $n+1$ preview effect was significant in all eye movement measures with longer time and less skipping when the $n+1$ preview was a pseudocharacter than an identity (all t or $|z| > 10.18$), replicating the standard preview effects reported in the literature (Rayner, 1998; 2009). The $n+2$ preview effect was also significant in all eye movement measures with longer time and less skipping of $n+1$ when the $n+2$ preview was a pseudocharacter than an identity preview (all t or $|z| > 2.41$). Interestingly, the $n+2$ preview interacted with the $n+1$ preview across all measures (all $|t|$ or $z > 2.07$, see Fig. 2). The planned contrasts showed a robust $n+2$ preview effect only when the $n+1$ preview was an identity (all t or $|z| > 2.03$, though this effect only approached significance in Go-past time, $t = 1.89$) rather than a pseudocharacter (all $|t|$

Table 3
LMM analyses for all regions in Experiment 1.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
The pretarget region (*n*)	(Intercept)	5.34	0.02	**330.71**	**<2e-16**	5.34	0.02	**334.60**	**<2e-16**	5.40	0.02	**268.66**	**<2e-16**
	n + 1 preview	0.00	0.01	0.10	0.92	0.00	0.01	0.62	0.54	0.01	0.01	1.16	0.25
	n + 2 preview	−0.01	0.01	−1.20	0.23	−0.01	0.01	−1.00	0.32	−0.01	0.01	−1.08	0.28
	Interaction	−0.03	0.02	−1.80	0.07	−0.02	0.02	−0.99	0.33	−0.03	0.02	−1.65	0.10
Constituent 1 (*n* + 1)	(Intercept)	5.39	0.02	**308.70**	**<0.001**	5.38	0.02	**293.91**	**<0.001**	5.45	0.02	**240.61**	**<0.001**
	n + 1 preview	0.17	0.02	**10.19**	**<0.001**	0.21	0.02	**12.43**	**<0.001**	0.27	0.02	**14.08**	**<0.001**
	n + 2 preview	0.04	0.01	**2.80**	**0.01**	0.04	0.01	**2.69**	**0.01**	0.05	0.02	**3.09**	**0.00**
	Interaction	−0.04	0.02	**−2.08**	**0.04**	−0.04	0.02	**−2.34**	**0.02**	−0.06	0.02	**−3.22**	**0.00**
	Contrast 1	0.03	0.01	**2.43**	**0.02**	0.03	0.01	**2.10**	**0.04**	0.04	0.02	**2.38**	**0.02**
	Contrast 2	−0.01	0.01	−0.48	0.63	−0.01	0.01	−0.53	0.60	−0.02	0.01	−1.42	0.16
Constituent 2 (*n* + 2)	(Intercept)	5.37	0.02	**297.07**	**<0.001**	5.36	0.02	**292.93**	**<0.001**	5.43	0.02	**258.17**	**<0.001**
	n + 1 preview	0.00	0.01	−0.12	0.91	−0.01	0.01	−0.67	0.51	0.00	0.02	0.03	0.98
	n + 2 preview	0.04	0.01	**3.10**	**0.00**	0.04	0.01	**2.87**	**0.00**	0.05	0.01	**3.40**	**0.00**
	Interaction	−0.03	0.02	−1.57	0.12	−0.03	0.02	−1.42	0.16	−0.04	0.02	−1.86	0.06
	Contrast 1	–	–	–	–	–	–	–	–	0.05	0.01	**3.41**	**0.00**
	Contrast 2	–	–	–	–	–	–	–	–	0.01	0.01	0.80	0.43
The whole region	(Intercept)	5.39	0.02	**309.07**	**<0.001**	5.41	0.02	**289.21**	**<0.001**	5.77	0.03	**189.32**	**<0.001**
	n + 1 preview	0.16	0.01	**11.39**	**<0.001**	0.18	0.02	**9.62**	**<0.001**	0.25	0.02	**16.25**	**<0.001**
	n + 2 preview	0.06	0.01	**5.19**	**<0.001**	0.08	0.02	**4.70**	**<0.001**	0.13	0.02	**8.11**	**<0.001**
	Interaction	−0.06	0.02	**−3.81**	**<0.001**	−0.08	0.03	**−3.20**	**0.00**	−0.13	0.02	**−6.14**	**<0.001**
	Contrast 1	0.05	0.01	**4.78**	**<0.001**	0.09	0.02	**5.20**	**<0.001**	0.13	0.02	**8.12**	**<0.001**
	Contrast 2	0.00	0.01	−0.18	0.86	0.00	0.02	0.02	0.98	0.00	0.02	−0.18	0.85

		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
The pretarget region (*n*)	(Intercept)	5.56	0.03	**201.87**	**<2e-16**	5.72	0.03	**173.02**	**< 2e-16**	−1.10	0.11	**−9.81**	**<2e-16**
	n + 1 preview	0.01	0.01	0.60	0.55	0.09	0.01	**7.36**	**0.00**	−0.06	0.08	−0.79	0.43
	n + 2 preview	−0.03	0.01	**−2.37**	**0.02**	0.00	0.01	−0.15	0.88	0.03	0.08	0.44	0.66
	Interaction	−0.03	0.03	−1.28	0.20	−0.06	0.02	**−2.26**	**0.02**	−0.03	0.11	−0.30	0.77
	Contrast 1	–	–	**–**	**–**	0.03	0.02	1.35	0.18	–	–	–	–
	Contrast 2	–	–	**–**	**–**	−0.03	0.02	−1.57	0.12	–	–	**–**	**–**
Constituent 1 (*n* + 1)	(Intercept)	5.59	0.03	**202.86**	**<0.001**	5.68	0.03	**183.56**	**<0.001**	−1.04	0.12	**−8.90**	**<0.001**
	n + 1 preview	0.34	0.03	**13.20**	**<0.001**	0.28	0.02	**13.20**	**<0.001**	−1.27	0.11	**−11.26**	**<0.001**
	n + 2 preview	0.05	0.02	**2.69**	**0.01**	0.05	0.02	**2.80**	**0.01**	−0.20	0.08	**−2.42**	**0.02**
	Interaction	−0.07	0.03	**−2.61**	**0.01**	−0.07	0.02	**−3.04**	**0.00**	0.41	0.13	**3.14**	**0.00**
	Contrast 1	0.04	0.02	1.89	0.06	0.04	0.02	**2.04**	**0.04**	−0.23	0.09	**−2.59**	**0.01**
	Contrast 2	−0.02	0.02	−1.07	0.28	−0.02	0.02	−1.47	0.14	0.16	0.09	1.79	0.07
Constituent 2 (*n* + 2)	(Intercept)	5.56	0.03	**194.64**	**<0.001**	5.64	0.03	**184.35**	**<0.001**	−0.99	0.10	**−9.86**	**<0.001**
	n + 1 preview	0.05	0.02	**2.41**	**0.02**	0.05	0.02	**2.45**	**0.01**	−0.20	0.08	**−2.51**	**0.01**
	n + 2 preview	0.09	0.02	**4.24**	**<0.001**	0.06	0.02	**3.38**	**0.00**	−0.26	0.08	**−3.04**	**0.00**
	Interaction	−0.07	0.03	**−2.35**	**0.02**	−0.08	0.03	**−2.90**	**0.00**	0.21	0.12	1.86	0.06
	Contrast 1	0.09	0.02	**4.19**	**<0.001**	0.06	0.02	**3.40**	**0.00**	−0.26	0.08	**−3.23**	**0.00**
	Contrast 2	0.02	0.02	0.94	0.35	−0.01	0.02	−0.68	0.50	−0.04	0.08	−0.53	0.59
The whole region	(Intercept)	5.96	0.04	**164.06**	**<0.001**	6.15	0.04	**154.49**	**<0.001**	−4.15	0.31	**−13.32**	**<0.001**
	n + 1 preview	0.36	0.02	**15.56**	**<0.001**	0.27	0.02	**14.71**	**<0.001**	−2.52	0.62	**−4.08**	**0.00**
	n + 2 preview	0.14	0.02	**8.27**	**<0.001**	0.12	0.02	**7.50**	**<0.001**	−0.60	0.19	**−3.08**	**0.00**
	Interaction	−0.14	0.02	**−5.93**	**<0.001**	−0.13	0.02	**−5.88**	**<0.001**	0.31	0.36	0.86	0.39
	Contrast 1	0.14	0.02	**7.57**	**<0.001**	0.12	0.02	**7.32**	**<0.001**	–	–	–	–
	Contrast 2	0.00	0.02	−0.10	0.92	−0.01	0.02	−0.55	0.59	–	–	–	–

Note. Significant terms featured in bold, and terms approaching significance are underlined. *b* = regression coefficient. Contrast 1 refers to the *n* + 2 preview effect when the *n* + 1 preview was an identity, and Contrast 2 refers to the *n* + 2 preview effect when the *n* + 1 preview was a pseudocharacter.

or $z < 1.80$). These results perfectly replicate the findings from Cutter et al. (2014) and Zang et al. (2021), and demonstrate that the availability of constituent *n* + 1 in the parafovea licensed preprocessing of *n* + 2 to a significant degree resulting in reliable *n* + 2 preview effects. By contrast, when constituent *n* + 1 was not available in the parafovea, preview effects were substantially diminished. These findings provide evidence that frequently used four-character idioms are processed as a whole in Chinese reading.

The second constituent (n + 2)

For the second constituent analyses, there was a reliable *n* + 2 preview effect in all measures (all *t* or $|z| > 2.86$), and a reliable *n* + 1 preview effect in TFD, Go-past time and SP (all $|t|$ or $|z| > 2.40$). Furthermore, the interaction between the *n* + 1 and *n* + 2 previews was significant in TFD and Go-past time (all $|t| > 2.34$), and this effect approached significance in GD and SP ($|t|$ or $z = 1.86$). The planned contrasts showed a significant *n* + 2 preview effect with longer fixations

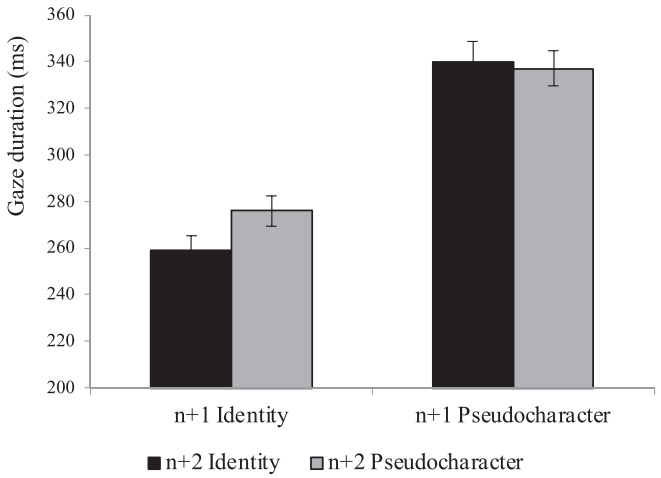

Fig. 2. $n + 2$ preview effects for GD on the first constituents when the $n + 1$ preview was an identity or a pseudocharacter in Experiment 1 (Error bars represent standard errors of the mean).

Table 4
The stroke number of each constituent character of the identity and its counterpart pseudocharacter preview in Experiment 2.

	C/P 1	C/P 2	C/P 3	C/P 4	C/P 1 + 2	C/P 3 + 4
Identity	7.3 (3.1)	8.5 (3.2)	7.7 (3.2)	7.6 (2.6)	15.8 (4.5)	15.3 (3.8)
Pseudocharacter	7.3 (3.1)	8.5 (3.2)	7.7 (3.2)	7.6 (2.6)	15.8 (4.5)	15.2 (3.7)

Standard deviations are provided in parentheses. C = Character; P = Pseudocharacter.

and less skipping for pseudocharacter than identity previews, but this effect was only reliable when $n + 1$ was an identity (all t or $|z| > 3.22$) rather than a pseudocharacter (all $|t|$ or $|z| < 0.95$). These results are consistent with the findings from the first constituent analyses.

The whole region

The whole region includes both the first ($n + 1$) and the second ($n + 2$) constituents. The pattern of results for this whole region was the same as that for the first constituent analyses, and the effects were robust across all the fixation time measures, though not skipping. This is not surprising as any four-character string is a relatively long segment of text in Chinese, and therefore, it is less likely to be skipped than shorter (e.g., two-character) regions. Importantly, there were reliable effects of $n + 1$ preview, $n + 2$ preview and an interaction in all fixation time measures (all $|t| > 3.19$) with significant $n + 2$ preview effects appearing only when the $n + 1$ was parafoveally available (all $t > 4.77$). Again, these results are entirely consistent with the analyses from the individual regions.

Overall, the results of Experiment 1 were consistent with those from Cutter et al (2014) and Zang et al (2021), indicating that the presence of the first constituent of a four character Chinese idiom licenses processing of the second, and that these highly familiar four-character idioms appear to be parafoveally processed as MCUs. To extend the findings from Experiment 1, Experiment 2 was conducted to examine whether highly familiar four-character phrases are processed parafoveally as single representations during reading.

Experiment 2

Method

Participants

One hundred and nine native Chinese-speaking students (90 females, mean age 21 years) from Tianjin Normal University participated in the experiment. They had normal or corrected-to-normal vision and were naive as to the purpose of the experiment.

Apparatus

The same apparatus was used as in Experiment 1.

Materials and design

Fifty-six four-character target phrases were used that were the same as those used in Experiment 2 of He et al. (2021). These strings consisted of two two-character words that were most often categorized as single four-character words in a prescreen test (see He et al., 2021 for details). Each target phrase was inserted into a sentence ranging from 16 to 22 characters in length. Note, though, 22 sentences originally used in He et al's study were edited slightly in the current experiment to avoid the pretarget word in those sentences being the very high frequency function word de "的" that is often skipped during Chinese reading which could have affected pre-processing of the target string (Zang et al., 2018). The same pre-screen procedure was carried out in Experiment 2 as in Experiment 1. All sentences were pre-screened for their naturalness, and predictability of the whole target strings given the preceding context and predictability of the second constituent given the preceding context including the first constituent. Specifically, the mean sentence naturalness was 1.4 (SD = 0.2, N = 20), and the target phrases were unpredictable (0.6 %, SD = 2.8 %, N = 20) given the preceding sentential context. The predictability of the second constituent given the prior context including the first constituent was 28 % (SD = 29 %, N = 20), though, as in Experiment 1, further analyses showed that it did not affect any findings of the study (see Table A3 in the Appendix). Finally, the conditional probability of the second constituent of phrases given the first was 18 % (SD = 22 %) and this did not influence the findings either (Table A4 in the Appendix).

As in Experiment 1, the boundary paradigm was used to orthogonally manipulate the preview of each constituent ($n + 1$, the first two-character word; and $n + 2$, the second two-character word) of the target phrase to be either an identity or a pseudocharacter preview. Each constituent character of the identity (phrase) and its counterpart pseudocharacter preview was controlled for the number of strokes (Table 4), all $|t| < 1.01$. An example set of sentences under different preview conditions is illustrated in Fig. 3.

Procedure

The basic procedure was the same as in Experiment 1, however, each participant in Experiment 2 read 90 sentences (56 experimental sentences, 28 filler sentences without display changes and 6 practice sentences) from one of the four files with conditions rotated across these files.

Results and discussion

Participants fully understood the sentences in Experiment 2 with 95 % mean comprehension accuracy, and they were unaware of any display changes and/or unable to report which characters had changed. The same data exclusion criteria as in Experiment 1 were used. Trials were removed if fewer than three fixations were made (0.7 %); or if the display change occurred too early (3.7 %) or late (8.9 %); or the display change was triggered by hooks, incorrect saccades or blinks (6.6 %). Finally, any observations for each measure that were above or below three standard deviations from each participant's mean were excluded prior to conducting the analyses (1.4 %, the average number of

n+1 preview	*n*+2 preview	Sentence
Identity	Identity	张雨桐认为|**安全系数**达到百分之二十才算合格。
	Pseudocharacter	张雨桐认为|**而芏系数**达到百分之二十才算合格。
Pseudocharacter	Identity	张雨桐认为|**安全亜葨**达到百分之二十才算合格。
	Pseudocharacter	张雨桐认为|**而芏亜葨**达到百分之二十才算合格。

Fig. 3. An example of the stimuli under different preview conditions used in Experiment 2. The preview of the first and the second constituent of the target phrase was manipulated with an invisible boundary (represented by the vertical line) placed preceding the target. Before the eyes crossed the boundary, the preview appeared at the position of the target string. Once the eyes crossed the boundary, the preview changed to the correct target (both the target and the preview are in bold for illustrative purposes, but were presented normally in the experiment). The English translation for the sentence is "Zhang Yutong believes that a **safety factor** of 20 percent is required to be considered qualified".

Table 5
Eye movement measures for all regions in Experiment 2.

Region	$n+1$ preview	$n+2$ preview	FFD	SFD	GD	Go-past	TFD	SP
The pretarget region (*n*)	Identity	Identity	220(4)	220(4)	241(5)	319(10)	348(10)	0.26(0.02)
		Pseudocharacter	224(4)	222(4)	243(5)	318(11)	361(12)	0.27(0.02)
	Pseudocharacter	Identity	217(4)	215(4)	242(5)	309(9)	373(11)	0.25(0.02)
		Pseudocharacter	219(4)	216(4)	242(5)	309(10)	392(15)	0.26(0.02)
Constituent 1 ($n+1$)	Identity	Identity	236(4)	237(5)	259(6)	321(10)	356(12)	0.21(0.02)
		Pseudocharacter	248(5)	251(6)	276(7)	344(11)	374(12)	0.18(0.02)
	Pseudocharacter	Identity	284(6)	295(7)	340(9)	430(13)	446(16)	0.13(0.02)
		Pseudocharacter	280(6)	291(7)	337(8)	446(15)	457(17)	0.12(0.01)
Constituent 2 ($n+2$)	Identity	Identity	228(4)	227(4)	250(6)	301(11)	340(12)	0.26(0.02)
		Pseudocharacter	239(4)	238(5)	270(6)	321(11)	374(14)	0.23(0.02)
	Pseudocharacter	Identity	234(5)	231(5)	258(6)	322(10)	345(13)	0.23(0.02)
		Pseudocharacter	235(5)	231(5)	259(7)	343(13)	354(14)	0.22(0.02)
The whole region	Identity	Identity	235(4)	232(5)	396(13)	498(16)	610(22)	0.02(0.01)
		Pseudocharacter	250(5)	261(8)	443(13)	552(18)	664(25)	0.02(0.00)
	Pseudocharacter	Identity	281(6)	292(8)	495(15)	663(22)	730(28)	0.01(0.00)
		Pseudocharacter	277(6)	291(9)	493(15)	688(27)	756(30)	0.01(0.00)

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

observations removed across all measures and regions). The same eye movement measures as in Experiment 1 were computed for the pretarget region (*n*, the preboundary two characters), the first constituent ($n+1$, the first two-character word), the second constituent ($n+2$, the second two-character word), and the whole target string whereby $n+1$ and $n+2$ comprised a single region. The means and standard deviations across all the regions are shown in Table 5. The fixed and random effects structure of LMMs and the trimming procedure was the same as in Experiment 1. Fixed effect estimations for all the eye movement measures across all regions are shown in Table 6.

The pretarget region (n)

The $n+1$ preview effect was significant in TFD with longer total fixations on the pretarget region when the $n+1$ preview was a pseudocharacter compared to an identity. This result replicates the findings from Experiment 1, as well as Zang et al., (2021,2023), suggesting a late orthographic influence of the $n+1$ preview on total times for word *n*. None of the other effects were reliable.

The first constituent (n + 1)

The $n+1$ and $n+2$ preview effects were significant in all eye movement measures with longer times and less skipping for pseudocharacter than identity previews on the first constituent (all t or $|z| > 2.12$). Importantly, the $n+2$ preview interacted with the $n+1$ preview across all fixation time measures (all $|t| > 1.99$, though as in Experiment 1, the interaction only approached significance in Go-past time, $|t| = 1.81$). The planned contrasts showed that this was due to a robust $n+2$ preview effect when $n+1$ was an identity (all $|t| > 3.17$, see Fig. 4) rather than a pseudocharacter (all $|t| < 1.62$). Again, these results align perfectly with those from Experiment 1 and the previous findings of Cutter et al. (2014) and Zang et al., (2021,2023), and indicate that the second constituent of the frequently used four-character phrases was processed to a significant and greater degree when the first constituent was parafoveally available compared with when it was not. Note, also that this effect occurred when the second constituent was positioned three to four characters away from the current fixation. These results provide further strong evidence for the MCU hypothesis.

The second constituent (n + 2)

For the second constituent analyses, there was a reliable $n+2$ preview effect in all fixation time measures (all $t > 2.67$), and a reliable $n+1$ preview effect in Go-past time and SP (all $|t|$ or $|z| > 2.31$). Furthermore, the interaction between the $n+1$ and $n+2$ previews was significant in GD and TFD (all $|t| > 2.03$) and approached significance in FFD ($|t| = 1.89$) and SFD ($|t| = 1.93$). The planned contrasts showed a significant $n+2$ preview effect with longer fixations for pseudocharacter than identity previews, but this effect was only reliable when $n+1$ was an identity (all $t > 2.67$) rather than a pseudocharacter preview (all $t < 1$). It appears that the effects observed for the first constituent analyses carried over to the second constituent, again, providing evidence for the MCU hypothesis.

Table 6
LMM analyses for all regions in Experiment 2.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
The pretarget region (*n*)	(Intercept)	5.34	0.02	**348.90**	**<2e-16**	5.41	0.02	**298.09**	**<2e-16**	5.34	0.02	**344.61**	**<2e-16**
	$n+1$ preview	−0.01	0.01	−1.59	0.11	0.00	0.01	−0.20	0.84	−0.01	0.01	−1.69	0.09
	$n+2$ preview	0.01	0.01	0.95	0.35	0.00	0.01	−0.30	0.77	0.00	0.01	0.12	0.91
	Interaction	−0.01	0.02	−0.66	0.51	−0.01	0.02	−0.28	0.78	−0.01	0.02	−0.52	0.60
Constituent 1 ($n+1$)	(Intercept)	5.40	0.02	**287.69**	**<0.001**	5.40	0.02	**275.38**	**<0.001**	5.47	0.02	**241.57**	**<0.001**
	$n+1$ preview	0.19	0.01	**12.43**	**<0.001**	0.22	0.02	**13.19**	**<0.001**	0.28	0.02	**15.89**	**<0.001**
	$n+2$ preview	0.05	0.01	**3.90**	**<0.001**	0.06	0.01	**4.39**	**<0.001**	0.07	0.02	**4.38**	**<0.001**
	Interaction	−0.07	0.02	**−3.95**	**<0.001**	−0.07	0.02	**−3.52**	**<0.001**	−0.08	0.02	**−3.80**	**<0.001**
	Contrast 1	0.05	0.01	**3.61**	**<0.001**	0.06	0.01	**4.19**	**<0.001**	0.07	0.02	**4.12**	**<0.001**
	Contrast 2	−0.02	0.01	−1.61	0.11	−0.01	0.02	−0.57	0.57	−0.01	0.02	−0.65	0.52
Constituent 2 ($n+2$)	(Intercept)	5.37	0.02	**273.03**	**<0.001**	5.37	0.02	**277.80**	**<0.001**	5.44	0.02	**225.76**	**<0.001**
	$n+1$ preview	0.01	0.01	0.99	0.32	0.01	0.01	0.51	0.61	0.02	0.02	0.92	0.36
	$n+2$ preview	0.04	0.01	**2.86**	**0.00**	0.04	0.01	**2.68**	**0.01**	0.06	0.02	**3.72**	**<0.001**
	Interaction	−0.04	0.02	−1.89	0.06	−0.04	0.02	−1.93	0.05	−0.06	0.02	**−2.45**	**0.01**
	Contrast 1	0.04	0.01	**2.86**	**0.00**	0.04	0.01	**2.68**	**0.01**	0.06	0.02	**3.79**	**<0.001**
	Contrast 2	0.00	0.01	0.22	0.83	0.00	0.01	0.00	1.00	0.01	0.02	0.31	0.76
The whole region	(Intercept)	5.40	0.02	**304.86**	**<0.001**	5.39	0.02	**248.81**	**<0.001**	5.84	0.03	**181.48**	**<0.001**
	$n+1$ preview	0.17	0.01	**11.51**	**<0.001**	0.19	0.02	**8.09**	**<0.001**	0.24	0.02	**13.12**	**<0.001**
	$n+2$ preview	0.06	0.01	**4.82**	**<0.001**	0.11	0.02	**5.35**	**<0.001**	0.13	0.02	**7.17**	**<0.001**
	Interaction	−0.08	0.02	**−4.49**	**<0.001**	−0.11	0.03	**−3.51**	**<0.001**	−0.13	0.02	**−5.37**	**<0.001**
	Contrast 1	0.06	0.01	**4.43**	**<0.001**	0.12	0.02	**5.61**	**<0.001**	0.13	0.02	**7.22**	**<0.001**
	Contrast 2	−0.02	0.01	−1.42	0.16	0.00	0.02	0.08	0.93	0.00	0.02	−0.19	0.85

		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
The pretarget region (*n*)	(Intercept)	5.59	0.03	**198.18**	**<2e-16**	5.73	0.03	**178.20**	**<2e-16**	−1.19	0.12	**−9.95**	**<2e-16**
	$n+1$ preview	−0.02	0.01	−1.44	0.15	0.06	0.01	**3.98**	**0.00**	−0.13	0.09	−1.44	0.15
	$n+2$ preview	−0.01	0.02	−0.51	0.61	0.02	0.02	1.34	0.18	−0.02	0.09	−0.27	0.79
	Interaction	0.00	0.03	0.04	0.97	0.02	0.03	0.56	0.57	0.09	0.13	0.73	0.47
Constituent 1 ($n+1$)	(Intercept)	5.62	0.03	**205.47**	**<0.001**	5.71	0.03	**185.79**	**<0.001**	−1.62	0.14	**−11.81**	**<0.001**
	$n+1$ preview	0.32	0.02	**13.56**	**<0.001**	0.25	0.02	**11.83**	**<0.001**	−0.97	0.15	**−6.57**	**0.00**
	$n+2$ preview	0.07	0.02	**3.24**	**0.00**	0.07	0.02	**3.39**	**0.00**	−0.23	0.11	**−2.13**	**0.03**
	Interaction	−0.05	0.03	−1.81	0.07	−0.05	0.03	**−2.00**	**0.05**	0.12	0.17	0.69	0.49
	Contrast 1	0.07	0.02	**3.18**	**0.00**	0.06	0.02	**3.29**	**0.00**	–	–	–	–
	Contrast 2	0.02	0.02	0.98	0.33	0.02	0.02	0.80	0.42	–	–	–	–
Constituent 2 ($n+2$)	(Intercept)	5.55	0.03	**185.65**	**<0.001**	5.67	0.03	**170.89**	**<0.001**	−1.21	0.12	**−10.39**	**<0.001**
	$n+1$ preview	0.06	0.02	**2.64**	**0.01**	0.00	0.02	0.12	0.90	−0.25	0.11	**−2.32**	**0.02**
	$n+2$ preview	0.06	0.02	**2.82**	**0.01**	0.08	0.02	**3.76**	**0.00**	−0.20	0.10	−1.93	0.05
	Interaction	−0.04	0.03	−1.15	0.25	−0.06	0.03	**−2.04**	**0.04**	0.19	0.14	1.34	0.18
	Contrast 1	–	–	–	–	0.08	0.02	**3.67**	**<0.001**	–	–	–	–
	Contrast 2	–	–	–	–	0.02	0.02	0.99	0.32	–	–	–	–
The whole region	(Intercept)	6.03	0.04	**167.44**	**<0.001**	6.24	0.04	**159.32**	**<0.001**	−5.78	0.63	**−9.12**	**<0.001**
	$n+1$ preview	0.33	0.02	**14.44**	**<0.001**	0.21	0.02	**11.12**	**<0.001**	−0.70	0.34	**−2.03**	**0.04**
	$n+2$ preview	0.13	0.02	**6.59**	**<0.001**	0.10	0.02	**5.75**	**<0.001**	0.21	0.71	0.30	0.77
	Interaction	−0.10	0.03	**−4.04**	**<0.001**	−0.06	0.02	**−2.77**	**0.01**	0.12	0.52	0.22	0.82
	Contrast 1	0.13	0.02	**6.64**	**<0.001**	0.10	0.02	**5.92**	**<0.001**	–	–	–	–
	Contrast 2	0.02	0.02	1.23	0.22	0.04	0.02	**2.14**	**0.03**	–	–	–	–

Note. Significant terms featured in bold, and terms approaching significance are underlined. b = regression coefficient. Contrast 1 refers to the $n+2$ preview effect when the $n+1$ preview was an identity, and Contrast 2 refers to the $n+2$ preview effect when the $n+1$ preview was a pseudocharacter.

The whole region

As in Experiment 1, the whole region includes both the first ($n+1$) and the second ($n+2$) constituents. Again, the pattern of results for the whole region was the same as that for the first constituent analyses, and effects were robust across all the fixation time measures but not skipping. Effects for skipping were weak very likely due to the low probability of skipping the four-character strings. Importantly, there were reliable effects of $n+1$ preview, $n+2$ preview and an interaction in all fixation time measures (all $|t| > 2.77$). The $n+2$ preview effect was more robust when a preview of $n+1$ was parafoveally available (all $t >$ 4.42) compared with when it was not (all $t < 1.43$). That the effect achieved significance in a late measure (TFD) as well as in the early measures, suggests that MCU effects continued developing cumulatively across first pass and second pass reading, $t = 2.14$).

To reiterate, in general, the results from both Experiments 1 and 2 pattern very similarly to previous research (Cutter et al., 2014; Zang et al., 2021, 2023), and both experiments produced greater $n+2$ preview benefit when the $n+1$ was an identity compared with when it was a pseudocharacter preview. These results strongly suggest during Chinese reading of highly familiar four-character phrases, the first

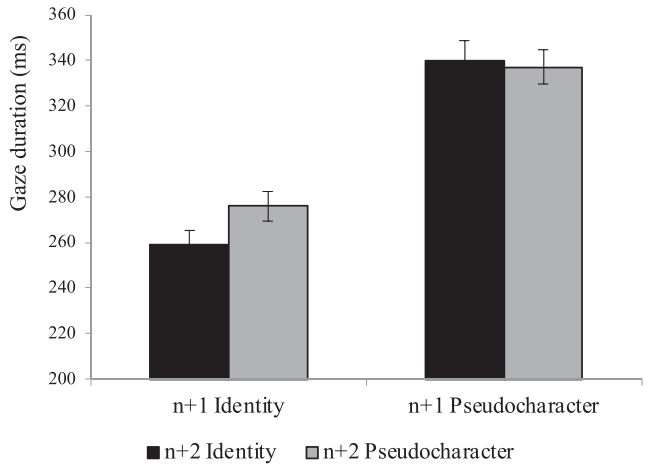

Fig. 4. $n + 2$ preview effects for GD on the first constituents when the $n + 1$ preview was an identity or a pseudocharacter in Experiment 2 (Error bars represent standard errors of the mean).

Table 7
Statistical properties for the target strings in Experiment 3.

	One-word string	Ambiguous string	Two-word string
No. of strokes	15.0(3.9)	15.6(3.7)	15.8(3.5)
Frequency (per million)	91(127)	79(123)	84(125)
Character 3	7.6(3.1)	7.8(2.9)	7.2(2.5)
Pseudocharacter 3	7.6(3.1)	7.8(2.9)	7.2(2.5)
Character 4	7.6(3.1)	7.7(2.7)	8.5(2.6)
Pseudocharacter 4	7.4(2.7)	7.7(2.7)	8.5(2.7)

constituent licenses processing of the second because these phrases are lexicalized, being represented as single units and are processed as MCUs parafoveally. To further consolidate this claim, Experiment 3 was conducted to examine whether the lexical status of four-character strings (i.e., whether they are, or are not, a MCU) modulates how they are processed.

Experiment 3

Method

Participants

One hundred and two native Chinese-speaking students (88 females, mean age 22 years) from Tianjin Normal University participated in the experiment. They had normal or corrected-to-normal vision and were naive as to the purpose of the experiment.

Apparatus

The same apparatus was used as in Experiment 1.

Materials and design

Forty-eight four-character target string triplets were selected from Experiment 2 of He et al. (2021) that were always comprised of two two-character words. However, in an off-line pre-screen procedure (see He et al., 2021), the first set of strings were predominantly categorized as being a single four-character word (i.e., they were a subset of the 56 stimuli used in the present Experiment 2; these strings were termed one-word strings as per He et al.). The second set of stimuli were unambiguously categorized as two separate two-character words (two-word strings), and the third set were categorized approximately equally often as a single four-character word and as two separate two-character words (ambiguous strings, for details of the pre-screen procedure see He et al., 2021). Each set of four-character strings shared the initial two-character word (i.e., the first constituent, $n + 1$), and the second two-character words of each triplet (i.e., the second constituent, $n + 2$) were matched for frequency and stroke complexity.

As in Experiment 2, 16 sentences originally used in He et al's study were edited slightly to avoid using the very high frequency function word de "的" as a pretarget. As a consequence, the second constituents in one ambiguous string and four two-word strings were replaced in order for each triplet string to fit into the same sentence frame. In order to ensure the reliability of segmentation preferences for these (slightly adapted) stimuli, we undertook the same offline pencil and paper task as conducted by He et al. Twenty participants were required to indicate the word boundaries within 48 sets of four-character target strings. The first set of the triplets were judged most often to be one-word strings (M = 86 %, SD = 5 %), the second set as two-word strings (M = 98 %, SD = 6 %), and for the third set, approximately half the participants segmented them as one-word strings (M = 58 %, SD = 9 %), whilst the other half segmented them as two-word strings. The segmentation proportion for each category was almost identical to those obtained by He et al.

Each triplet target string was embedded in a sentence frame ranging from 16 to 25 characters in length with identical sentence context prior to the target string. All sentences were pre-screened for their naturalness on a 5-point scale (1 = very natural) by 18 participants and the mean sentence naturalness for the one-word string was 1.4 (SD = 0.2), the two-word string was 1.7 (SD = 0.2) and the ambiguous string was 1.6 (SD = 0.3). Although all sentences across the three types of strings were almost equally natural to read, the statistical analysis showed a difference among the three types ($F = 18.7p < .05$). Therefore, it was subsequently statistically controlled as a covariate in a set of LMM analyses of the eye movement data. The pattern of these results was identical to those reported here indicating that naturalness did not contribute to our effects (see Table A5 in the Appendix).

As in Experiments 1 and 2, predictability of the whole target strings given the preceding context and predictability of the second constituent given the preceding context including the first constituent were also assessed. The target strings were unpredictable given the preceding sentential context (for the one-word strings, M = 0.7 %, SD = 3 %; for the two-word strings and the ambiguous strings, M = 0 %; N = 20). The predictability of the second constituent given the prior context including the first constituent was 26 % (SD = 27 %) for the one-word strings, and 0 % for the other two types of strings (N = 20). Similarly, the conditional probability of the second constituent of phrases given the first for the one-word strings (M = 16.8 %, SD = 20 %) was higher than that for the two-word strings (M = 0 %) and the ambiguous strings (M = 0.6 %, SD = 1.2 %, $F = 31$). Again, further LMM analyses showed that the two variables exerted no robust influence on the results of Experiment 3 (Table A6 and A7 in the Appendix).

The boundary paradigm was used to orthogonally manipulate the preview of the second constituent ($n + 2$, the second two-character word) of the target string to be either a pseudocharacter or the identity. The second constituents across the three types of strings were controlled for the number of strokes ($F = 1.3$) and word frequency ($F = 0.11$, Table 7), and each constituent character of the identity (the third and fourth character of the four-character target string) and its counterpart pseudocharacter preview were controlled for the number of strokes, all $|t| < 1.01$. An example set of sentences under the different experimental conditions is illustrated in Fig. 5.

Procedure

The basic procedure was the same as in Experiments 1 and 2, however, each participant in Experiment 3 read 78 sentences (48 experimental sentences, 24 filler sentences without display changes and 6 practice sentences) from one of the six files with conditions rotated across these files.

Preview type	String type	sentence
Identity	One-word string	张雨桐认为|**安全系数**达到百分之二十才算合格。
	Ambiguous string	张雨桐认为|**安全需求**得到满足才是最重要的。
	Two-word string	张雨桐认为|**安全就是**要高于一切并要始终高于一切。
Pseudocharacter	One-word string	张雨桐认为|**安全亜詎**达到百分之二十才算合格。
	Ambiguous string	张雨桐认为|**安全簽壱**得到满足才是最重要的。
	Two-word string	张雨桐认为|**安全雏羡**要高于一切并要始终高于一切。

Fig. 5. An example of the stimuli under the different experimental conditions used in Experiment 3. The first constituent was identical across the three types of strings and the preview of the second constituent was manipulated with an invisible boundary (represented by the vertical line) positioned prior to the target. Before the eyes crossed the boundary, the preview appeared at the position of the target string. Once the eyes crossed the boundary, the preview changed to the correct target (the target strings and previews are in bold for illustration, but were presented normally in the experiment). The English translation for the sentence is "Zhang Yutong believes that a **safety factor** of 20 percent is required to be considered qualified/ Zhang Yutong believes that having **safety requirement** met is the most important thing/ Zhang Yutong believes that **safety is** above all else and should be always above all else".

Table 8
Eye movement measures for all regions in Experiment 3.

Region	String Type	Preview Type	FFD	SFD	GD	Go-past	TFD	SP
The pretarget region (*n*)	1-Word String	Identity	220(4)	221(4)	242(5)	340(13)	363(14)	0.28(0.02)
		Pseudocharacter	217(4)	217(4)	238(5)	347(14)	371(15)	0.29(0.02)
	Ambiguous String	Identity	220(4)	221(5)	245(6)	339(13)	363(12)	0.28(0.02)
		Pseudocharacter	215(5)	215(5)	237(6)	336(13)	386(18)	0.29(0.02)
	2-Word String	Identity	221(4)	221(4)	243(5)	338(13)	382(15)	0.29(0.02)
		Pseudocharacter	223(4)	222(4)	253(6)	344(13)	396(15)	0.30(0.02)
Constituent 1 ($n + 1$)	1-Word String	Identity	222(4)	221(4)	246(5)	318(12)	371(16)	0.23(0.02)
		Pseudocharacter	235(5)	237(5)	272(6)	370(16)	404(15)	0.20(0.02)
	Ambiguous String	Identity	235(5)	238(6)	265(7)	338(12)	429(18)	0.23(0.02)
		Pseudocharacter	242(5)	244(6)	284(7)	366(14)	436(16)	0.19(0.02)
	2-Word String	Identity	238(5)	241(6)	280(8)	356(13)	434(19)	0.22(0.02)
		Pseudocharacter	241(5)	241(5)	273(7)	381(17)	446(18)	0.24(0.02)
Constituent 2 ($n + 2$)	1-Word String	Identity	226(4)	227(5)	252(7)	318(12)	363(14)	0.23(0.02)
		Pseudocharacter	226(5)	229(5)	253(6)	325(11)	359(14)	0.24(0.02)
	Ambiguous String	Identity	242(5)	239(5)	279(6)	390(17)	445(18)	0.20(0.02)
		Pseudocharacter	242(5)	243(5)	281(6)	386(15)	429(17)	0.19(0.02)
	2-Word String	Identity	238(5)	237(6)	276(9)	401(17)	446(18)	0.23(0.02)
		Pseudocharacter	242(5)	239(6)	278(7)	412(17)	472(16)	0.17(0.02)
The whole region	1-Word String	Identity	222(4)	217(5)	395(12)	516(19)	668(29)	0.04(0.01)
		Pseudocharacter	235(5)	245(7)	423(14)	571(21)	690(27)	0.03(0.01)
	Ambiguous String	Identity	236(5)	244(8)	470(17)	612(23)	818(35)	0.04(0.01)
		Pseudocharacter	242(5)	250(11)	491(17)	645(23)	820(34)	0.03(0.01)
	2-Word String	Identity	236(5)	240(10)	448(15)	631(29)	835(37)	0.05(0.01)
		Pseudocharacter	239(4)	227(6)	470(16)	664(26)	875(36)	0.04(0.01)

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

Results and discussion

Participants fully understood the sentences with 94 % mean comprehension accuracy. Participants were unaware of any display changes and/or unable to report which characters had changed. The same data exclusion criteria as in Experiment 1 were used. Trials were removed if fewer than three fixations were made (0.1 %); or if the display change occurred too early (3.3 %) or late (8.2 %); or the display change was triggered by hooks, incorrect saccades or blinks (12.1 %). Finally, any observations for each measure that were above or below three standard deviations from each participant's mean were excluded prior to conducting the analyses (1.4 %, the average number of observations removed across all measures and regions). The same eye movement measures as in Experiments 1 and 2 were reported for the pretarget region (*n*, the preboundary two characters), the first constituent ($n + 1$, the first two-character word), the second constituent ($n + 2$, the second two-character word), and the whole target string whereby $n + 1$ and $n + 2$ comprised a single region. The means and standard deviations across all the regions are shown in Table 8. In the LMM analyses, String Type, Preview and their interaction were treated as fixed factors. For the string type, contrasts were carried out, with comparisons of ambiguous strings vs one-word strings, two-word strings vs ambiguous strings, two-word strings vs one-word strings. Participants and items were treated as crossed random factors. The trimming procedure was the same as in Experiments 1 and 2. Fixed effect estimations for all the eye movement measures across all regions are shown in Table 9

The pretarget region (n)

There was solely a significant difference between one-word and two-

Table 9
LMM analyses for all regions in Experiment 3.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
The pretarget region (*n*)	Intercept	5.34	0.01	**379.24**	**<0.001**	5.34	0.01	**378.68**	**<0.001**	5.42	0.02	**316.01**	**<0.001**
	Preview	−0.01	0.01	−0.68	0.49	−0.01	0.01	−1.30	0.20	0.00	0.01	−0.37	0.71
	Ambiguous vs. 1-word	−0.01	0.01	−0.85	0.39	−0.01	0.01	−0.87	0.39	0.00	0.02	0.09	0.93
	2-word vs. ambiguous	0.02	0.01	1.71	0.09	0.02	0.01	1.45	0.15	0.02	0.02	1.58	0.12
	2-word vs. 1-word	0.01	0.01	0.86	0.39	0.01	0.01	0.58	0.56	0.03	0.02	1.66	0.10
	Preview × ambiguous vs. 1-word	−0.01	0.02	−0.31	0.76	−0.01	0.03	−0.27	0.79	−0.01	0.03	−0.25	0.81
	Preview × 2-word vs. ambiguous	0.01	0.02	0.50	0.62	0.01	0.03	0.51	0.61	0.04	0.03	1.30	0.19
	Preview × 2-word vs. 1-word	0.00	0.02	0.19	0.85	0.01	0.03	0.24	0.81	0.03	0.03	1.06	0.29
Constituent 1 ($n+1$)	Intercept	5.41	0.02	**340.24**	**<0.001**	5.42	0.02	**331.33**	**<0.001**	5.52	0.02	**275.24**	**<0.001**
	Preview	0.03	0.01	**2.21**	**0.03**	0.03	0.01	**2.38**	**0.02**	0.05	0.01	**3.42**	**0.00**
	Ambiguous vs. 1-word	0.04	0.01	**3.31**	**0.00**	0.05	0.01	**3.68**	**<0.001**	0.06	0.02	**3.68**	**<0.001**
	2-word vs. ambiguous	0.00	0.01	−0.07	0.95	−0.01	0.01	−0.86	0.39	0.00	0.02	0.10	0.92
	2-word vs. 1-word	0.04	0.01	**3.22**	**0.00**	0.04	0.01	**2.77**	**0.01**	0.06	0.02	**3.75**	**<0.001**
	Preview × ambiguous vs. 1-word	0.00	0.03	0.17	0.87	−0.02	0.03	−0.58	0.56	−0.04	0.03	−1.19	0.24
	Preview × 2-word vs. ambiguous	−0.03	0.03	−1.24	0.22	−0.04	0.03	−1.36	0.17	−0.07	0.03	**−2.10**	**0.04**
	Preview × 2-word vs. 1-word	−0.03	0.03	−1.07	0.29	−0.06	0.03	−1.94	0.05	−0.11	0.03	**−3.27**	**0.00**
Constituent 2 ($n+2$)	Intercept	5.41	0.01	**382.20**	**<0.001**	5.40	0.01	**370.46**	**< 2e-16**	5.51	0.02	**318.92**	**<0.001**
	Preview	0.00	0.01	0.24	0.81	0.01	0.01	0.77	0.44	0.01	0.01	0.74	0.46
	Ambiguous vs. 1-word	0.06	0.01	**4.63**	**<0.001**	0.06	0.02	**4.08**	**0.00**	0.10	0.02	**5.84**	**<0.001**
	2-word vs. ambiguous	−0.01	0.01	−0.57	0.57	−0.02	0.02	−1.04	0.30	−0.01	0.02	−0.74	0.46
	2-word vs. 1-word	0.06	0.01	**4.05**	**<0.001**	0.04	0.02	**2.57**	**0.01**	0.09	0.02	**5.06**	**<0.001**
	Preview × ambiguous vs. 1-word	0.01	0.03	0.26	0.80	0.02	0.03	0.59	0.56	0.02	0.03	0.63	0.53
	Preview × 2-word vs. ambiguous	0.00	0.03	0.02	0.98	−0.01	0.03	−0.27	0.79	−0.02	0.03	−0.67	0.50
	Preview × 2-word vs. 1-word	0.01	0.03	0.28	0.78	0.01	0.03	0.31	0.75	0.00	0.03	−0.04	0.97
The whole region	Intercept	5.41	0.02	**357.27**	**<0.001**	5.39	0.02	**286.05**	**<0.001**	5.95	0.03	**194.81**	**<0.001**
	Preview	0.02	0.01	**2.05**	**0.04**	0.03	0.02	1.60	0.11	0.06	0.02	**3.71**	**<0.001**
	Ambiguous vs. 1-word	0.04	0.01	**3.56**	**<0.001**	0.04	0.02	1.78	0.08	0.14	0.02	**7.25**	**<0.001**
	2-word vs. ambiguous	−0.01	0.01	−0.63	0.53	−0.03	0.02	−1.18	0.24	−0.04	0.02	**−2.25**	**0.03**
	2-word vs. 1-word	0.04	0.01	**2.91**	**0.00**	0.01	0.02	0.59	0.56	0.10	0.02	**4.95**	**<0.001**
	Preview × ambiguous vs. 1-word	−0.01	0.02	−0.50	0.62	−0.07	0.04	−1.69	0.09	−0.02	0.04	−0.60	0.55
	Preview × 2-word vs. ambiguous	−0.02	0.02	−1.01	0.31	−0.04	0.04	−0.82	0.41	−0.02	0.04	−0.54	0.59
	Preview × 2-word vs. 1-word	−0.04	0.02	−1.50	0.13	−0.11	0.04	**−2.56**	**0.01**	−0.05	0.04	−1.13	0.26

		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
The pretarget region (*n*)	Intercept	5.65	0.03	**182.79**	**<0.001**	5.75	0.04	**161.17**	**<0.001**	−1.08	0.11	**−9.72**	**<2e-16**
	Preview	0.01	0.02	0.65	0.51	0.02	0.02	1.37	0.17	0.04	0.07	0.65	0.52
	Ambiguous vs. 1-word	−0.01	0.02	−0.24	0.81	0.01	0.02	0.65	0.52	0.04	0.08	0.43	0.67
	2-word vs. ambiguous	0.00	0.02	−0.04	0.97	0.03	0.02	1.56	0.12	0.03	0.08	0.37	0.71
	2-word vs. 1-word	−0.01	0.02	−0.28	0.78	0.04	0.02	**2.20**	**0.03**	0.07	0.08	0.80	0.43
	Preview × ambiguous vs. 1-word	−0.01	0.04	−0.34	0.74	0.04	0.04	0.95	0.34	0.02	0.17	0.09	0.93
	Preview × 2-word vs. ambiguous	0.04	0.04	0.94	0.35	−0.01	0.04	−0.24	0.81	0.00	0.17	0.02	0.99
	Preview × 2-word vs. 1-word	0.03	0.04	0.60	0.55	0.03	0.04	0.71	0.48	0.02	0.17	0.11	0.91
Constituent 1 ($n+1$)	Intercept	5.71	0.03	**204.38**	**<0.001**	5.87	0.04	**167.07**	**<0.001**	−1.56	0.12	**−12.95**	**<0.001**
	Preview	0.09	0.02	**5.07**	**<0.001**	0.05	0.02	**3.20**	**0.00**	−0.16	0.08	−1.94	0.05
	Ambiguous vs. 1-word	0.05	0.02	**2.38**	**0.02**	0.12	0.02	**5.79**	**<0.001**	−0.05	0.10	−0.53	0.60
	2-word vs. ambiguous	0.03	0.02	1.40	0.16	0.01	0.02	0.72	0.47	0.14	0.10	1.42	0.16
	2-word vs. 1-word	0.08	0.02	**3.76**	**<0.001**	0.13	0.02	**6.48**	**<0.001**	0.09	0.10	0.88	0.38
	Preview × ambiguous vs. 1-word	−0.08	0.04	−1.86	0.06	−0.07	0.04	−1.73	0.08	0.02	0.20	0.07	0.94

(*continued on next page*)

Table 9 (*continued*)

Region	Effect	FFD				SFD				GD			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
	Preview × 2-word vs. ambiguous	−0.01	0.04	−0.25	0.81	0.02	0.04	0.42	0.67	0.30	0.20	1.51	0.13
	Preview × 2-word vs. 1-word	−0.09	0.04	**−2.09**	**0.04**	−0.05	0.04	−1.31	0.19	0.32	0.20	1.58	0.11
Constituent 2 ($n + 2$)	Intercept	5.72	0.03	**216.53**	**<0.001**	5.87	0.03	**186.28**	**<0.001**	−1.54	0.10	**−14.78**	**< 2e-16**
	Preview	0.02	0.02	1.01	0.31	0.00	0.02	0.19	0.85	−0.11	0.10	−1.11	0.27
	Ambiguous vs. 1-word	0.15	0.02	**6.63**	**<0.001**	0.18	0.02	**8.99**	**<0.001**	−0.27	0.10	**−2.77**	**0.01**
	2-word vs. ambiguous	0.04	0.02	1.65	0.10	0.05	0.02	**2.42**	**0.02**	0.03	0.10	0.32	0.75
	2-word vs. 1-word	0.19	0.02	**8.21**	**<0.001**	0.23	0.02	**11.31**	**<0.001**	−0.24	0.10	**−2.44**	**0.01**
	Preview × ambiguous vs. 1-word	0.00	0.05	−0.11	0.92	−0.02	0.04	−0.52	0.60	−0.16	0.20	−0.83	0.41
	Preview × 2-word vs. ambiguous	−0.01	0.05	−0.32	0.75	0.11	0.04	**2.62**	**0.01**	−0.27	0.20	−1.34	0.18
	Preview × 2-word vs. 1-word	−0.02	0.05	−0.42	0.68	0.09	0.04	**2.06**	**0.04**	−0.44	0.20	**−2.20**	**0.03**
The whole region	Intercept	6.23	0.04	**168.13**	**<0.001**	6.50	0.04	**164.02**	**<0.001**	−4.27	0.25	**−16.93**	**<0.001**
	Preview	0.09	0.02	**5.18**	**<0.001**	0.04	0.02	**2.45**	**0.02**	−0.21	0.18	−1.19	0.23
	Ambiguous vs. 1-word	0.15	0.02	**7.27**	**<0.001**	0.20	0.02	**11.31**	**<0.001**	0.12	0.23	0.50	0.61
	2-word vs. ambiguous	0.01	0.02	0.29	0.77	0.03	0.02	1.86	0.06	0.28	0.21	1.35	0.18
	2-word vs. 1-word	0.16	0.02	**7.50**	**<0.001**	0.23	0.02	**13.10**	**<0.001**	0.40	0.22	1.83	0.07
	Preview × ambiguous vs. 1-word	−0.07	0.04	−1.67	0.10	−0.03	0.04	−0.75	0.45	−0.06	0.46	−0.12	0.90
	Preview × 2-word vs. ambiguous	0.00	0.04	0.04	0.97	0.02	0.04	0.52	0.61	0.13	0.42	0.31	0.75
	Preview × 2-word vs. 1-word	−0.07	0.04	−1.62	0.11	−0.01	0.04	−0.23	0.82	0.08	0.43	0.18	0.86

Note. Significant terms featured in bold, and terms approaching significance are underlined. b = regression coefficient.

word strings in TFD. Readers spent longer processing pretargets when two-word strings were presented in the parafovea compared with one-word strings. Given that TFD includes the time spent rereading this region and is therefore a relatively late measure of processing (probably reflecting integrative rather than initial recognition processes), this result probably indicates that two-word strings take longer to process than the one-word strings (see He et al., 2021) and there may be some sensitivity to such processing difficulty based on parafoveal preview prior to direct fixation. The fact that no other effects were reliable presumably indicates that the number of characters between the pretarget region and the preview of the second constituent of the target string was sufficient to prevent any influence from it.

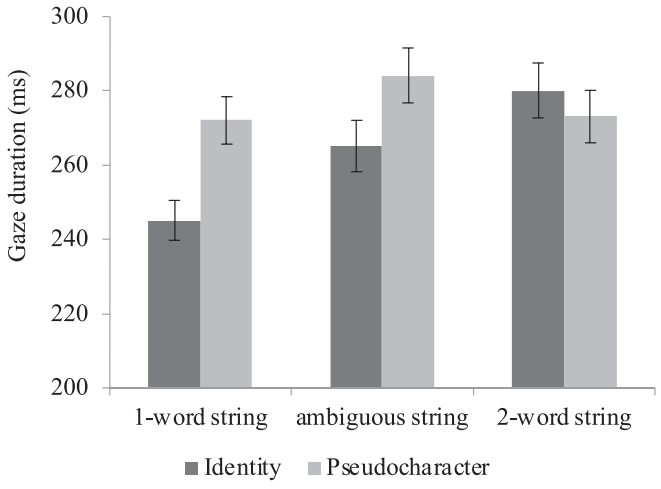

Fig. 6. $n + 2$ preview effects for three types of strings on the first constituents in GD in Experiment 2 (Error bars represent standard errors of the mean).

The first constituent ($n + 1$)

In the first constituent analyses, all reading times were shorter for one-word strings compared to ambiguous (all $t > 2.37$) and two-word strings (all $t > 2.76$), demonstrating a processing advantage for one-word strings relative to the latter two strings. This finding is consistent with that of He et al., (2021). There was no difference between the latter two strings (all $t < 1.41$). There was a robust preview effect in all fixation time measures such that readers spent longer (all $t > 2.20$) and skipped the first constituent numerically less often ($|z| = 1.94$) when the second constituent preview was a pseudocharacter compared with when it was an identity. Importantly, the preview effect interacted with the type of string in fixation time measures including SFD, GD, Go-past and TFD (note, though, the interaction between preview type and the contrast for the ambiguous versus 1-word string only approached significance in Go-past and TFD, for details see Table 9 and Fig. 6). The planned contrasts showed that the preview effect for the second constituent was greatest for the one-word strings (SFD: $b = 0.06$, $SE = 0.02$, $t = 2.74$; GD: $b = 0.10$, $SE = 0.02$, $t = 4.21$; Go-past time: $b = 0.15$, $SE = 0.03$, $t = 4.74$; TFD: $b = 0.09$, $SE = 0.03$, $t = 3.22$), intermediate for the ambiguous strings (SFD: $b = 0.04$, $SE = 0.02$, $t = 1.83$, $p = .07$; GD: $b = 0.06$, $SE = 0.02$, $t = 2.64$; Go-past time: $b = 0.07$, $SE = 0.03$, $t = 2.19$; TFD: $b = 0.02$, $SE = 0.03$, $t = 0.88$) and least for the two-word strings (all $t < 1.77$). These results are entirely in line with our predictions. The findings demonstrate that readers preprocess the second constituent of a four-character string in the parafovea to the greatest extent when the second constituent along with the first form a single lexical unit, numerically less when they are treated ambiguously, that is by some readers as a single word and by others as two separate words, and least when the target string is unambiguously treated as two separate words. In other words, the lexical status of multi-character strings substantially influences how they are processed.

The second constituent ($n + 2$)

Similar to the first constituent analyses, readers spent less time and skipped the second constituents more often for one-word strings

Table A1
Fixed effect estimations for all the eye movement measures when predictability of the second constituents of idioms given the preceding context including the first constituents was included as a covariate in Experiment 1.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.39	0.02	**308.67**	**<0.001**	5.38	0.02	**293.83**	**<0.001**	5.45	0.02	**240.74**	**<0.001**
	$n+1$ preview	0.17	0.02	**10.21**	**<0.001**	0.21	0.02	**12.46**	**<0.001**	0.27	0.02	**14.10**	**<0.001**
	$n+2$ preview	0.04	0.01	**2.80**	**0.01**	0.04	0.01	**2.70**	**0.01**	0.05	0.02	**3.10**	**0.00**
	Interaction	−0.04	0.02	**−2.09**	**0.04**	−0.04	0.02	**−2.35**	**0.02**	−0.07	0.02	**−3.23**	**0.00**
	Contrast1	0.03	0.01	**2.43**	**0.02**	0.03	0.01	**2.10**	**0.04**	0.04	0.02	**2.38**	**0.02**
	Contrast2	−0.01	0.01	−0.48	0.63	−0.01	0.01	−0.53	0.60	−0.02	0.01	−1.42	0.16
	Prediction	0.03	0.02	1.19	0.24	0.03	0.03	1.10	0.27	0.05	0.04	1.43	0.15
Constituent 2 ($n+2$)	(Intercept)	5.37	0.02	**297.18**	**<0.001**	5.36	0.02	**293.22**	**<0.001**	5.43	0.02	**258.41**	**<0.001**
	$n+1$ preview	0.00	0.01	−0.12	0.91	−0.01	0.01	−0.67	0.50	0.00	0.02	0.03	0.98
	$n+2$ preview	0.04	0.01	**3.11**	**0.00**	0.04	0.01	**2.88**	**0.00**	0.05	0.01	**3.41**	**0.00**
	Interaction	−0.03	0.02	−1.57	0.12	−0.03	0.02	−1.42	0.16	−0.04	0.02	−1.86	0.06
	Contrast1	–	–	–	–	–	–	–	–	0.05	0.01	**3.41**	**0.00**
	Contrast2	–	–	–	–	–	–	–	–	0.01	0.01	0.80	0.43
	Prediction	−0.03	0.03	−1.08	0.28	−0.03	0.03	−1.16	0.25	−0.05	0.04	−1.31	0.19
The whole region	(Intercept)	5.39	0.02	**309.00**	**<0.001**	5.40	0.02	**301.50**	**<0.001**	5.77	0.03	**195.80**	**<0.001**
	$n+1$ preview	0.16	0.01	**11.39**	**<0.001**	0.18	0.02	**8.81**	**<0.001**	0.25	0.02	**13.25**	**<0.001**
	$n+2$ preview	0.06	0.01	**5.18**	**<0.001**	0.08	0.02	**4.88**	**<0.001**	0.13	0.02	**7.09**	**<0.001**
	Interaction	−0.06	0.02	**−3.80**	**<0.001**	−0.09	0.03	**−3.11**	**0.00**	−0.13	0.02	**−5.43**	**<0.001**
	Contrast1	0.05	0.01	**4.78**	**<0.001**	0.09	0.02	**5.20**	**<0.001**	0.13	0.02	**8.12**	**<0.001**
	Contrast2	0.00	0.01	−0.18	0.86	0.00	0.02	0.02	0.98	0.00	0.02	−0.18	0.85
	Prediction	0.02	0.02	0.95	0.34	−0.06	0.03	**−1.97**	**0.05**	−0.05	0.04	−1.43	0.15
		Go-past				TFD				SP			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.59	0.03	**203.75**	**<0.001**	5.68	0.03	**183.68**	**<0.001**	−1.07	0.13	**−8.48**	**<0.001**
	$n+1$ preview	0.34	0.03	**13.23**	**<0.001**	0.28	0.02	**13.21**	**<0.001**	−1.08	0.09	**−11.40**	**<0.001**
	$n+2$ preview	0.05	0.02	**2.70**	**0.01**	0.05	0.02	**2.80**	**0.01**	−0.23	0.09	**−2.58**	**0.01**
	Interaction	−0.07	0.03	**−2.62**	**0.01**	−0.07	0.02	**−3.05**	**0.00**	0.39	0.13	**2.97**	**0.00**
	Contrast1	0.04	0.02	1.89	0.06	0.04	0.02	**2.04**	**0.04**	−0.23	0.09	**−2.59**	**0.01**
	Contrast2	−0.02	0.02	−1.07	0.28	−0.02	0.02	−1.47	0.14	0.16	0.09	1.79	0.07
	Prediction	0.11	0.05	**2.04**	**0.04**	0.08	0.06	1.23	0.22	−0.33	0.22	−1.49	0.14
Constituent 2 ($n+2$)	(Intercept)	5.57	0.03	**194.88**	**<0.001**	5.64	0.03	**185.01**	**<0.001**	−0.99	0.10	**−9.92**	**<0.001**
	$n+1$ preview	0.05	0.02	**2.41**	**0.02**	0.05	0.02	**2.45**	**0.01**	−0.21	0.08	**−2.43**	**0.02**
	$n+2$ preview	0.09	0.02	**4.26**	**<0.001**	0.06	0.02	**3.40**	**0.00**	−0.25	0.08	**−3.00**	**0.00**
	Interaction	−0.07	0.03	**−2.35**	**0.02**	−0.08	0.03	**−2.91**	**0.00**	0.21	0.12	1.79	0.07
	Contrast1	0.09	0.02	**4.19**	**<0.001**	0.06	0.02	**3.40**	**0.00**	−0.26	0.08	**−3.23**	**0.00**
	Contrast2	0.02	0.02	0.94	0.35	−0.01	0.02	−0.68	0.50	−0.04	0.08	−0.53	0.59
	Prediction	−0.07	0.05	−1.35	0.18	−0.11	0.05	**−2.06**	**0.04**	0.62	0.19	**3.30**	**0.00**
The whole region	(Intercept)	5.96	0.04	**163.90**	**< 2e-16**	6.15	0.04	**154.41**	**<0.001**	−4.41	0.33	**−13.27**	**<0.001**
	$n+1$ preview	0.36	0.02	**15.56**	**< 2e-16**	0.27	0.02	**14.71**	**<0.001**	−1.19	0.23	**−5.15**	**0.00**
	$n+2$ preview	0.14	0.02	**8.27**	**< 2e-16**	0.12	0.02	**7.50**	**<0.001**	−0.61	0.20	**−3.09**	**0.00**
	Interaction	−0.14	0.02	**−5.93**	**0.00**	−0.13	0.02	**−5.88**	**<0.001**	0.33	0.35	0.95	0.34
	Contrast1	0.14	0.02	**7.57**	**<0.001**	0.12	0.02	**7.32**	**<0.001**	–	–	–	–
	Contrast2	0.00	0.02	−0.10	0.92	−0.01	0.02	−0.55	0.59	–	–	–	–
	Prediction	0.01	0.06	0.12	0.90	−0.04	0.06	−0.69	0.49	0.48	0.35	1.35	0.18

Note. Significant terms featured in bold, and terms approaching significance are underlined. b = regression coefficient. Contrast 1 refers to the $n+2$ preview effect when the $n+1$ preview was an identity, and Contrast 2 refers to the $n+2$ preview effect when the $n+1$ preview was a pseudocharacter.

compared with ambiguous and two-word strings (all t or $|z| > 2.43$). There were no differences between the latter two strings in all measures (all $|t| < 1.66$) except the TFD ($t = 2.42$) with longer total reading times for the two-word strings than the ambiguous strings. Again, these results replicate He et al's findings. There was also an interaction between preview and the target string with a significant preview effect on the second constituent for the two-word strings in SP ($b = 0.33$, $SE = 0.14$, $z = 2.31$) and TFD ($b = 0.06$, $SE = 0.03$, $t = 2.19$), but not for the one-word and ambiguous strings (all t or $z < 1.33$). These results were not predicted but presumably occurred because the second constituent was processed to some extent for the one-word and ambiguous strings prior to their fixation, whereas the $n+2$ preview for the two-word string was not processed to the same degree. Therefore, effects associated with $n+2$ preview only appear later when readers decide whether the second constituent should be skipped or fixated.

The whole region

Again, readers spent less time fixating one-word strings than ambiguous and two-word strings (all $t > 2.90$ in all reading time measures except in SFD), replicating findings from He et al.'s study. There

Table A2
Fixed effect estimations for all the eye movement measures when conditional probability of idioms was included as a covariate in Experiment 1.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.39	0.02	**299.62**	**< 2e-16**	5.38	0.02	**294.67**	**< 2e-16**	5.43	0.03	**212.57**	**< 2e-16**
	$n+1$ preview	0.18	0.02	**11.79**	**< 2e-16**	0.21	0.02	**12.47**	**< 2e-16**	0.27	0.02	**14.10**	**< 2e-16**
	$n+2$ preview	0.04	0.01	**3.07**	**0.00**	0.04	0.01	**2.71**	**0.01**	0.05	0.02	**3.10**	**0.00**
	Interaction	−0.04	0.02	**−2.55**	**0.01**	−0.04	0.02	**−2.37**	**0.02**	−0.07	0.02	**−3.24**	**0.00**
	Contrast1	0.03	0.01	**2.43**	**0.02**	0.03	0.01	**2.10**	**0.04**	0.04	0.02	**2.38**	**0.02**
	Contrast2	−0.01	0.01	−0.48	0.63	−0.01	0.01	−0.53	0.60	−0.02	0.01	−1.42	0.16
	CP	0.10	0.04	**2.25**	**0.03**	0.09	0.04	**2.04**	**0.04**	0.12	0.06	1.90	0.06
Constituent 2 ($n+2$)	(Intercept)	5.37	0.02	**291.25**	**< 2e-16**	5.37	0.02	**292.96**	**< 2e-16**	5.43	0.02	**258.20**	**< 2e-16**
	$n+1$ preview	0.00	0.01	−0.12	0.90	−0.01	0.01	−0.67	0.51	0.00	0.02	0.02	0.98
	$n+2$ preview	0.04	0.01	**3.05**	**0.00**	0.04	0.01	**2.88**	**0.00**	0.05	0.01	**3.40**	**0.00**
	Interaction	−0.03	0.02	−1.52	0.13	−0.03	0.02	−1.42	0.15	−0.04	0.02	−1.86	0.06
	Contrast1	–	–	–	–	–	–	–	–	0.05	0.01	**3.41**	**0.00**
	Contrast2	–	–	–	–	–	–	–	–	0.01	0.01	0.80	0.43
	CP	−0.03	0.05	−0.63	0.53	−0.03	0.05	−0.66	0.51	−0.06	0.06	−0.96	0.34
The whole region	(Intercept)	5.39	0.02	**309.39**	**< 2e-16**	5.40	0.02	**300.65**	**< 2e-16**	5.77	0.03	**189.27**	**< 2e-16**
	$n+1$ preview	0.16	0.01	**11.39**	**< 2e-16**	0.18	0.02	**8.77**	**0.00**	0.25	0.02	**16.25**	**< 2e-16**
	$n+2$ preview	0.06	0.01	**5.19**	**0.00**	0.08	0.02	**4.86**	**0.00**	0.13	0.02	**8.11**	**0.00**
	Interaction	−0.06	0.02	**−3.81**	**0.00**	−0.09	0.03	**−3.11**	**0.00**	−0.13	0.02	**−6.14**	**0.00**
	Contrast1	0.05	0.01	**4.78**	**0.00**	0.09	0.02	**5.20**	**0.00**	0.13	0.02	**8.12**	**0.00**
	Contrast2	0.00	0.01	−0.18	0.86	0.00	0.02	0.02	0.98	0.00	0.02	−0.19	0.85
	CP	0.05	0.04	1.43	0.16	0.00	0.05	−0.02	0.99	0.06	0.07	0.82	0.41

		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.59	0.03	**202.97**	**< 2e-16**	5.68	0.03	**179.19**	**<2e-16**	−1.07	0.13	**−8.48**	**< 2e-16**
	$n+1$ preview	0.34	0.03	**13.22**	**< 2e-16**	0.28	0.02	11.24	**<2e-16**	−1.08	0.09	**−11.37**	**< 2e-16**
	$n+2$ preview	0.05	0.02	**2.70**	**0.01**	0.05	0.02	**2.37**	**0.02**	−0.23	0.09	**−2.59**	**0.01**
	Interaction	−0.07	0.03	**−2.61**	**0.01**	−0.07	0.03	**2.60**	**0.01**	0.39	0.13	**2.96**	**0.00**
	Contrast1	0.04	0.02	1.89	0.06	0.04	0.02	**2.04**	**0.04**	−0.03	0.01	**−2.61**	**0.01**
	Contrast2	−0.02	0.02	−1.07	0.28	−0.02	0.02	−1.47	0.14	0.02	0.01	1.72	0.08
	CP	0.11	0.10	1.11	0.27	0.06	0.11	0.52	0.61	−0.42	0.40	−1.05	0.30
Constituent 2 ($n+2$)	(Intercept)	5.57	0.03	**195.48**	**< 0.001**	5.64	0.03	**185.58**	**< 2e-16**	−1.04	0.12	**−8.66**	**< 2e-16**
	$n+1$ preview	0.05	0.02	**2.40**	**0.02**	0.05	0.02	**2.46**	**0.01**	−0.20	0.08	**−2.52**	**0.01**
	$n+2$ preview	0.09	0.02	**4.25**	**< 0.001**	0.06	0.02	**3.39**	**0.00**	−0.26	0.08	**−3.04**	**0.00**
	Interaction	−0.07	0.03	**−2.35**	**0.02**	−0.08	0.03	**−2.91**	**0.00**	0.21	0.12	1.86	0.06
	Contrast1	0.09	0.02	**4.19**	**0.00**	0.06	0.02	**3.40**	**0.00**	−0.05	0.01	**−3.24**	**0.00**
	Contrast2	0.02	0.02	0.94	0.35	−0.01	0.02	−0.68	0.50	−0.01	0.01	−0.50	0.62
	CP	−0.17	0.10	−1.82	0.07	−0.24	0.09	**−2.54**	**0.01**	0.26	0.35	0.74	0.46
The whole region	(Intercept)	5.96	0.04	**163.91**	**< 2e-16**	6.15	0.04	**154.52**	**< 2e-16**	−4.15	0.31	**−13.33**	**< 2e-16**
	$n+1$ preview	0.36	0.02	**15.56**	**< 2e-16**	0.27	0.02	**14.71**	**< 2e-16**	−2.51	0.62	**−4.08**	**0.00**
	$n+2$ preview	0.14	0.02	**8.27**	**< 2e-16**	0.12	0.02	**7.51**	**0.00**	−0.60	0.19	**−3.07**	**0.00**
	Interaction	−0.14	0.02	**−5.93**	**0.00**	−0.13	0.02	**−5.88**	**0.00**	0.31	0.36	0.85	0.39
	Contrast1	0.14	0.02	**7.57**	**0.00**	0.12	0.02	**7.32**	**0.00**	–	–	–	–
	Contrast2	0.00	0.02	−0.10	0.92	−0.01	0.02	−0.55	0.59	–	–	–	–
	CP	0.01	0.11	0.12	0.91	−0.13	0.11	−1.17	0.25	−0.19	0.65	−0.29	0.77

Note. Significant terms featured in bold, and terms approaching significance are underlined. *b* = regression coefficient. CP = Conditional Probability. Contrast 1 refers to the $n+2$ preview effect when the $n+1$ preview was an identity, and Contrast 2 refers to the $n+2$ preview effect when the $n+1$ preview was a pseudocharacter.

were generally no differences between the latter two strings (all $t < 1.18$, though there were slightly longer total reading times for the two-word than the ambiguous strings and an opposite pattern in GD). A possible explanation for this result is that forming a decision as to whether an ambiguous string is a single word, or two, takes time during the first pass reading, and of course, performing lexical identification twice as compared with just once takes more time.

Relative to the identity preview, readers spent more time when they received a pseudocharacter preview (all $t > 2.05$ in all fixation time measures except SFD). Finally, there was an interaction between the preview and the contrast for the 2-word versus 1-word string in SFD, and planned contrasts showed a robust preview benefit for the one-word string ($b = 0.09$, $SE = 0.03$, $t = 3.20$) rather than the two-word strings (all $t < 0.61$). Thus, as in the preceding experiments, readers pre-processed the second constituent to a greater extent for the one-word strings compared to the two-word strings. It is perhaps not surprising, given that commonly used four-character strings are often treated or categorised as single words, that parafoveal processing of these strings, similar to the processing of words, tends to occur to a greater extent than for multiple, separate, words (Juhasz, Pollatsek, Hyönä, Drieghe &

Table A3
Fixed effect estimations for all the eye movement measures when predictability of the second constituents of phrases given the preceding context including the first constituents was included as a covariate in Experiment 2.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.40	0.02	**287.31**	**<0.001**	5.40	0.02	**274.96**	**<0.001**	5.47	0.02	**241.30**	**<0.001**
	$n+1$ preview	0.19	0.01	**12.43**	**<0.001**	0.22	0.02	**13.19**	**<0.001**	0.28	0.02	**15.89**	**<0.001**
	$n+2$ preview	0.05	0.01	**3.90**	**<0.001**	0.06	0.01	**4.39**	**<0.001**	0.07	0.02	**4.38**	**<0.001**
	Interaction	−0.07	0.02	**−3.95**	**<0.001**	−0.07	0.02	**−3.52**	**<0.001**	−0.08	0.02	**−3.80**	**<0.001**
	Contrast1	0.05	0.01	**3.61**	**<0.001**	0.06	0.01	**4.19**	**<0.001**	0.07	0.02	**4.12**	**<0.001**
	Contrast2	−0.02	0.01	−1.61	0.11	−0.01	0.02	−0.57	0.57	−0.01	0.02	−0.65	0.52
	Prediction	0.00	0.02	−0.07	0.95	0.00	0.03	−0.09	0.93	−0.02	0.03	−0.54	0.59
Constituent 2 ($n+2$)	(Intercept)	5.37	0.02	**288.94**	**<0.001**	5.37	0.02	**275.54**	**<0.001**	5.44	0.02	**230.26**	**<0.001**
	$n+1$ preview	0.01	0.01	0.92	0.36	0.01	0.02	0.51	0.61	0.02	0.02	0.92	0.36
	$n+2$ preview	0.04	0.01	**2.82**	**0.01**	0.04	0.02	**2.51**	**0.01**	0.06	0.02	**3.72**	**<0.001**
	Interaction	−0.04	0.02	−1.87	0.06	−0.04	0.02	−1.73	0.08	−0.06	0.02	**−2.45**	**0.01**
	Contrast1	0.04	0.01	**2.86**	**0.00**	0.04	0.01	**2.68**	**0.01**	0.06	0.02	**3.79**	**<0.001**
	Contrast2	0.00	0.01	0.22	0.83	0.00	0.01	0.00	1.00	0.01	0.02	0.31	0.76
	Prediction	−0.09	0.03	**−2.73**	**0.01**	−0.08	0.03	**−2.44**	**0.02**	−0.13	0.04	**−3.02**	**0.00**
The whole region	(Intercept)	5.40	0.02	**324.47**	**<0.001**	5.40	0.02	**248.37**	**<0.001**	5.84	0.03	**182.52**	**<0.001**
	$n+1$ preview	0.17	0.02	**10.42**	**<0.001**	0.19	0.02	**8.09**	**<0.001**	0.24	0.02	**13.13**	**<0.001**
	$n+2$ preview	0.06	0.01	**4.50**	**0.00**	0.11	0.02	**5.35**	**<0.001**	0.13	0.02	**7.17**	**<0.001**
	Interaction	−0.08	0.02	**−3.99**	**0.00**	−0.11	0.03	**−3.51**	**<0.001**	−0.13	0.02	**−5.37**	**<0.001**
	Contrast1	0.06	0.01	**4.43**	**<0.001**	0.12	0.02	**5.61**	**<0.001**	0.13	0.02	**7.22**	**<0.001**
	Contrast2	−0.02	0.01	−1.42	0.16	0.00	0.02	0.08	0.93	0.00	0.02	−0.19	0.85
	Prediction	−0.01	0.02	−0.63	0.53	0.00	0.03	−0.04	0.97	−0.11	0.05	**−2.01**	**0.04**
		Go-past				TFD				SP			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.61	0.03	**200.91**	**<0.001**	5.71	0.03	**185.98**	**<0.001**	−1.62	0.14	**−11.81**	**< 2e-16**
	$n+1$ preview	0.32	0.02	**16.13**	**<0.001**	0.26	0.02	**11.83**	**<0.001**	−0.97	0.15	**−6.57**	**0.00**
	$n+2$ preview	0.07	0.02	**3.31**	**0.00**	0.07	0.02	**3.39**	**0.00**	−0.23	0.11	**−2.13**	**0.03**
	Interaction	−0.05	0.03	−1.89	0.06	−0.05	0.03	**−2.00**	**0.05**	0.12	0.17	0.68	0.49
	Contrast1	0.07	0.02	**3.18**	**0.00**	0.06	0.02	**3.29**	**0.00**	–	–	–	–
	Contrast2	0.02	0.02	0.98	0.33	0.01	0.02	0.80	0.42	–	–	–	–
	Prediction	−0.03	0.05	−0.72	0.47	−0.06	0.05	−1.19	0.24	−0.05	0.22	−0.21	0.83
Constituent 2 ($n+2$)	(Intercept)	5.55	0.03	**189.33**	**<0.001**	5.67	0.03	**177.33**	**<0.001**	−1.21	0.11	**−10.55**	**<2e-16**
	$n+1$ preview	0.06	0.02	**2.44**	**0.02**	0.00	0.02	0.11	0.92	−0.25	0.11	**−2.31**	**0.02**
	$n+2$ preview	0.06	0.02	**2.85**	**0.00**	0.08	0.02	**3.75**	**<0.001**	−0.20	0.10	−1.92	0.05
	Interaction	−0.03	0.03	−1.13	0.26	−0.06	0.03	**−2.03**	**0.04**	0.19	0.14	1.33	0.18
	Contrast1	–	–	–	–	0.08	0.02	**3.67**	**<0.001**	–	–	–	–
	Contrast2	–	–	–	–	0.02	0.02	0.99	0.32	–	–	–	–
	Prediction	−0.17	0.06	**−2.91**	**0.00**	−0.23	0.06	**−4.03**	**<0.001**	0.54	0.22	**2.47**	**0.01**
The whole region	(Intercept)	6.03	0.04	**168.86**	**<0.001**	6.24	0.04	**162.17**	**<0.001**	−5.22	0.48	**−10.89**	**<0.001**
	$n+1$ preview	0.33	0.02	**14.46**	**<0.001**	0.21	0.02	**11.12**	**<0.001**	−1.15	0.70	−1.64	0.10
	$n+2$ preview	0.13	0.02	**6.59**	**<0.001**	0.10	0.02	**5.76**	**<0.001**	−0.52	0.32	−1.65	0.10
	Interaction	−0.10	0.03	**−4.04**	**<0.001**	−0.06	0.02	**−2.77**	**0.01**	0.05	0.52	0.09	0.93
	Contrast1	0.13	0.02	**6.64**	**<0.001**	0.10	0.02	**5.92**	**<0.001**	–	–	–	–
	Contrast2	0.02	0.02	1.23	0.22	0.04	0.02	**2.14**	**0.03**	—	–	–	–
	Prediction	−0.14	0.06	**−2.21**	**0.03**	−0.20	0.06	**−3.15**	**0.00**	0.52	0.49	1.07	0.28

Note. Significant terms featured in bold, and terms approaching significance are underlined. *b* = regression coefficient. Contrast 1 refers to the $n+2$ preview effect when the $n+1$ preview was an identity, and Contrast 2 refers to the $n+2$ preview effect when the $n+1$ preview was a pseudocharacter.

Rayner, 2009). These results demonstrate, again, that there is processing advantage of the one-word strings in both the fovea and parafovea during Chinese reading, and the findings are entirely complementary to those we obtained in Experiments 1 and 2, as well as the previous research by Cutter et al (2014) and Zang et al., (2021,2023). Together, these results from the three experiments provide strong evidence for the claim that frequently used idioms and phrases can be represented lexically as MCUs and are processed in the parafovea (and in the fovea) as single lexicalized representations. It is clearly the case that lexical processing in Chinese can be operationalized over linguistic units that are larger than a single word unit.

General discussion

In three boundary experiments, we set out to investigate whether frequently occurring Chinese four-character idioms and four-character phrases are represented and processed foveally and parafoveally as single lexicalised MCUs, this even though they are relatively long and

Table A4
Fixed effect estimations for all the eye movement measures when conditional probability of phrases was included as a covariate in Experiment 2.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.40	0.02	**287.32**	**< 2e-16**	5.40	0.02	**275.00**	**< 2e-16**	5.47	0.02	**241.24**	**< 2e-16**
	$n+1$ preview	0.19	0.01	**12.43**	**< 2e-16**	0.22	0.02	**13.19**	**< 2e-16**	0.28	0.02	**15.89**	**< 2e-16**
	$n+2$ preview	0.05	0.01	**3.90**	**0.00**	0.06	0.01	**4.39**	**0.00**	0.07	0.02	**4.38**	**0.00**
	Interaction	−0.07	0.02	**−3.95**	**0.00**	−0.07	0.02	**−3.53**	**0.00**	−0.08	0.02	**−3.80**	**0.00**
	Contrast1	0.05	0.01	**3.61**	**0.00**	0.06	0.01	**4.19**	**0.00**	0.07	0.02	**4.12**	**0.00**
	Contrast2	−0.02	0.01	−1.61	0.11	−0.01	0.02	−0.57	0.57	−0.01	0.02	−0.65	0.52
	CP	0.00	0.03	**−0.08**	0.94	−0.01	0.04	−0.24	0.81	−0.01	0.04	−0.24	0.81
Constituent 2 ($n+2$)	(Intercept)	5.37	0.02	**291.68**	**< 2e-16**	5.37	0.02	**278.55**	**< 2e-16**	5.44	0.02	**238.96**	**< 2e-16**
	$n+1$ preview	0.01	0.01	0.94	0.35	0.01	0.02	0.53	0.60	0.02	0.02	0.91	0.36
	$n+2$ preview	0.04	0.01	**2.82**	**0.00**	0.04	0.02	**2.51**	**0.01**	0.06	0.02	**3.70**	**0.00**
	Interaction	−0.04	0.02	−1.87	0.06	−0.04	0.02	−1.72	0.09	−0.06	0.02	**−2.48**	**0.01**
	Contrast1	0.04	0.01	**2.86**	**0.00**	−0.04	0.01	**2.68**	**0.01**	0.06	0.02	**3.79**	**0.00**
	Contrast2	0.00	0.01	0.22	0.83	0.00	0.01	0.00	1.00	0.01	0.02	0.31	0.76
	CP	−0.14	0.04	**−3.56**	**0.00**	−0.14	0.04	**−3.36**	**0.00**	−0.18	0.06	**−3.21**	**0.00**
The whole region	(Intercept)	5.40	0.02	**324.85**	**< 2e-16**	5.40	0.02	**248.62**	**< 2e-16**	5.84	0.03	**183.39**	**< 2e-16**
	$n+1$ preview	0.17	0.02	**10.41**	**< 2e-16**	0.19	0.02	**8.09**	**0.00**	0.24	0.02	**13.13**	**< 2e-16**
	$n+2$ preview	0.06	0.01	**4.49**	**0.00**	0.11	0.02	**5.31**	**0.00**	0.13	0.02	**7.17**	**0.00**
	Interaction	−0.08	0.02	**−3.99**	**0.00**	−0.11	0.03	**−3.52**	**0.00**	−0.13	0.02	**−5.37**	**0.00**
	Contrast1	0.06	0.01	**4.43**	**0.00**	0.12	0.02	**5.61**	**0.00**	0.13	0.02	**7.22**	**0.00**
	Contrast2	−0.02	0.01	−1.42	0.16	0.00	0.02	0.08	0.93	0.00	0.02	−0.19	0.85
	CP	−0.03	0.03	−1.00	0.32	−0.05	0.04	−1.10	0.28	−0.17	0.07	**−2.46**	**0.02**
		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
Constituent 1 ($n+1$)	(Intercept)	5.62	0.03	**197.21**	**< 2e-16**	5.71	0.03	**185.38**	**< 2e-16**	−1.66	0.15	**−11.37**	**< 2e-16**
	$n+1$ preview	0.32	0.03	**12.27**	**< 2e-16**	0.26	0.02	**11.83**	**< 2e-16**	−0.78	0.12	**−6.61**	**0.00**
	$n+2$ preview	0.07	0.02	**3.01**	**0.00**	0.07	0.02	**3.39**	**0.00**	−0.24	0.12	**−2.00**	**0.05**
	Interaction	−0.05	0.03	1.71	0.09	−0.05	0.03	**−2.00**	**0.05**	0.13	0.17	0.78	0.43
	Contrast1	0.07	0.02	**3.18**	**0.00**	0.06	0.02	**3.29**	**0.00**	–	–	–	–
	Contrast2	0.02	0.02	0.98	0.33	0.02	0.02	0.80	0.42	–	–	–	–
	CP	−0.02	0.06	0.40	0.69	−0.01	0.07	−0.20	0.84	−0.13	0.30	−0.45	0.65
Constituent 2 ($n+2$)	(Intercept)	5.55	0.03	**187.30**	**< 2e-16**	5.67	0.03	**176.88**	**< 2e-16**	−1.21	0.11	**−10.86**	**< 2e-16**
	$n+1$ preview	0.06	0.03	**2.32**	**0.02**	0.00	0.02	0.09	0.93	−0.25	0.11	**−2.31**	**0.02**
	$n+2$ preview	0.06	0.02	**2.57**	**0.01**	0.08	0.02	**3.73**	**0.00**	−0.19	0.10	−1.90	0.06
	Interaction	−0.04	0.03	−1.05	0.30	−0.06	0.03	**−2.02**	**0.04**	0.18	0.14	1.31	0.19
	Contrast1	–	–	–	–	0.08	0.02	**3.67**	**0.00**	–	–	–	–
	Contrast2	–	–	–	–	0.02	0.02	0.99	0.32	–	–	–	–
	CP	−0.24	0.08	**−3.09**	**0.00**	−0.27	0.08	**−3.37**	**0.00**	1.12	0.26	**4.26**	**0.00**
The whole region	(Intercept)	6.03	0.04	**169.43**	**< 2e-16**	6.24	0.04	**161.48**	**< 2e-16**	−5.39	0.47	**−11.57**	**<2e-16**
	$n+1$ preview	0.33	0.02	**14.45**	**< 2e-16**	0.22	0.02	**11.12**	**< 2e-16**	−0.65	0.33	**−1.96**	**0.05**
	$n+2$ preview	0.13	0.02	**6.58**	**0.00**	0.10	0.02	**5.75**	**0.00**	−0.53	0.32	−1.67	0.10
	Interaction	−0.10	0.03	**−4.03**	**0.00**	−0.06	0.02	**−2.76**	**0.01**	0.08	0.52	0.15	0.88
	Contrast1	0.13	0.02	**6.65**	**0.00**	0.10	0.02	**5.92**	**0.00**	–	–	–	–
	Contrast2	0.02	0.02	1.23	0.22	0.04	0.02	**2.14**	**0.03**	–	–	–	–
	CP	−0.20	0.08	**−2.50**	**0.02**	−0.23	0.08	**−2.75**	**0.01**	0.52	0.64	0.82	0.42

Note. Significant terms featured in bold, and terms approaching significance are underlined. *b* = regression coefficient. CP = Conditional Probability. Contrast 1 refers to the $n+2$ preview effect when the $n+1$ preview was an identity, and Contrast 2 refers to the $n+2$ preview effect when the $n+1$ preview was a pseudocharacter.

extend substantially into the parafoveal region. Our results from all three experiments are simple and straightforward: substantially increased parafoveal processing of the second constituents ($n+2$) of idioms and highly familiar phrases occurred when the first constituents ($n+1$) were available compared to when they were not in the parafovea, indicating that the first constituents activate processing of the whole lexical units, that is, MCUs, and thus license pre-processing of the second constituents to a greater extent. In addition, more parafoveal processing of the second constituents occurred when the first and second constituents were more likely to be judged as forming a single word, rather than multiple separate words, indicating that the lexical status of multicharacter strings does determine whether they are, or are not, processed as MCUs. It is also important to note that these effects cannot be, or at least cannot entirely be, attributed to predictability of the second constituent given the first, or the conditional probability of the two constituents of one type of target string relative to the other. As all these variables were pre-screened and formally quantified as covariates in the LMM analyses, and all the critical effects remained, suggesting that the processing advantage in foveal and parafoveal processing for the four-character idioms and highly familiar phrases extends beyond these variables (see Zang et al., 2021 for further discussion).

One might argue that the more pronounced $n+2$ preview effect

Table A5
Fixed effect estimations for all the eye movement measures when sentence naturalness was included as a covariate in Experiment 3.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Constituent 1 ($n+1$)	Intercept	5.41	0.02	**337.75**	**<0.001**	5.42	0.02	**329.45**	**<0.001**	5.52	0.02	**274.15**	**<0.001**
	Preview	0.03	0.01	**2.21**	**0.03**	0.03	0.01	**2.37**	**0.02**	0.05	0.01	**3.42**	**0.00**
	Ambiguous vs. 1-word	0.04	0.01	**2.67**	**0.01**	0.05	0.02	**3.12**	**0.00**	0.05	0.02	**3.08**	**0.00**
	2-word vs. ambiguous	0.00	0.01	−0.36	0.72	−0.02	0.02	−1.09	0.28	0.00	0.02	−0.14	0.89
	2-word vs. 1-word	0.03	0.02	**2.12**	**0.03**	0.03	0.02	1.86	0.06	0.05	0.02	**2.68**	**0.01**
	Preview × ambiguous vs. 1-word	0.00	0.03	0.17	0.86	−0.02	0.03	−0.58	0.56	−0.04	0.03	−1.18	0.24
	Preview × 2-word vs. ambiguous	−0.03	0.03	−1.24	0.22	−0.04	0.03	−1.38	0.17	−0.07	0.03	**−2.11**	**0.04**
	Preview × 2-word vs. 1-word	−0.03	0.03	−1.07	0.29	−0.06	0.03	−1.95	0.05	−0.11	0.03	**−3.27**	**0.00**
	Sentence naturalness	0.04	0.03	1.36	0.18	0.03	0.03	1.11	0.27	0.04	0.04	1.10	0.27
Constituent 2 ($n+2$)	Intercept	5.41	0.01	**382.34**	**<0.001**	5.40	0.01	**370.91**	**<0.001**	5.51	0.02	**319.33**	**<0.001**
	Preview	0.00	0.01	0.25	0.80	0.01	0.01	0.74	0.46	0.01	0.01	0.76	0.45
	Ambiguous vs. 1-word	0.06	0.01	**4.13**	**<0.001**	0.06	0.02	**3.80**	**<0.001**	0.09	0.02	**5.01**	**<0.001**
	2-word vs. ambiguous	−0.01	0.01	−0.79	0.43	−0.02	0.02	−1.18	0.24	−0.02	0.02	−1.12	0.26
	2-word vs. 1-word	0.05	0.02	**3.14**	**0.00**	0.04	0.02	2.43	0.02	0.07	0.02	**3.62**	**<0.001**
	Preview × ambiguous vs. 1-word	0.01	0.03	0.26	0.80	0.02	0.03	0.56	0.58	0.02	0.03	0.63	0.53
	Preview × 2-word vs. ambiguous	0.00	0.03	0.02	0.99	−0.01	0.03	−0.24	0.81	−0.02	0.03	−0.68	0.50
	Preview × 2-word vs. 1-word	0.01	0.03	0.27	0.79	0.01	0.03	0.31	0.75	0.00	0.03	−0.05	0.96
	Sentence naturalness	0.03	0.03	1.14	0.25	0.02	0.03	0.54	0.59	0.06	0.03	1.85	0.07
The whole region	Intercept	5.41	0.02	**356.00**	**< 2e-16**	5.39	0.02	**286.57**	**<0.001**	5.95	0.03	**194.48**	**<0.001**
	Preview	0.02	0.01	**2.05**	**0.04**	0.03	0.02	1.62	0.11	0.06	0.02	**3.72**	**<0.001**
	Ambiguous vs. 1-word	0.04	0.01	**3.04**	**0.00**	0.04	0.02	1.93	0.05	0.14	0.02	**6.71**	**<0.001**
	2-word vs. ambiguous	−0.01	0.01	−0.84	0.40	−0.02	0.02	−1.01	0.31	−0.05	0.02	**−2.28**	**0.02**
	2-word vs. 1-word	0.03	0.01	**2.04**	**0.04**	0.02	0.02	0.87	0.39	0.09	0.02	**4.09**	**<0.001**
	Preview × ambiguous vs. 1-word	−0.01	0.02	−0.50	0.62	−0.07	0.04	−1.64	0.10	−0.02	0.04	−0.60	0.55
	Preview × 2-word vs. ambiguous	−0.02	0.02	−1.01	0.31	−0.04	0.04	−0.82	0.41	−0.02	0.04	−0.54	0.59
	Preview × 2-word vs. 1-word	−0.04	0.02	−1.51	0.13	−0.10	0.04	**−2.51**	**0.01**	−0.05	0.04	−1.13	0.26
	Sentence naturalness	0.03	0.02	1.05	0.29	−0.03	0.04	−0.76	0.45	0.02	0.04	0.40	0.69

		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
Constituent 1 ($n+1$)	Intercept	5.71	0.03	**205.70**	**<0.001**	5.87	0.03	**169.03**	**< 2e-16**	−1.56	0.12	**−12.97**	**<2e-16**
	Preview	0.09	0.02	**5.07**	**<0.001**	0.05	0.02	**3.22**	**0.00**	−0.16	0.08	−1.94	0.05
	Ambiguous vs. 1-word	0.03	0.02	1.43	0.15	0.09	0.02	**4.40**	**0.00**	−0.06	0.11	−0.60	0.55
	2-word vs. ambiguous	0.02	0.02	0.83	0.41	0.00	0.02	0.01	0.99	0.13	0.10	1.31	0.19
	2-word vs. 1-word	0.05	0.03	**2.02**	**0.04**	0.09	0.02	**4.02**	**<0.001**	0.07	0.11	0.63	0.53
	Preview × ambiguous vs. 1-word	−0.08	0.04	−1.85	0.07	−0.07	0.04	−1.73	0.08	0.01	0.20	0.07	0.94
	Preview × 2-word vs. ambiguous	−0.01	0.04	−0.27	0.79	0.02	0.04	0.40	0.69	0.30	0.20	1.51	0.13
	Preview × 2-word vs. 1-word	−0.09	0.04	**−2.10**	**0.04**	−0.05	0.04	−1.33	0.19	0.32	0.20	1.58	0.12
	Sentence naturalness	0.11	0.05	**2.33**	**0.02**	0.13	0.04	**3.00**	**0.00**	0.07	0.21	0.33	0.74
Constituent 2 ($n+2$)	Intercept	5.72	0.03	**218.18**	**<0.001**	5.87	0.03	**188.56**	**<0.001**	−1.54	0.10	**−14.79**	**<2e-16**
	Preview	0.02	0.02	1.03	0.30	0.00	0.02	0.21	0.84	−0.10	0.08	−1.25	0.21
	Ambiguous vs. 1-word	0.14	0.02	**5.80**	**<0.001**	0.16	0.02	**7.64**	**<0.001**	−0.26	0.10	**−2.49**	**0.01**
	2-word vs. ambiguous	0.03	0.02	1.26	0.21	0.04	0.02	1.73	0.08	0.04	0.10	0.41	0.68
	2-word vs. 1-word	0.17	0.03	**6.42**	**<0.001**	0.20	0.02	**8.55**	**<0.001**	−0.22	0.11	−1.94	0.05
	Preview × ambiguous vs. 1-word	−0.01	0.05	−0.11	0.91	−0.02	0.04	−0.53	0.60	−0.16	0.20	−0.82	0.41
	Preview × 2-word vs. ambiguous	−0.01	0.05	−0.32	0.75	0.11	0.04	**2.59**	**0.01**	−0.27	0.20	−1.34	0.18
	Preview × 2-word vs. 1-word	−0.02	0.05	−0.43	0.67	0.08	0.04	**2.02**	**0.04**	−0.43	0.20	**−2.19**	**0.03**
	Sentence naturalness	0.08	0.05	1.58	0.11	0.13	0.04	**2.89**	**0.00**	−0.10	0.19	−0.49	0.62
The whole region	Intercept	6.23	0.04	**168.61**	**<0.001**	6.50	0.04	**165.46**	**<0.001**	−3.03	0.11	**−29.82**	**<0.001**
	Preview	0.09	0.02	**5.19**	**<0.001**	0.04	0.02	**2.47**	**0.01**	−0.21	0.17	−1.24	0.21
	Ambiguous vs. 1-word	0.14	0.02	**6.41**	**<0.001**	0.18	0.02	**9.60**	**<0.001**	0.12	0.22	0.54	0.59
	2-word vs. ambiguous	0.00	0.02	−0.01	0.99	0.02	0.02	1.09	0.28	0.28	0.20	1.42	0.16
	2-word vs. 1-word	0.14	0.02	**5.79**	**<0.001**	0.20	0.02	**9.66**	**<0.001**	0.40	0.22	1.79	0.07
	Preview × ambiguous vs. 1-word	−0.07	0.04	−1.67	0.10	−0.03	0.04	−0.76	0.45	−0.10	0.43	−0.25	0.80

(*continued on next page*)

Table A5 (*continued*)

Region	Effect	FFD				SFD				GD			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
	Preview × 2-word vs. ambiguous	0.00	0.04	0.03	0.98	0.02	0.04	0.49	0.63	0.22	0.39	0.57	0.57
	Preview × 2-word vs. 1-word	−0.07	0.04	−1.63	0.10	−0.01	0.04	−0.27	0.79	0.12	0.43	0.29	0.78
	Sentence naturalness	0.06	0.05	1.26	0.21	0.12	0.04	**3.08**	**0.00**	−0.03	0.38	−0.08	0.94

Note. Significant terms featured in bold, and terms approaching significance are underlined. b = regression coefficient.

observed in Experiments 1 and 2 cannot be solely attributed to the incorrect $n + 1$ previews disrupting the lexical processing of the MCU. Instead, it is possible that the implausibility of the first constituent within the sentence context causes difficulty in its semantic integration. In a clever study by Staub, Rayner, Pollatsek, Hyönä and Majewski (2017), the plausibility of the first constituent of a noun-noun compound (e.g., *cafeteria manager*) was manipulated by varying the combination of the preceding verb (*visited* versus *talked to*) and the first constituent (the plausible condition: *visited the cafeteria*; and the implausible condition, *talked to the cafeteria*).[2] Note, the whole compound (*cafeteria manager*) always remained plausible in context. Staub et al. found increased reading times for the first noun when it was implausible. This indicates that English readers might prefer to employ a word-based processing strategy, incrementally and rapidly processing and integrate the first constituent into the sentence context. This finding might occur due to compounds such as "cafeteria manager" not being represented as a single unit.

Contrary to the findings of Staub et al., there is prior research on Chinese reading, using the same paradigm and manipulations, and these studies have showed a different pattern. Specifically, when the whole target string was plausible within the context, the plausibility of the first constituent did not significantly impact the reading of two-character words (Yang, Staub, Li, Wang, & Rayner, 2012), three-character words (Zhou & Li, 2021), or familiar four-character phrases comprised of two two-character words (Wang, Yang, Biemann, & Li, 2023). However, for unfamiliar four-character phrases, gaze durations were shorter when the first constituent was plausible compared to when it was implausible (Wang et al., 2023; Yao et al., 2022). These findings suggest that Chinese readers tend to process familiar multiple constituent words and phrases as a whole, rather than initially segmenting constituent characters/words into independent smaller units before then integrating those smaller units at a later stage of reading. If this is the case, in the present experiments, it is very unlikely that the contextual compatibility of the first constituent of the MCUs influenced the magnitude of the $n + 2$ preview effect under identity and pseudocharacter $n + 1$ previews. Furthermore, in Experiment 3, the first constituent remained consistently identical across the different types of target strings. This ensured that the preceding sentence context up to the first constituent was always identical. Consequently, readers had no opportunity to integrate the first two-character word into the sentence differentially across the different preview conditions. In sum, it seems unlikely that issues of plausibility contributed to our effects.

Our results provide strong and consistent evidence to support the MCU Hypothesis stipulating that linguistic units like highly familiar four-character idioms and phrases are processed as lexicalized single units (MCUs). It appears that as readers make fixations during natural reading, visual and cognitive processes can be operationalised over linguistic units corresponding to both single and multi-word representations that are stored in the lexicon. Recall earlier, that our consideration of MCUs in respect of parafoveal processing, preview benefits, formation of oculomotor commitments (i.e., decisions of where and when to move the eyes) and the identification of multiple words via a single lexical representation reflects theorizing beyond existing theoretical frameworks on language use and processing (Bod, 2006; Bybee, 2006; see also Arnon & Snider, 2010; Conklin & Schmitt, 2008, 2012; Shaoul & Westbury, 2011; Siyanova-Chanturia et al., 2011). To this extent, if readers process linguistic units in the way we have suggested, then there are significant implications for theories of on-line oculomotor decision making during reading.

As discussed in the Introduction, over previous decades, one of the most contentious debates with respect to current models of eye movement control in reading has centered on whether lexical processing occurs serially or in parallel. Recall, E-Z Reader posits that attention required to support lexical processing is allocated to only one word at a time (e.g., Reichle et al., 1998; Reichle, 2011), whereas SWIFT (Engbert et al., 2002; Engbert & Kliegl, 2011) and OB1-Reader (Snell et al., 2018) assume that attention is allocated over multiple words to support simultaneous lexical processing in parallel.

At a general level, we do not consider that the results fit so neatly with the parallel processing account, although this might initially seem odd, given that our results suggest that under certain circumstances, multiple words are identified simultaneously. However, a key aspect of our findings is that effects across matched strings were differential in respect of whether parafoveal character strings were, or were not, MCUs. On our understanding, parallel accounts do not differentiate between strings on this basis, and therefore, one might expect that matched strings under all our conditions should have been processed in parallel. This was clearly not the case, and therefore, to us, it is unclear how this approach might explain the effects.

In contrast, it is possible that the current findings may align with the serial processing position, but broadly, this would only be the case under the assumption that multiple word units might be represented alongside individual words in the mental lexicon. Recently, a newly proposed computational model, Über-Reader, has integrated word identification, sentence and discourse processing components into the E-Z Reader framework, whilst retaining the fundamental assumption that lexical processing and identification occurs serially, one word at a time (Reichle, 2021). In a simulation using Über-Reader, Reichle and Schotter (2020) found that word pairs can be recalled but accurate recall is only limited to short and high frequent word sequences like "in the". They argued that given that the majority of English words are more than three letters long, and given the highly productive nature of language, it is unlikely for most word sequences to be encountered frequently enough to be represented in memory. They also suggest that "the parallel identification of two or more words would thus likely be limited to sequences like 'in the', as well as perhaps idioms or commonly used phrases" (p.169). Nevertheless, the Über-Reader has not yet simulated how identification of frequently co-occurring word pairs influences eye movement patterns during natural sentence reading. Despite the current impasse that exists between serial and parallel lexical processing positions, the MCU hypothesis offers potential to move forward, at least, to some degree. Serial lexical processing can be maintained, but multiple words can be processed in parallel when these words comprise a lexicalized unit with which readers are familiar.

In our view, our findings fit most neatly with the Chinese Reading Model (Li & Pollatsek, 2020), though again, we feel modification to

[2] Whilst the example we provide here is that used by Staub et al., in fact, several of the stimuli adopted by Staub differed in nature to this. Regardless, for the sake of clarity, here we adhere to an example that was consistent with the claims that Staub et al made in their paper.

Table A6
Fixed effect estimations for all the eye movement measures when predictability of the second constituents of target strings given the preceding context including the first constituents was included as a covariate in Experiment 3.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Constituent 1 ($n+1$)	Intercept	5.41	0.02	**339.94**	**<0.001**	5.42	0.02	**331.10**	**<0.001**	5.52	0.02	**274.61**	**<0.001**
	Preview	0.03	0.01	**2.21**	**0.03**	0.03	0.01	**2.38**	**0.02**	0.05	0.01	**3.41**	**0.00**
	Ambiguous vs. 1-word	0.04	0.02	**2.27**	**0.02**	0.05	0.02	**2.87**	**0.00**	0.05	0.02	**2.28**	**0.02**
	2-word vs. ambiguous	0.00	0.01	−0.07	0.95	−0.01	0.01	−0.86	0.39	0.00	0.02	0.10	0.92
	2-word vs. 1-word	0.04	0.02	2.21	0.03	0.04	0.02	2.16	0.03	0.05	0.02	**2.35**	**0.02**
	Preview × ambiguous vs. 1-word	0.00	0.03	0.16	0.87	−0.02	0.03	−0.58	0.56	−0.04	0.03	−1.19	0.23
	Preview × 2-word vs. ambiguous	−0.03	0.03	−1.23	0.22	−0.04	0.03	−1.36	0.17	−0.07	0.03	**−2.10**	**0.04**
	Preview × 2-word vs. 1-word	−0.03	0.03	−1.07	0.29	−0.06	0.03	−1.94	0.05	−0.11	0.03	**−3.27**	**0.00**
	Predictability	−0.02	0.04	−0.56	0.58	0.00	0.04	−0.05	0.96	−0.05	0.05	−0.95	0.34
Constituent 2 ($n+2$)	Intercept	5.41	0.01	**381.75**	**<0.001**	5.40	0.01	**371.74**	**<0.001**	5.51	0.02	**321.64**	**<0.001**
	Preview	0.00	0.01	0.24	0.81	0.01	0.01	0.77	0.44	0.01	0.01	0.73	0.47
	Ambiguous vs. 1-word	0.05	0.02	**3.00**	**0.00**	0.05	0.02	**2.50**	**0.01**	0.07	0.02	**3.26**	**0.00**
	2-word vs. ambiguous	−0.01	0.01	−0.57	0.57	−0.02	0.02	−1.03	0.30	−0.01	0.02	−0.74	0.46
	2-word vs. 1-word	0.04	0.02	**2.53**	**0.01**	0.03	0.02	1.41	0.16	0.06	0.02	**2.66**	**0.01**
	Preview × ambiguous vs. 1-word	0.01	0.03	0.26	0.79	0.02	0.03	0.59	0.55	0.02	0.03	0.64	0.52
	Preview × 2-word vs. ambiguous	0.00	0.03	0.02	0.98	−0.01	0.03	−0.27	0.79	−0.02	0.03	−0.66	0.51
	Preview × 2-word vs. 1-word	0.01	0.03	0.29	0.77	0.01	0.03	0.32	0.75	0.00	0.03	−0.01	0.99
	Predictability	−0.06	0.04	−1.38	0.17	−0.06	0.04	−1.48	0.14	−0.13	0.05	**−2.40**	**0.02**
The whole region	Intercept	5.41	0.02	**357.30**	**<0.001**	5.39	0.02	**285.63**	**<0.001**	5.95	0.03	**194.81**	**<0.001**
	Preview	0.02	0.01	**2.05**	**0.04**	0.03	0.02	1.61	0.11	0.06	0.02	**3.71**	**<0.001**
	Ambiguous vs. 1-word	0.03	0.02	**2.18**	**0.03**	0.03	0.03	1.25	0.21	0.10	0.03	**3.96**	**<0.001**
	2-word vs. ambiguous	−0.01	0.01	−0.63	0.53	−0.03	0.02	−1.18	0.24	−0.04	0.02	**−2.24**	**0.03**
	2-word vs. 1-word	0.03	0.02	1.67	0.09	0.01	0.03	0.30	0.77	0.06	0.03	**2.19**	**0.03**
	Preview × ambiguous vs. 1-word	−0.01	0.02	−0.50	0.61	−0.07	0.04	−1.69	0.09	−0.02	0.04	−0.60	0.55
	Preview × 2-word vs. ambiguous	−0.02	0.02	−1.01	0.31	−0.04	0.04	−0.82	0.41	−0.02	0.04	−0.54	0.59
	Preview × 2-word vs. 1-word	−0.04	0.02	−1.51	0.13	−0.11	0.04	**−2.56**	**0.01**	−0.05	0.04	−1.14	0.26
	Predictability	−0.04	0.04	−1.05	0.30	−0.02	0.06	−0.27	0.79	−0.17	0.06	**−2.72**	**0.01**

		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
Constituent 1 ($n+1$)	Intercept	5.71	0.03	**203.66**	**<0.001**	5.87	0.04	**165.94**	**< 2e-16**	−1.56	0.12	**−12.96**	**<2e-16**
	Preview	0.09	0.02	**5.06**	**<0.001**	0.05	0.02	**3.19**	**0.00**	−0.16	0.08	−1.92	0.06
	Ambiguous vs. 1-word	0.03	0.03	1.25	0.21	0.08	0.03	**3.01**	**0.00**	−0.20	0.13	−1.56	0.12
	2-word vs. ambiguous	0.03	0.02	1.40	0.16	0.01	0.02	0.74	0.46	0.14	0.10	1.41	0.16
	2-word vs. 1-word	0.07	0.03	**2.34**	**0.02**	0.09	0.03	**3.59**	**<0.001**	−0.06	0.13	−0.45	0.65
	Preview × ambiguous vs. 1-word	−0.08	0.04	−1.87	0.06	−0.07	0.04	−1.75	0.08	0.01	0.20	0.04	0.96
	Preview × 2-word vs. ambiguous	−0.01	0.04	−0.24	0.81	0.02	0.04	0.43	0.67	0.30	0.20	1.51	0.13
	Preview × 2-word vs. 1-word	−0.09	0.04	**−2.10**	**0.04**	−0.05	0.04	−1.32	0.19	0.31	0.20	1.55	0.12
	Predictability	−0.06	0.07	−0.95	0.34	−0.15	0.06	**−2.30**	**0.02**	−0.58	0.31	−1.86	0.06
Constituent 2 ($n+2$)	Intercept	5.72	0.03	**215.36**	**<0.001**	5.87	0.03	**185.54**	**<0.001**	−1.54	0.10	**−14.99**	**< 2e-16**
	Preview	0.02	0.02	1.00	0.32	0.00	0.02	0.18	0.86	−0.10	0.08	−1.27	0.20
	Ambiguous vs. 1-word	0.11	0.03	**3.59**	**<0.001**	0.12	0.03	**4.73**	**<0.001**	−0.04	0.13	−0.29	0.77
	2-word vs. ambiguous	0.04	0.02	1.65	0.10	0.05	0.02	**2.43**	**0.02**	0.03	0.10	0.32	0.75
	2-word vs. 1-word	0.14	0.03	**4.86**	**<0.001**	0.17	0.03	**6.61**	**<0.001**	0.00	0.13	−0.03	0.97
	Preview × ambiguous vs. 1-word	0.00	0.05	−0.10	0.92	−0.02	0.04	−0.51	0.61	−0.16	0.20	−0.80	0.42
	Preview × 2-word vs. ambiguous	−0.01	0.05	−0.32	0.75	0.11	0.04	**2.65**	**0.01**	−0.27	0.20	−1.33	0.18
	Preview × 2-word vs. 1-word	−0.02	0.05	−0.41	0.68	0.09	0.04	**2.09**	**0.04**	−0.43	0.20	**−2.16**	**0.03**
	Predictability	−0.20	0.07	**−2.74**	**0.01**	−0.25	0.07	**−3.83**	**<0.001**	0.88	0.27	**3.19**	**0.00**
The whole region	Intercept	6.23	0.04	**167.39**	**<0.001**	6.50	0.04	**162.99**	**<0.001**	−3.31	0.11	**−29.77**	**<2e-16**
	Preview	0.09	0.02	**5.23**	**<0.001**	0.04	0.01	**2.46**	**0.01**	−0.21	0.17	−1.24	0.21
	Ambiguous vs. 1-word	0.09	0.03	**3.43**	**0.00**	0.12	0.02	**5.41**	**<0.001**	0.23	0.27	0.85	0.39
	2-word vs. ambiguous	0.01	0.02	0.30	0.77	0.03	0.02	1.89	0.06	0.28	0.19	1.44	0.15
	2-word vs. 1-word	0.10	0.03	**3.65**	**<0.001**	0.16	0.02	**6.85**	**<0.001**	0.51	0.26	**1.97**	**0.05**

(*continued on next page*)

Table A6 (*continued*)

Region	Effect	FFD				SFD				GD			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
	Preview × ambiguous vs. 1-word	−0.07	0.04	−1.68	0.09	−0.03	0.03	−0.77	0.44	−0.11	0.43	−0.25	0.80
	Preview × 2-word vs. ambiguous	0.00	0.04	0.03	0.98	0.02	0.04	0.51	0.61	0.22	0.39	0.57	0.57
	Preview × 2-word vs. 1-word	−0.07	0.04	−1.64	0.10	−0.01	0.04	−0.25	0.80	0.12	0.40	0.29	0.77
	Predictability	−0.23	0.07	**−3.44**	**0.00**	−0.30	0.06	**−5.20**	**<0.001**	0.44	0.59	0.74	0.46

Note. Significant terms featured in bold, and terms approaching significance are underlined. b = regression coefficient.

incorporate units corresponding to MCUs in the mental lexicon would be a necessity. Recall earlier, that the CRM does not take a clear position regarding whether lexical identification is a serial or parallel process during reading. It assumes that lexical identification occurs via an interactive activation process such that all characters within perceptual span are processed simultaneously. The characters that might comprise one or more words, activate all those consistent possible words, and these activated words then compete with each other until one wins the competition and is, thus, identified and simultaneously segmented from the upcoming sentential character stream. In relation to the present study, presumably the first constituent of a MCU is initially processed in the parafovea. Simultaneously, this processing triggers the activation of all the lexical entries consistent with that initial constituent, that is, the individual characters that directly comprise the first constituent, alongside any MCUs of which the constituent is a part. Activation of these entries at the character and the subsequent word/MCU level alongside mutual inhibition between activated entries, leads to the identification of the word or MCU (see Cutter et al., 2014). Importantly, Zhou and Li (2021) have shown that when there exists a segmentation ambiguity such that the upcoming string might be short (e.g., one or two characters in length) or long (e.g., three or four characters in length), then it is most likely that readers will initially segment the string in favor of the longer string since the lexical representation corresponding to that string will accrue a greater degree of activation than will the shorter string (c.f., the heuristic implemented in the revised model of E-Z Reader by Yu, Liu & Reichle, 2021). In our view, this also fits neatly with the current findings in that we have shown that when strings are lexicalised as MCUs, they are more likely initially processed as a single unit than when they are not.

On the assumption that MCUs, that is units comprised of more than one word (e.g., *teddy bear*), might be lexically represented as single elements that may be activated in a manner similar to single words (e.g., *dog*), then it is reasonable to assume that (1) more than one word can be identified simultaneously, and (2) for unspaced character based languages like Chinese, those units will be simultaneously segmented. It is in this way that we consider that the CRM complements the MCU hypothesis neatly (see also Gu et al., 2023). To be clear, the way in which any particular sequence of words will be processed is determined by the way they are represented lexically. Of course, further computational modelling work is required to evaluate whether the CRM can be extended to capture how visual, linguistic and oculomotor control processes appear to operate in relation to MCUs during Chinese reading. A further open question concerns how such processing assumptions might be captured in computational accounts of eye movement control during reading of alphabetic languages.

Our results advance understanding of language processing, in particular, what visual and linguistic units in character based languages like Chinese are lexicalized in memory, essentially, in what form are they represented. Next, let us consider word spaced alphabetic languages like English. In English, idioms like “kick the bucket” have been suggested to be represented as a single element in the lexicon due to its meaning not being easily constructed from the meaning of its constituent words (Libben & Titone, 2008; Swinney & Cutler, 1979; Titone & Connine, 1999). And, as noted earlier, spaced compound words also appear to operate as MCUs in English (Cutter et al., 2014). These possibilities raise the question of what other multi-word phrases might operate similarly? That is to say, might the concept of MCU processing extend beyond idioms and spaced compounds in English? For example, are determiners such as “the”, “a”, and “an” processed as individual words in and of themselves in definite and indefinite noun phrases like “the woman”, “a woman” and “an ape”? Or instead, might it be the case that they are lexically represented and identified as noun phrase MCUs? Furthermore, consider prepositional verbs such as “care for”, “wash up”, “believe in”, or phrasal verbs like “took off” in “The plane took off”, or “knocked on” in “We knocked on the door”, or even “come up with” in “We need to come up with an answer”. To us, these seem likely MCU candidates, particularly non-separable prepositional and phrasal verbs. Similarly, if we consider prepositional phrases more generally, then the prepositions themselves require interpretation in relation to the noun phrase with which they appear in order that they might be interpreted appropriately (e.g., “in” or “by” in “in the garden”, or “by the apprentice”). Might some prepositional phrases, as with articles, be represented and identified as MCUs in relation to accompanying noun phrases? This seems to be particularly appropriate in relation to prepositional phrases in which the preposition may take a different meaning contingent on the noun phrase with which it co-occurs (“by” means something different in “by the greenhouse” compared to “by the apprentice”). There is clearly some dependency in these cases and unified consideration of the phrase is a necessity for successful interpretation (see Cutter, Martin, & Sturt, 2020; Hennecke, 2022). Thus, this raises the possibility that prepositions and articles might license processing of a contentful noun as per MCU processing described earlier (i.e., in a manner similar to the way that”-teddy” licenses processing of “bear”). And we note again that Über-Reader (Reichle & Schotter, 2020) seems to accurately identify pairs of words if those constituent words are short and of high frequency, though it is yet to be evaluated regarding how processing of these common short word sequences influences eye movements during sentence reading.

Extending these reflections, we might consider alphabetic languages with other characteristics. For example, in some languages (e.g., Finnish) prepositions do not sit alone as individual words but they are bound to nouns, that is, they appear in the written form as part of a single word – it seems highly likely that nouns and prepositions are identified together (as MCUs) in Finnish. Moreover, Finnish (like German and Turkish) is an agglutinate language, and for such languages there might be a processing situation that is a counterpart to that observed in relation to MCUs in English. And we note, here, that even though Finnish is an alphabetic language, and in many ways, fundamentally different to character based languages like Chinese, in the sense that both languages often have sequences of unspaced linguistic units of text that require segmentation for meaningful interpretation, there is commonality between both languages. For example, in Finnish, it is unlikely that readers process a long word (e.g., lumipallosotatantere meaning “snowball fight field”) as a single unit, as there is more orthographic information than can be processed effectively in a single fixation. It seems much more likely that readers process these words via “units” (perhaps, morphemes) that are lexically represented (e.g., Bertram & Hyönä, 2003; Hyönä, Bertram, & Pollatsek, 2004). To this extent, if we assume processing occurs in line with lexically represented

Table A7
Fixed effect estimations for all the eye movement measures when conditional probability of target strings was included as a covariate in Experiment 3.

		FFD				SFD				GD			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Constituent 1 ($n+1$)	Intercept	5.41	0.02	**338.43**	**<2e-16**	5.42	0.02	**329.99**	**<2e-16**	5.52	0.02	**273.90**	**< 2e-16**
	Preview	0.03	0.01	**2.22**	**0.03**	0.03	0.01	**2.39**	**0.02**	0.05	0.01	**3.42**	**0.00**
	Ambiguous vs. 1-word	0.03	0.02	1.95	0.05	0.04	0.02	**2.47**	**0.01**	0.04	0.02	1.92	0.06
	2-word vs. ambiguous	0.00	0.01	−0.09	0.93	−0.01	0.01	−0.88	0.38	0.00	0.02	0.08	0.94
	2-word vs. 1-word	0.03	0.02	1.85	0.06	0.03	0.02	1.69	0.09	0.04	0.02	1.96	0.05
	Preview × ambiguous vs. 1-word	0.00	0.03	0.16	0.88	−0.02	0.03	−0.59	0.56	−0.04	0.03	−1.20	0.23
	Preview × 2-word vs. ambiguous	−0.03	0.03	−1.23	0.22	−0.04	0.03	−1.36	0.17	−0.07	0.03	**−2.10**	**0.04**
	Preview × 2-word vs. 1-word	−0.03	0.03	−1.07	0.28	−0.06	0.03	−1.94	0.05	−0.11	0.03	**−3.27**	**0.00**
	Conditional Probability	−0.08	0.05	−1.54	0.12	−0.06	0.06	−1.14	0.26	−0.14	0.07	**−2.08**	**0.04**
Constituent 2 ($n+2$)	Intercept	5.41	0.01	**384.50**	**< 2e-16**	5.40	0.01	**378.36**	**<2e-16**	5.51	0.02	**324.03**	**< 2e-16**
	Preview	0.00	0.01	0.31	0.76	0.01	0.01	0.74	0.46	0.01	0.01	0.82	0.41
	Ambiguous vs. 1-word	0.05	0.02	**2.83**	**0.00**	0.04	0.02	**2.41**	**0.02**	0.07	0.02	**3.39**	**0.00**
	2-word vs. ambiguous	−0.01	0.01	−0.61	0.54	−0.02	0.02	−1.29	0.20	−0.01	0.02	−0.80	0.42
	2-word vs. 1-word	0.04	0.02	**2.27**	**0.02**	0.02	0.02	1.26	0.21	0.05	0.02	**2.68**	**0.01**
	Preview × ambiguous vs. 1-word	0.00	0.03	0.18	0.86	0.02	0.03	0.57	0.57	0.02	0.03	0.53	0.60
	Preview × 2-word vs. ambiguous	0.00	0.03	0.03	0.98	−0.01	0.03	−0.24	0.81	−0.02	0.03	−0.65	0.51
	Preview × 2-word vs. 1-word	0.01	0.03	0.20	0.84	0.01	0.03	0.32	0.75	0.00	0.03	−0.12	0.91
	Conditional Probability	−0.13	0.05	**−2.35**	**0.02**	−0.13	0.06	**−2.39**	**0.02**	−0.21	0.07	**−3.00**	**0.00**
The whole region	Intercept	5.41	0.02	**356.47**	**<2e-16**	5.39	0.02	**285.95**	**<2e-16**	5.95	0.03	**195.50**	**< 2e-16**
	Preview	0.02	0.01	**2.06**	**0.04**	0.03	0.02	1.60	0.11	0.06	0.02	**3.75**	**0.00**
	Ambiguous vs. 1-word	0.03	0.01	**2.24**	**0.03**	0.03	0.02	1.11	0.27	0.10	0.02	**4.10**	**0.00**
	2-word vs. ambiguous	−0.01	0.01	−0.66	0.51	−0.03	0.02	−1.20	0.23	−0.05	0.02	**−2.34**	**0.02**
	2-word vs. 1-word	0.02	0.01	1.66	0.10	0.00	0.02	0.06	0.95	0.05	0.02	**2.09**	**0.04**
	Preview × ambiguous vs. 1-word	−0.01	0.02	−0.52	0.60	−0.07	0.04	−1.68	0.09	−0.03	0.04	−0.63	0.53
	Preview × 2-word vs. ambiguous	−0.02	0.02	−1.00	0.32	−0.04	0.04	−0.82	0.41	−0.02	0.04	−0.53	0.60
	Preview × 2-word vs. 1-word	−0.04	0.02	−1.52	0.13	−0.11	0.04	**−2.55**	**0.01**	−0.05	0.04	−1.15	0.25
	Conditional Probability	−0.07	0.05	−1.44	0.15	−0.06	0.07	−0.81	0.42	−0.29	0.08	**−3.62**	**0.00**

		Go-past				TFD				SP			
Region	Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*z*	*p*
Constituent 1 ($n+1$)	Intercept	5.71	0.03	**203.71**	**< 2e-16**	5.87	0.04	**166.50**	**< 2e-16**	−1.52	0.12	**−12.51**	**<2e-16**
	Preview	0.09	0.02	**5.07**	**0.00**	0.05	0.02	**3.21**	**0.00**	−0.16	0.08	−1.93	0.05
	Ambiguous vs. 1-word	0.04	0.03	1.57	0.12	0.10	0.02	**3.96**	**0.00**	−0.14	0.12	−1.21	0.23
	2-word vs. ambiguous	0.03	0.02	1.39	0.17	0.01	0.02	0.69	0.49	0.14	0.10	1.38	0.17
	2-word vs. 1-word	0.07	0.03	**2.70**	**0.01**	0.11	0.02	**4.49**	**0.00**	−0.01	0.12	−0.05	0.96
	Preview × ambiguous vs. 1-word	−0.08	0.04	−1.87	0.06	−0.07	0.04	−1.75	0.08	0.01	0.20	0.06	0.95
	Preview × 2-word vs. ambiguous	−0.01	0.04	−0.24	0.81	0.02	0.04	0.43	0.67	0.30	0.20	1.52	0.13
	Preview × 2-word vs. 1-word	−0.09	0.04	**−2.10**	**0.04**	−0.05	0.04	−1.32	0.19	0.32	0.20	1.58	0.11
	Conditional Probability	−0.06	0.09	−0.73	0.47	−0.12	0.08	−1.48	0.14	−0.59	0.41	−1.43	0.15
Constituent 2 ($n+2$)	Intercept	5.72	0.03	**216.77**	**< 2e-16**	5.87	0.03	**187.34**	**< 2e-16**	−1.54	0.10	**−14.88**	**< 2e-16**
	Preview	0.02	0.02	1.09	0.27	0.00	0.02	0.23	0.82	−0.10	0.08	−1.27	0.20
	Ambiguous vs. 1-word	0.11	0.03	**4.03**	**0.00**	0.14	0.02	**5.82**	**0.00**	−0.10	0.12	−0.85	0.40
	2-word vs. ambiguous	0.04	0.02	1.59	0.11	0.05	0.02	**2.35**	**0.02**	0.04	0.10	0.38	0.70
	2-word vs. 1-word	0.15	0.03	**5.27**	**0.00**	0.19	0.02	**7.67**	**0.00**	−0.06	0.12	−0.51	0.61
	Preview × ambiguous vs. 1-word	−0.01	0.05	−0.20	0.84	−0.02	0.04	−0.55	0.58	−0.16	0.20	−0.80	0.43
	Preview × 2-word vs. ambiguous	−0.01	0.05	−0.31	0.76	0.11	0.04	**2.64**	**0.01**	−0.27	0.20	−1.35	0.18
	Preview × 2-word vs. 1-word	−0.02	0.05	−0.50	0.62	0.08	0.04	**2.05**	**0.04**	−0.43	0.20	**−2.17**	**0.03**
	Conditional Probability	−0.28	0.10	**−2.90**	**0.00**	−0.28	0.09	**−3.24**	**0.00**	1.04	0.36	**2.91**	**0.00**
The whole region	Intercept	6.23	0.04	**167.61**	**< 2e-16**	6.50	0.04	**163.89**	**< 2e-16**	−4.15	0.23	**−18.00**	**<2e-16**

(*continued on next page*)

Table A7 (*continued*)

Region	Effect	FFD				SFD				GD			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
	Preview	0.09	0.02	**5.24**	**0.00**	0.04	0.02	**2.49**	**0.01**	−0.22	0.18	−1.25	0.21
	Ambiguous vs. 1-word	0.11	0.02	**4.35**	**0.00**	0.16	0.02	**7.54**	**0.00**	0.20	0.27	0.75	0.46
	2-word vs. ambiguous	0.00	0.02	0.22	0.83	0.03	0.02	1.78	0.07	0.30	0.21	1.47	0.14
	2-word vs. 1-word	0.11	0.03	**4.46**	**0.00**	0.19	0.02	**8.88**	**< 2e-16**	0.50	0.26	1.93	0.05
	Preview × ambiguous vs. 1-word	−0.07	0.04	−1.71	0.09	−0.03	0.04	−0.78	0.44	−0.09	0.45	−0.19	0.85
	Preview × 2-word vs. ambiguous	0.00	0.04	0.05	0.96	0.02	0.04	0.52	0.60	0.20	0.41	0.50	0.62
	Preview × 2-word vs. 1-word	−0.07	0.04	−1.65	0.10	−0.01	0.04	−0.25	0.80	0.12	0.43	0.28	0.78
	Conditional Probability	−0.26	0.09	**−3.05**	**0.00**	−0.24	0.07	**−3.30**	**0.00**	0.61	0.75	0.81	0.42

Note. Significant terms featured in bold, and terms approaching significance are underlined. b = regression coefficient.

units that may not always correspond to a single word (i.e., lexically represented units may be a word, more than one word, or sub-units of a single word), then the MCU Hypothesis offers flexibility with respect to how visual and linguistic processing might occur over portions of text during reading. It is not simply that multiple words may be processed as a single unit, but also, larger single multi-unit words may be broken down into smaller units. Again, this possibility exists irrespective of the word spacing (or agglutination) characteristics of a language. Also, we acknowledge that this suggestion may be considered similar to the Word Grouping Hypothesis (Drieghe, Pollatsek, Staub & Rayner, 2008; Radach, 1996), however, we note that the Word Grouping Hypothesis is primarily focused on issues of saccadic targeting, not lexical identification and more general oculomotor control and lexical processing commitments. Nonetheless, there is clearly some linkage between these suggestions.

In line with our theorising, Huettig, Audring and Jackendoff (2022) recently put forward the notion of the "extended lexicon", specifying that the lexicon contains a large number of multiword items including idioms, cliches, frequent fixed expressions, collocations, and even syntactic constructions (e.g., "that gem of a result", meaning that result, which was a gem). They argued that these items must be learned and stored in the lexicon. Indeed, there is a line of psycholinguistic research that seems to support this view (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Siyanova-Chanturia et al., 2011; Tremblay, Derwing, Libben, & Westbury, 2011). For example, Bannard and Matthews (2008) showed that two- to three year old children were faster and better at repeating more frequent phrases (e.g., *a drink of milk*) than less frequent ones (e.g., *a drink of tea*). In British child-directed speech, the phrase "a drink of milk" occurs more frequently than "a drink of tea", though it is important to note that when considering constituents of these phrases, their frequencies are equivalent, that is, *tea* is as common as *milk*. In a follow up experiment, Arnon and Snider (2010) demonstrated that readers are sensitive to frequency information on phrases or multi-word units, and that this holds not only for "special", highly common phrases (e.g., *don't have to worry*), but also for phrases across the entire frequency range. Furthermore, Siyanova-Chanturia et al. (2011) found that both native and non-native speakers, across all proficiency levels, read frequent phrases faster than infrequent phrases. Native speakers and more proficient non-native speakers read binomial phrases (e.g., *bride and groom*) faster than their reversed forms (e.g., *groom and bride*), and showed sensitivity to the frequency of phrase occurrence in the English language, whereas less proficient non-native speakers had comparable reading speeds for both forms of phrase. These studies indicate that readers appear to have stored representations for frequently encountered multi-word sequences or phrases, and every occurrence of a particular form contributes to its degree of enrichment in the reader's mental lexicon (see also Bod, 2006; Bybee, 2006). Further work is required to establish the extent to which strings of this type and beyond are indeed represented lexically in this way, and whether this leads them to be processed as MCUs in natural reading. What should be very clear, however, is that if visual and linguistic processing is operationalized over multiple constituents that are lexically represented as single elements in the manner that we are suggesting, then this will have significant implications for theoretical accounts of where, and when, we move our eyes during natural reading.

At a more general level, our consideration of MCU processing, lexical identification and eye movement control during natural reading raises a question of whether current theoretical models of lexical processing realistically capture how such processing actually occurs during natural reading. Almost all current models of lexical identification seek to explain isolated word recognition (even the Interactive Activation model implemented within the Chinese Reading Model put forward by Li & Pollatsek, 2020, derives from a model of isolated word identification, McClelland & Rumelhart, 1981). These models stipulate that words are activated based on visual stimulation that is constant and uniform over the entire period during which the word is identified. However, in our view, identifying a word, or more precisely, perhaps a multi-constituent unit, in reading is not a constant and uniform process; lexical identification during reading is episodic, that is, visual information directly activates stored lexical representations and this is delivered during discrete time periods (fixations). And for lexical identification to occur during natural reading, there are most often at least two delivery episodes (the first when the lexical unit is in the parafovea and the second when the lexical unit is in the fovea, that is, when it is fixated). Furthermore, the nature (quality) of the visual information that is delivered to the lexical processing system, has been unequivocally demonstrated to affect the ease with which a word is identified (visual degradation effects, e.g., Jordan, McGowan, & Paterson, 2012). And the quality of the visual information that is available to be processed in respect of the identity of a word (or MCU) directly determines the speed with which that lexical representation is identified (and, note that, at least arguably, this aspect of processing itself impacts the decision of when to move our eyes during reading). Furthermore, the critical point to make here is that the nature of that visual input changes qualitatively from one episode to another (i.e., one fixation to another) during natural reading. When a lexical unit is processed parafoveally, visual information is degraded and impoverished whereas when a lexical unit is directly fixated, visual information is rich and clear – to be clear, this is not processing with smooth progression and continuity, but instead processing occurs according to step changes in the quality of visual input with these step changes occurring across the period during which a lexical unit is identified. Of course, it is well established that change in the quality of visual information available about a word from one "episode" (fixation) to another is directly related to a number of factors (e.g., distance away from an upcoming word; length of the preceding word, position of the fixation within the word itself, length of the word itself, etc. See Rayner, 2009). Thus, quite critically, these observations and theoretical suggestions together lead us to the view that unless theoretical/computational models of word identification during natural reading reflect the episodic nature of the delivery of visual information

to the lexical processing system, and thereby, capture stepped differences in the quality of the visual input across those episodes, then those lexical identification models cannot provide an accurate account of this process in respect of natural reading. And we consider that this is the case in respect of lexical representations corresponding to individual words, more than one word, or sub-units of a single word.

To summarize, the present study provides strong support for the MCU hypothesis, demonstrating that lexical identification (via parafoveal and foveal processing) appears to be operationalized over MCUs. The MCU Hypothesis offers opportunity for theoretical progression beyond the current serial versus parallel lexical identification impasse and offers an opportunity for reformulation of theoretical research questions. We suggest that lexical identification in natural reading is episodic and with qualitative changes in the nature of the visual input across those episodes, and therefore, more ecologically valid computational models of lexical identification might be included in models of natural reading (because of quite direct linkage between lexical processing and oculomotor decisions). Since saccadic eye movements deliver the visual information that is required for lexical identification (and other linguistic processing), accounts of natural reading must reflect how eye movements constrain linguistic processing (and we acknowledge that this is a non-trivial aspect of any theoretical account of natural reading). Finally, and most importantly from our point of view, models of eye movement control during natural reading must reflect how oculomotor and linguistic processing are operationalized over lexical units that may be words or MCUs.

CRediT authorship contribution statement

Chuanli Zang: Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Shuangshuang Wang:** Investigation, Formal analysis. **Xuejun Bai:** Supervision. **Guoli Yan:** Supervision. **Simon P. Liversedge:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data sets, stimuli and analysis scripts are publicly available at: https://osf.io/6wuhs/

Acknowledgments

We acknowledge support from Economic and Social Research Council Grant (ES/R003386/1).

Appendix

Tables A1–A7

References

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*(6), 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.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241–248.

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1277–1287.

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.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48*, 615–634.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review, 23*(3), 291–320.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*, 711–733.

Chinese Academy of Social Sciences, Institute of Linguistics. (2016). Modern Chinese Dictionary (7th Edition). Beijing: The Commercial Press.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and non-native 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.

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2017). Reading sentences of uniform word length: Evidence for the adaptation of the preferred saccade length during reading. *Journal of Experimental Psychology: Human Perception and Performance, 43*(11), 1895–1911.

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2018). Reading sentences of uniform word length—II: Very rapid adaptation of the preferred saccade length. *Psychonomic Bulletin & Review, 25*(4), 1435–1440.

Cutter, M. G., Martin, A. E., & Sturt, P. (2020). Capitalization interacts with syntactic complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(6), 1146–1164.

Di Sciullo, A., & Williams, E. (1987). *On the Definition of Word*. Cambridge, MA: MIT Press.

Drieghe, D., Pollatsek, A., Staub, A., & Rayner, K. (2008). The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1552–1560.

Engbert, R., & Kliegl, R. (2011). Parallel graded attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (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.

Gu, J., Zhou, J., Bao, Y., Liu, J., Perea, M., & Li, X. (2023). The effect of transposed-character distance in Chinese reading. *Journal of Experimental Psychology: Learning Memory & Cognition, 49*(3), 464–476.

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, 112*, 662–689.

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.

Huang, B., & Liao, X. (2007). *Modern Chinese* (4th Edition, p. 266). Beijing: Higher Education Press.

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition, 224*, Article 105050.

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

Juhasz, B. J., Pollatsek, A., Hyönä, J., Drieghe, D., & Rayner, K. (2009). Parafoveal processing within and between words. *Quarterly Journal of Experimental Psychology, 62*(7), 1356–1376. https://doi.org/10.1080/17470210802400010

Jordan, T. R., McGowan, V. A., & Paterson, K. B. (2012). Reading with a filtered fovea: The influence of visual quality at the point of fixation during reading. *Psychonomic Bulletin & Review, 19*, 1078–1084.

Hennecke, I. (2022). Prepositional constituents in multi-word units: An experimental reading study of the French preposition de. *Linguistics, 60*(6), 1785–1810.

Hyönä, J., Bertram, R., & Pollatsek, A. (2004). Are long compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory & Cognition, 32*, 523–532.

Lexicon of Common Words in Contemporary Chinese Research Team. (2008). *Lexicon of common words in contemporary Chinese*. Beijing, China: The Commercial Press.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye movement control during Chinese reading. *Psychological Review, 127*(6), 1139–1162.

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology, 58*(4), 525–552.

Li, X., & Su, X. (2021). *Lexicon of Common Words in Contemporary Chinese* (2nd Edition). Beijing: The Commercial Press.

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

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition, 36*(6), 1103–1121.

Liu, P., Li, W., Lin, N., & Li, X. (2013). Do Chinese readers follow the national standard rules for word segmentation during reading? *PLoS One1, 8*, e55440.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375–407.
McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578–586.
Packard, J. L. (2003). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.
Radach, R. (1996). *Blickbewegungen beim Lesen: Psychologische Aspekte der Determination von Fixationspositionen [Eye movements in reading: Psychological factors that determine fixation locations]*. Münster, Ger-many: Waxmann.
Rayner, K. (1975). The perceptual span and peripheral cues during reading. *Cognitive Psychology, 7*, 65–81.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.
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.
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*(1), 230–245.
Reichle, E. D. (2011). Serial-Attention Models of Reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 767–786). New York, NY: Oxford University Press.
Reichle, E. D. (2021). *Computational models of reading: A handbook*. New York: Oxford University Press.
Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125–157.
Reichle, E.D. & Schotter, E.R. (2020). A computational analysis of the constraints on parallel word identification. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. 164-170.
Schotter, E. R., & Payne, B. R. (2019). Eye movements and comprehension are important to reading. *Trends in Cognitive Sciences, 23*(10), 811–812.
Shaoul, C., & Westbury, C. F. (2011). Formulaic sequences: Do they exist and do they matter? *The Mental Lexicon, 6*(1), 171–196.
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.
Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning Memory and Cognition, 33*(6), 1162–1169.
Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behavior research methods, 50*(6), 2606–2629.
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.
Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*, 569–613.
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 & Review, 24*(3), 666–689.
Wang, J., Yang, J., Biemann, C, & Li, X. (2023). Insight to mechanism of semantic processing of lexicalized and novel compound words: an eye movement study. *Journal of Experimental Psychology: Learning, Memory and Cognition, 49*(11), 1812–1822.
Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction*. London: Bloomsbury Publishing Plc.
Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
Wulff, S., & Titone, D. (2014). Introduction to the special issue, bridging the methodological divide: Linguistic and psycholinguistic approaches to formulaic language. *The Mental Lexicon, 9*, 371–376.
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., Staub, A., Li, N., Wang, S., & Rayner, K. (2012). Plausibility effects when reading one- and two-character words in Chinese: Evidence from eye movements. *Journal of Experimental Psychology: Learning Memory and Cognition, 38*(6), 1801–1809.
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.
Yao, P., Alkhammash, R., & Li, X. (2022). Plausibility and Syntactic Reanalysis in Processing Novel Noun-noun Combinations During Chinese Reading: Evidence From Native and Non-native Speakers. *Scientific Studies of Reading, 26(5),* 390-408.
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.
Yu, L., Liu, Y., & Reichle, E. D. (2021). A corpus-based versus experimental examination of word- and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of experimental psychology. General, 150*(8), 1612–1641.
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, S. P. (2020). Word skipping in Chinese reading: The role of high-frequency preview and syntactic felicity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 603–620.
Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2021). Foveal and Parafoveal Processing of Chinese Three-character Idioms in Reading. *Journal of Memory and Language, 119*, Article 104243.
Zang, C., Fu, Y., Du, H., Bai, X., Yan, G., & Liversedge. (2023). Preview effects in reading of Chinese two-constituent words, idioms and phrases. *Journal of Experimental Psychology: Learning, Memory and Cognition.* https://doi.org/10.1037/xlm0001234
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.
Zhou, J., & Li, X. (2021). On the segmentation of Chinese incremental words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(8), 1353–1368.