Journal of Experimental Psychology: Learning, Memory, and Cognition

Processing Multiconstituent Units: Preview Effects During Reading of Chinese Words, Idioms, and Phrases

Chuanli Zang, Ying Fu, Hong Du, Xuejun Bai, Guoli Yan, and Simon P. Liversedge

Online First Publication, May 25, 2023. https://dx.doi.org/10.1037/xlm0001234

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001234

Processing Multiconstituent Units: Preview Effects During Reading of Chinese Words, Idioms, and Phrases

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

[1] School of Psychology and Computer Science, University of Central Lancashire

[2] Key Research Base of Humanities and Social Sciences of the Ministry of Education, Faculty of Psychology, Tianjin Normal University

Arguably, the most contentious debate in the field of eye movement control in reading has centered on whether words are lexically processed serially or in parallel during reading. Chinese is character-based and unspaced, meaning the issue of how lexical processing is operationalized across potentially ambiguous, multicharacter strings is not straightforward. We investigated Chinese readers' processing of frequently occurring multiconstituent units (MCUs), that is, linguistic units composed of more than a single word, that might be represented lexically as a single representation. In Experiment 1, we manipulated the linguistic category of a two-constituent Chinese string (word, MCU, or phrase) and the preview of its second constituent (identical or pseudocharacter) using the boundary paradigm with the boundary located before the two-constituent string. A robust preview effect was obtained when the second constituent, alongside the first, formed a word or MCU, but not a phrase, suggesting that frequently occurring MCUs are lexicalized and processed parafoveally as single units during reading. In Experiment 2, we further manipulated the phrase type of a two-constituent but three-character Chinese string (idiom with a one-character modifier and a two-character noun, or matched phrase) and the preview of the second constituent noun (identity or pseudocharacter). A greater preview effect was obtained for idioms than phrases, indicating that idioms are processed to a greater extent in the parafovea than matched phrases. Together, the results of these two experiments suggest that lexical identification processes in Chinese can be operationalized over linguistic units that are larger than an individual word.

Keywords: multiconstituent units, preview effects, eye movements, Chinese reading

One of the most controversial issues with regard to reading research is whether words are lexically processed serially or in parallel, that is, are multiple words encoded and identified simultaneously during reading? A considerable amount of research on the reading of alphabetic languages has investigated this issue using a variety of different tasks, yet despite considerable effort to settle whether lexical processing occurs serially or in parallel during natural reading, the matter remains under debate. In this paper, we will focus on this issue and consider it in relation to reading in a nonalphabetic language, namely Chinese. We do this because the characteristics of written Chinese are such that significant issues in relation to this question arise that simply do not for other (e.g., alphabetic) languages.

Chuanli Zang https://orcid.org/0000-0002-9573-4968

Simon P. Liversedge https://orcid.org/0000-0002-8579-8546

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

All data files, materials, and analysis scripts are available at: https://osf.io/wk3pj/.

Correspondence concerning this article should be addressed to Chuanli Zang, School of Psychology and Computer Science, University of Central Lancashire, Preston, Lancashire PR1 2HE, United Kingdom. Email: czang@uclan.ac.uk

In a recent opinion article, Snell and Grainger (2019) argued that readers are parallel processors, citing data from a flanker paradigm in which readers were required to make a semantic or syntactic categorization of a foveal target word, with a semantically (Snell, Declerck, & Grainger, 2018) or syntactically (Snell et al., 2017) congruent/incongruent word flanking its left side and right side. The target and flanking words were presented simultaneously for 170 ms before disappearing. Snell, Grainger, and colleagues found that response times were influenced by the semantic or syntactic congruency of flanking words, such that reaction times were shorter in the congruent compared to the incongruent conditions. They argued that as the display duration of 170 ms is shorter than the average time required to lexically identify a word, this demonstrates that multiple words are processed simultaneously.

In a follow-up piece, White et al. (2019) provided behavioral and neurophysiological evidence against the view of Snell and Grainger (2019). In a series of experiments using semantic categorization and lexical decision tasks, White et al. required participants to focus their attention on one word in some trials but distribute their attention to two words in other trials. When participants could accurately identify one word (80% correct) at a given time (mean = 84 ms), in the same amount of time, they were unable to identify both words (presumably, when attention was divided between both words). Furthermore, White et al. recorded functional magnetic resonance imaging responses to examine any possible neuropsychological markers of parallel word processing. They found that when two words were presented simultaneously, only one of them received attention with activation emerging in the anterior visual word form

area-2 (located at the interface of the visual and linguistic processing systems), reflecting a unique influence of the lexical frequency of the attended word in each trial. On the basis of these findings, White et al. claimed that only one word can be fully identified at a time, and thus lexical access is serial. These two recent studies illustrate the topicality of this issue in relation to the process of reading, demonstrating also that this debate remains alive and vibrant now as it has been for over a decade.

It is apparent that the nature of processing of words in the flanker paradigm, semantic categorization and lexical decision tasks is very likely quite different from the nature of processing that occurs during natural reading, as task requirements and task differences may induce modes of cognitive operation that are not engaged when readers process words in sentences naturally (Schotter & Payne, 2019). Therefore, the degree to which findings from such artificial tasks pertain directly to the serial versus parallel lexical processing debate in respect of natural reading may be regarded (by some at least) as unconvincing. Arguably, the most compelling, ecologically valid, empirical evidence derives from situations in which sentence reading occurs naturally. Thus, a further aim of the present study was to use a natural reading paradigm to determine whether there are circumstances in which lexical processing operates serially and, during the same experiment with the same participants, other circumstances in which words are processed in parallel. If such a demonstration was possible, then this would serve two purposes: First, it would illustrate that any account of serial versus parallel reading would have to be flexible and noncategorical (i.e., an account must explain why sometimes words are processed serially, and on other occasions, they are processed in parallel); second, any such demonstration and account might provide an opportunity to reconcile (at least to some degree) currently conflicting positions regarding serial versus parallel lexical processing in reading. To be clear, the current experiment investigated how lexical processing is operationalized across individual words, and beyond, during natural reading in a language with inherent lexical boundary ambiguity, and for which word segmentation is a necessity.

Currently, several influential computational models of oculomotor control during reading exist and offer alternative perspectives in relation to this theoretical issue. The E–Z Reader model (e.g., Reichle et al., 1998) assumes that word identification is a strictly serial process that operates such that attention is allocated sequentially from one word to the next, thereby allowing readers to keep track of word order for sentence comprehension. Therefore, words of a sentence are lexically processed only one at a time. Lexical processing of parafoveal word $n + 1$ occurs only after the current foveal word n is fully recognized. Similarly, lexical processing of word $n + 2$ begins only after lexical processing of $n + 1$ is completed. This claim has come into question due to observations of lexical parafoveal-on-foveal effects, that is, an influence of the lexical properties of the yet to be fixated word $n + 1$ on ongoing processing of the fixated word n (see Drieghe, 2011 for a review of studies showing such effects; see also Brothers et al., 2017 for evidence against lexical parafoveal on foveal effects) and preview effects of word $n + 2$ (e.g., Vasilev & Angele, 2017 for a review). Note also, though, that simulations of the E–Z Reader model have demonstrated the preview effects of word $n + 2$, though these were shown to be small in size (Schotter et al., 2014). It is important to note that, empirically, lexical parafoveal-on-foveal effects are often quite small in magnitude and do not occur consistently within the literature (again, see Drieghe, 2011), and such effects have been most often reported in corpus-based studies (e.g., Kennedy & Pynte, 2005; Kliegl et al., 2006). Parafoveal-on-foveal effects have been obtained much less frequently in carefully controlled experiments in which variables were orthogonally manipulated. For example, Brothers et al. (2017) conducted four well-controlled experiments with sufficient power and undertook a Bayesian meta-analysis combining their data with data from earlier studies, failing to obtain evidence of parafoveal-on-foveal effects in any of their investigations.

In contrast to the E–Z Reader model, the SWIFT model (Engbert et al., 2002) posits that attention is spatially distributed within the perceptual span (McConkie & Rayner, 1975), about the point of fixation, over multiple words, and thus, more than one word can be lexically processed, and potentially identified, in parallel. Of course, this aspect of the SWIFT model means that parafoveal-on-foveal effects are not at all problematic for it. Indeed, according to the SWIFT model, words to the right of the fixated word should be identified prior to that under direct fixation quite regularly. However, it is important to note that while SWIFT can account for parafoveal-on-foveal effects, its capability to recognize words out of sentential order has been criticized in that it is not immediately clear how word order information is maintained in order to allow for incremental interpretation during sentence comprehension (Reichle et al., 2009). A more recent, parallel, model that attempts to handle the issue of word order encoding, the OB1-reader (see Snell, van Leipsig, et al., 2018), includes mechanisms for mapping activated words onto possible spatial locations in a sentence-level representation via feedback with respect to individual words based on word length information, as well as syntactic and semantic information in the visual input. This model has been applied, primarily, to data derived from alphabetic reading situations, and it has yet to be subject to comprehensive empirical scrutiny. However, an obvious question arises with respect to its spatial mapping system when nonalphabetic languages such as Chinese are considered in which word spacing is absent, word length variability is very reduced, word boundary ambiguity is prevalent, and free word order reigns.

It is probably fair to say that the theoretical debate between serial versus parallel processing has currently reached a point where groups of researchers generally advocate one, or other, of the two alternative theoretical accounts. Fairly well-entrenched positions have been adopted, and data sets are generally proffered forth that are consistent with the views that are held. To date, there has been little movement forward in respect of resolving the debate by seeking an account that might accommodate (at least to some extent) results that are traditionally viewed to favor one position as well as those that favor the alternative position. To this extent, to us, it feels like something of an impasse has been reached. The purpose of the present paper is to try to take initial steps to develop an alternative perspective that might offer the potential to (at least partially) account for data from both sides of the debate.

Let us now consider eye movement control in relation to nonalphabetic languages. It remains a fact that the majority of research that has investigated whether words are processed serially or in parallel has been limited to reading in alphabetic languages like English or German, in which the word is a salient and clear visual and linguistic unit. Words (in most situations) are defined by spaces on each side, and understandably, in almost all models of lexical identification, they are the primary elements, or representations, featuring centrally and over which processes are operationalized (see Zang, 2019 for

more discussions). As noted earlier, unlike alphabetic languages, Chinese is a character-based, unspaced language. One or more characters comprise words, but there are no visual cues such as spaces to demarcate word boundaries in a text. There are also no visual or lexical indicators to mark each word's syntactic property, and word order is relatively free. Furthermore, there is sometimes ambiguity concerning the concept of a word in Chinese, and it is not straightforward for readers to discriminate words from other linguistic units such as phrases (e.g., Bai et al., 2008; He et al., 2021; Hoosain, 1992; Liu et al., 2013). Note that despite these characteristics, there is considerable evidence demonstrating that words in Chinese are psychologically real and play an important and fundamental role during reading (e.g., Bai et al., 2008; Li et al., 2013, 2014; see also Li et al., 2015; Li & Pollatsek, 2020; Zang et al., 2011 for reviews). These characteristics of written Chinese provide challenges for current models of eye movement control generally (largely because most were initially designed to explain eye movements in alphabetic languages), and more specifically, they make the issue of whether words are identified serially or in parallel during Chinese reading a more complex topic to disentangle.

Recently, Li & Pollatsek (2020) have proposed the Chinese reading model (CRM) specifically developed to explain eye movement control during Chinese reading. Because this model was developed with the characteristics of written Chinese in mind, it much more naturally engages and accounts for issues inherent in this written orthography (e.g., word segmentation). The CRM is composed of a word identification module and an eye movement control module, both of which work interactively. In relation to the word identification module, Li and Pollatsek adopted the interactive activation framework (McClelland & Rumelhart, 1981), and the model assumes that word segmentation and identification occur as part of a unified process. All characters within the perceptual span (one to the left and three to the right of the fixated character) are processed in parallel and directly activate character units in a character-activation map, and these, in turn, activate all the possible associated words (though character/word position specificity is maintained). The word units compete, exerting mutual inhibition until a single "winner" results, and at this point, a word is identified and simultaneously segmented from the character stream to the right of fixation. Upon word identification, the eyes saccade forward in the text targeting the character beyond the right boundary of the word that has just been identified, at which point the competition starts afresh. In this way, lexical identification and word segmentation occur for each word, sequentially along the text, until all the words in a sentence are identified. The activation of word and character units and lexical identification of words drive the eyes forward through the text. Thus, the model is built around an "engine" that is the process of word identification. However, as it stands, the model does not allow Chinese readers to identify more than a single word at a time. That is to say, in its current form, the units over which visual, linguistic, and oculomotor control processes are operationalized are single, individual, words, meaning that this model does not have any mechanism to explain how words might be processed and identified in parallel. We will return to this issue later.

As we indicated earlier, recently, we have developed a new perspective in relation to issues of serialism and parallelism in respect of oculomotor control during reading which we have termed the multiconstituent unit hypothesis (MCU hypothesis). In developing the MCU hypothesis, we aimed to offer an account that may provide a solution to the current serialism versus parallelism impasse (Zang, 2019). In addition, we felt that the hypothesis might lend itself well to issues intrinsic to many nonalphabetic languages such as Chinese with word boundary ambiguity (see He et al., 2021). The MCU hypothesis rests on a very simple idea, namely, that frequently used linguistic units composed of more than a single word may be represented, and therefore identified, lexically as single representations (e.g., Conklin & Schmitt, 2008, 2012; Shaoul & Westbury, 2011; Siyanova, 2010; Siyanova-Chanturia et al., 2011; Titone & Connine, 1999; Wray, 2002; Wulff & Titone, 2014). As we have considered, parallel processing of multiple words is consistent with the parallel processing framework (e.g., the SWIFT model), but cannot be reconciled with the serial processing framework such as the E–Z Reader model. However, if units corresponding to multiple words (MCUs) may be represented lexically, and therefore, their constituent words processed and identified simultaneously, then any demonstration of parallel processing of the constituents of a MCU would remain compatible with any account in which serial processing occurs. Thus, according to the MCU hypothesis, lexical processing is operationalized serially and sequentially over adjacent lexical units in a sentence, and these units might be individual words, or they might be MCUs. Furthermore, lexical identification operates on the basis of the familiarity of the units with respect to stored lexical representations (some of which are MCUs). In this way, the MCU hypothesis offers a potential explanation regarding why on some occasions during processing lexical identification appears to operate serially (with words being processed individually, one by one), and on other occasions, words appear to be identified in parallel.

Cutter et al. (2014) provided evidence that English spaced compounds (e.g., *teddy bear*) operate as MCUs during English reading. In their experiments, participants were required to read sentences containing spaced compounds composed of two frequently co-occurring constituent words, and the preview of each constituent (visible as an identity vs. masked by a nonword) was manipulated by the classical gaze contingent boundary paradigm (Rayner, 1975), with the invisible boundary located prior to the first constituent of the compound. The preview was replaced by the target word once readers' eyes crossed the boundary. Using this paradigm, it is possible to determine the extent to which parafoveal words or constituents of spaced compounds are processed prior to fixation. Cutter et al. found a reliable word $n + 2$ preview benefit, but only when the preview of word $n + 1$ was visible. In other words, processing of the second constituent occurred only if the first constituent was present and thus licensed the processing of the whole spaced compound as a MCU. Cutter et al.'s results are entirely consistent with the MCU hypothesis, suggesting that such processing may be operational (under some circumstances at least) during alphabetic reading.

Given that the written Chinese language is dense, unspaced, extensively ambiguous with respect to word boundaries, and indefinite with respect to the lexical status of words, it represents an excellent testbed language in which to demonstrate MCU effects. Any such demonstration would provide further independent support for the MCU hypothesis and would illustrate generality of effects across languages with markedly different orthographic characteristics. Beyond this, evidence for MCU-based processing might also offer a potentially valuable explanation of how processing might occur given significant inter-reader disagreement on word boundaries in Chinese (He et al., 2021).

Recently, Zang et al. (2021) provided evidence that Chinese idioms with a two-character modifier and a one-character noun structure ("2 + 1" modifier and noun, or MN idioms; e.g., 乌纱帽, "乌纱" means *black gauze*, "帽" means *cap*, "乌纱帽" means *an official post*) are processed as MCUs during natural reading. In their Experiment 1, idioms and matched phrases with "2 + 1" MN structure and with identical modifiers were selected as target strings. The preview of the noun was manipulated using the boundary paradigm. They found a greater preview benefit effect for idioms than for matched phrases. To extend the findings of Experiment 1, in Experiment 2 they adopted a similar paradigm used by Cutter et al. (2014) and manipulated the preview of both the modifier and the noun of idioms and matched phrases. Again, they found a reliable preview benefit effect from the noun when the modifier was present, and this effect was more pronounced for idioms than for matched phrases. Both experiments provide compelling evidence that "2 + 1" MN idioms can be lexicalized and accessed as a single representation. Indeed, idioms with "2 + 1" MN structure are quite like long words due to the significant constraint the two-character modifier exerts over the subsequent one-character noun. However, it remains an empirical question as to whether other types of idioms (e.g., "1 + 2" MN idioms that have less constraint from their modifiers to nouns than those of the "2 + 1" MN idioms) and phrases are also parafoveally processed as a single unit during natural sentence reading.

In the present study, we investigated the extent to which upcoming words, idioms, and phrases are processed in Chinese reading. Specifically, in Experiment 1, we were interested to know whether two constituent target strings (two characters in length) that are a single word, two separate words, or a string that is potentially lexicalized as a MCU in Chinese might be processed differently during natural reading. The critical issue concerned whether patterns of eye movement behavior for MCU strings were more comparable to those for two-character target words than they were to those observed for matched two-character, two-word phrases. We explored whether any such processing differences occurred in relation to both parafoveal and foveal processing by using the boundary paradigm (Rayner, 1975). Thus, in Experiment 1, we manipulated the linguistic category of a Chinese two-constituent string (a word, a phrase, or a MCU all carefully matched) in a boundary paradigm experiment with the boundary situated immediately prior to the two-character target string. In order to extend the findings of Experiment 1, in Experiment 2, we similarly used the boundary paradigm to examine whether two constituent target strings (three characters in length) that are idioms with a one-character modifier and two-character noun structure, or a matched phrase, might be parafoveally processed differently during natural reading.

One final issue concerns the criteria based on which we selected the items for our categories of stimuli. In Experiment 1, with respect to the first two categories of stimuli, that is, words and phrases, it is important to note that there are several different linguistic definitions or categorizations regarding the concept of a word or a phrase in Chinese (see Packard, 2003). In the present study, words and phrases were identified according to a syntactic analysis of the Chinese language, and our categorization was then verified through strict prescreening procedures by specialist Chinese Language and Linguistics experts from Tianjin Normal University (for full details, see Method section). The syntactic definition of a word, as specified by Packard (2003), is currently the most common linguistic characterization, and most closely coincides with the commonly shared notion of what a "word" is for most Chinese native speakers (again, see Packard, 2003). Specifically, our word stimuli were selected such that they were the "smallest independently useable part of language or that part of the sentence that can be used independently" (Packard, 2003, p. 16). We adopted the following criterion to distinguish between a word and a phrase (composed of multiple constituents and which may, potentially, be a MCU): if the constituents of the string are free morphemes and if the meaning of the string as a whole is not specialized (and therefore derivable in a compositional fashion from the meanings of its constituents), then it is a phrase. However, if any constituents in the string are not free morphemes or the meaning of the string is specialized (i.e., not compositionally derivable), then it is a word. Thus, we ensured that the constituent characters in our target phrase strings were free morphemes. For example, 木 偶 meaning *puppet* is composed of 木 (meaning *wood*) and the character 偶. Here, only the first character is a free morpheme that can stand as an independent syntactic form (a word), while the second character is not (i.e., the second character cannot appear on its own as a word in the language). Thus, under this definition, the character string 木 偶 is a word, and could not be categorized as a phrase. Further, if a linguistic unit had a specialized meaning, it too was categorized as a word. For example, 白 菜 (meaning *Chinese cabbage*) consists of two characters, each of which can be a word on its own, and each of which is a morpheme. The first character and morpheme is a modifying adjective (白 meaning *white*) and the second is a noun (菜 meaning *vegetable*), but neither of the constituents can be substituted in order to maintain the two-character string's special meaning (i.e., the particular vegetable that is *Chinese cabbage*). Consequently, 白 菜 would be categorized as a word due to the lack of compositionality. In contrast, 白桶 (meaning *white bucket*) would be categorized as a phrase because it consists of two constituent characters that are free morphemes, and they are also two single-character words, a modifying adjective (白 meaning *white*) and a noun (桶 meaning *bucket*). Critically, the noun and adjective can be substituted (e.g., 红 桶 means *red bucket*, and 白 墙 means *white wall*), and as such, the character string conveys compositional meaning. Consequently, to reiterate, 白 桶 is categorized as a phrase (see Fan, 1981; Packard, 2003; Wang, 1995).

Our third category of target string, the MCUs, is extremely important, theoretically, to the present study. Over recent decades, there has been significant consideration of how sequences of words that occur quite often together within a language are represented and linguistically processed. Such word sequences have been referred to by a number of terms in the literature over the years (e.g., multiword sequences, prefabricated chunks, lexical bundles, to name but a few, see Siyanova, 2010; Shaoul & Westbury, 2011). Here we adopt the term MCU to avoid committing to the word as the unit of granularity by which the term may be applied. A very important point, one which may be self-evident, is that the present consideration of MCUs as units of language that may be represented and stored lexically is not a new idea that we have generated. As we have indicated, such ideas are well established within the literature (e.g., Conklin & Schmitt, 2008, 2012; Shaoul & Westbury, 2011; Siyanova, 2010; Siyanova-Chanturia et al., 2011; Titone & Connine, 1999; Wray, 2002; Wulff & Titone, 2014), and good evidence exists supporting the notion that MCUs may be represented lexically. However, this is not to say that our current approach is without novelty. In fact, the novelty that we offer in our theorizing here is in how we consider the role of lexicalized MCUs in relation

to the operationalization of foveal and parafoveal processing over them during natural reading as an explanation to account for why words appear to be processed serially and sequentially on some occasions, while on other occasions they appear to be processed in parallel.

There are many different forms of word sequences that have the potential to be MCUs and to be represented lexically: phrasal, prepositional, and multiword verbs, spaced compounds, modifier-noun combinations, idioms and proverbs, markers within discourse, collocations, binomials and conjoined binomials, real names and product names, etc. In Experiment 1, we elected to use frequently used two-character word pairs that formed a phrase. To return to our example above, we used stimuli like 白 墙 (*white wall*), an adjective-noun pair that occurs very frequently in Chinese, forming a very recognizable two-character unit, and thus, a target string that is an excellent candidate MCU. In Experiment 2, we elected to use frequently used Chinese idioms with a one-character modifier and a two-character noun ("1 + 2" MN idioms; e.g., 铁饭碗, "铁" means *metal*, "饭碗" means *rice bowl*, "铁饭碗" means *a secure job*) and matched phrases ("1 + 2" MN phrases; e.g., 铁衣架, "铁" means *iron*, "衣架" means *clothes hanger*). To be clear, the MCUs used in Experiment 1 and the idioms used in Experiment 2 were character strings composed of two words where each constituent word co-occurred very frequently in the language and together formed a linguistic unit familiar to native Chinese readers (cf. Li et al., 2009 in which idioms were considered as long words by default).

In both experiments, we kept the first constituent of our target strings identical across conditions while we manipulated the preview of their second constituent (identical or pseudocharacter preview) using the boundary paradigm with the boundary located before the target string. Each set of target strings was embedded in the middle of each sentence. The context preceding the target strings was identical and neutral. It should be noted that in Cutter et al. (2014), the spaced compounds had a fairly high transitional probability (.42), meaning that the first constituent appeared 42% of the time as part of the whole compound in a corpus (Frisson et al., 2005; McDonald & Shillcock, 2003). Furthermore, they were embedded in fairly predictive sentence contexts maximizing the chances of them being processed more efficiently in the parafovea. Specifically, the whole compound was 33% predictable from the preceding context up to the pretarget word, and the second constituent was 97% predictable from the preceding context up to and including the first constituent. Cutter et al. argued that the high predictability of a second constituent given the preceding context including the first, might contribute, or even may be required, for early parafoveal processing of a second constituent as part of a lexicalized MCU. In the present study, we more directly considered the potential role of predictability of the whole target string and of the second constituent given the preceding context, including the first, as well as the potential role of transitional probability (see Frisson, et al., 2005; McDonald & Shillcock, 2003) in relation to the lexical licensing process, and we therefore controlled for these effects experimentally and statistically (see Method section for more details). We predicted that if Chinese readers process a two-constituent word, or a MCU, or an idiom as a single lexical unit, then they should preprocess the second constituent to a greater degree (greater preview effect) than they process the second constituent of a matched phrase. We based this prediction on the findings of Cutter et al. (2014) and Zang et al. (2021).

Open Practices

All data sets, materials, and analysis scripts are publicly available at: https://osf.io/wk3pj/. None of the experiments were preregistered.

Experiment 1

Method

Participants

One hundred and forty-four students (mean age = 21 years, SD = 2.5; Male = 22, Female = 122) were recruited at Tianjin Normal University. They were all native Chinese speakers with normal or corrected-to-normal vision. They were the regarding the purpose of the experiment and signed an informed consent form before taking part in the experiment. All of them received monetary compensation for their participation.

Apparatus

Eye movements were recorded with an SR Research EyeLink1000 plus eye tracker with a sampling rate of 1,000 Hz. Viewing was binocular, but only the right eye was monitored. Sentences were presented on a 24-inch ASUS VG248QE monitor with a refresh rate of 144 Hz and a screen resolution of 1,920 × 1,080 pixels. Stimuli were presented in Song font in black on a white background. The viewing distance of the participant to the monitor was approximately 64 cm. At this distance, each Chinese character subtended approximately 1.1° of visual angle.

Materials and Design

We selected a set of 66 two-character words, two-character MCUs, and two-character phrase triplets. Within each triplet, the first constituent of the two-character target strings was identical (the mean number of strokes 9, $SD = 3$; and the mean frequency 270 per million, $SD = 327$), and the second constituents were matched on character frequency and number of strokes ($F < 1.52$, $p > .05$). All the target strings were rated on a five-point scale for their linguistic category (1 = *definitely a phrase*; 5 = *definitely a word*) by 45 junior or graduate students majoring in Chinese Language and Literature. The two-character words were most likely to be categorized as words (M = 4.12, $p < .001$), whereas both MCUs ($M = 2.05$) and phrases (M = 1.94) were rated as phrases (see Table 1, $p > .05$). The target strings were also rated on a five-point scale for their familiarity (1 = *very unfamiliar*; 5 = *very familiar*) by 45 university students who did not take part in the eye tracking experiment. The mean scores were higher for words ($M = 4.03$) and MCUs ($M = 3.98$) than phrases ($M = 2.43$, all $p < .001$), and there was no difference between the former two conditions ($p > .05$).

We constructed 66 sentence frames, with each set of target strings embedded in the middle of each sentence. The context preceding the target strings was identical and neutral (see Figure 1). All the sentences were pre-screened for naturalness and contextual predictability. For the naturalness norms, 45 university students who did not take part in the eye-tracking experiment were required to rate sentence naturalness on a five-point scale (1 = *very unnatural*, 5 = *very natural*), and there was no difference across the three linguistic category conditions ($F < 1$). For the predictability norms, a separate group of 30 participants was required to conduct a sentence completion task. Half of

Table 1
Statistical Properties for the Two-Character Words, MCUs, and Phrases

	Words		MCUs		Phrases	
Characteristics of the target string and its constituents	*M*	*SD*	*M*	*SD*	*M*	*SD*
Second character's frequency (per million)	168	148	189	167	164	144
Second character's strokes	8	2	8	3	8	2
Linguistic category rating	4.12	0.52	2.05	0.43	1.94	0.45
Target string's familiarity	4.03	0.45	3.98	0.40	2.43	0.50
Sentence naturalness	3.92	0.25	3.92	0.28	3.88	0.29
Target string's predictability	0.00	0.02	0.00	0.00	0.00	0.00
Second character's predictability	0.10	0.21	0.05	0.10	0.00	0.18
Transitional probability	0.02	0.02	0.01	0.02	0.00	0.00

Note. MCUs = multiconstituent units; *SD* = standard deviation.

them assessed the predictability of the target strings given the preceding sentences up to the targets, and the other half assessed the predictability of the second constituents given the preceding sentences up to and including the first constituents. The mean predictability of the target strings was very low, though the mean predictability of the second constituents based on the preceding context including the first constituents was slightly higher for words ($M = 0.10$), lower for MCUs ($M = 0.05$) and lowest for phrases ($M = 0.00$, all $p < .09$). We also computed the transitional probability for each linguistic category based on a published online corpus (http://ccl.pku.edu.cn:8080/ccl_corpus/), that is, the probability of the second constituent given the first using the equation (p[constituent2|constituent1] = frequency [constituent1, constituent2]/frequency[constituent1], Frisson, et al., 2005; McDonald & Shillcock, 2003). The probability of the first constituent appearing as part of words ($M = 0.015$) and MCUs (0.011) was slightly higher than that of phrases (0.00) within the corpus, but there was no difference between the former two conditions ($p > .05$). However, both the predictability of the second constituents and the transitional probability did not exert an influence on the results, see details in the additional analysis.

Using the gaze-contingent boundary paradigm (Rayner, 1975), the preview of the second constituent of the two-constituent target string was manipulated with the invisible boundary placed before the target string. When readers' eyes crossed the boundary, an identical or a pseudocharacter preview (i.e., an unrelated character that appears extremely rarely in the Chinese language, and which, consequently, participants categorized as a pseudocharacter in a prescreen test, see Zang et al., 2020) was replaced by the target character. The pseudocharacter previews did not share any of the radicals of the target characters, and the number of strokes of the pseudocharacter previews was matched with the targets in the three target category conditions. Thus, we adopted a 3 (Linguistic Category of the Target String: a two-character word, a frequently used two-character MCU, or a two-character phrase) × 2 (Preview of the Second Constituent: identity or pseudocharacter) within-participant design. We constructed six files, with each file containing 66 experimental sentences, 20 filler sentences without display changes, and eight practice sentences presented at the beginning of the experiment. The experimental conditions were rotated across files according to a Latin square, but sentences in each condition were presented randomly within a file. Each sentence was read only once by each participant. There were 30 comprehension questions requiring participants to answer correctly with a yes/no response.

Procedure

Upon arrival, each participant was presented with an information sheet and written consent form, then seated comfortably in front of

Figure 1
An Example of Sentences With the Target Strings and Previews Used in Experiment 1

Target string	Preview type	Sentence
Two-character word	Identity	映入眼帘的那片|**绿洲**使探险家们看到了生存的希望。
	Pseudocharacter	映入眼帘的那片|**绿墝**使探险家们看到了生存的希望。
Two-character MCU	Identity	映入眼帘的那片|**绿草**使探险家们看到了生存的希望。
	Pseudocharacter	映入眼帘的那片|**绿墝**使探险家们看到了生存的希望。
Two-character phrase	Identity	映入眼帘的那片|**绿烟**是为了舞台效果而设计出来的。
	Pseudocharacter	映入眼帘的那片|**绿墝**是为了舞台效果而设计出来的。

Note. The vertical line represents the position of the invisible boundary. As soon as readers' eyes crossed the boundary, the preview changed to the target character (the target strings are in bold in the example but were presented normally in the experiment). The English translation for the first two sentences is "The *oasis/green grass* in sight gave the explorers hope of survival" and for the third sentence it is "The *green smoke* in sight was designed for stage effects."

the eye tracker with their head against a chin and forehead rests to minimize head movements. At the start of the experiment, participants completed a calibration procedure that involved fixating each one of a horizontal array of three points in turn, until an average calibration error of below 0.25° was achieved. Once the calibration was successful, the sentences were presented in turn. Each trial started with a drift correction dot presented on the left side of the screen. Participants were instructed to fixate the dot, which would trigger the onset of a sentence with the first character replacing the dot. Then participants read the sentence for comprehension and pressed the response keys on the button box to terminate the display after they finished reading the sentence. When a comprehension question appeared after a sentence, participants gave Y/N answers to the questions by pressing the response keys. The experiment lasted approximately 20–30 min.

Power Analysis

As mentioned earlier, in a directly related study, Cutter et al. (2014) have reported a reliable modulatory effect of word $n+1$ availability on the preview effect for word $n+2$, and their effect size (Cohen's d) was 0.33 for first-pass reading times (gaze duration, GD). Based on Westfall (2015) and an average effect size of 0.45, the power of our current sample size (144 participants and 66 sets of target string triplets in total) is estimated to be between 0.825 ($d = 0.33$) and 0.977 ($d = 0.45$), that is to say, we have sufficient power to establish an effect of average size in our study.

Results and Discussion

Data from six participants were excluded from the analyses: four sets of data due to low comprehension accuracy (below 80%) and the other two due to participants making a large number of blinks during recording. For the remaining 138 participants (revised power estimate = 0.812–0.973), their overall comprehension rate was 94%, indicating that they read and understood the sentences. All fixations shorter than 80 ms or longer than 1,200 ms were discarded. Trials were removed due to the following reasons: (a) tracker loss or fewer than three fixations were made in total (0.1%); (b) blinks occurred during display changes or during a fixation on the target region, or a display change occurred in an early or delayed manner (13.7%); and (c) measures of eye movements were above or below three SDs from the participant's mean (1.3%). In total, we removed 15.1% of the data prior to conducting the analyses.

We carried out analyses for the first character, the second character, the whole two-constituent character string, as well as the pretarget word. For each region, we computed the following eye movement measures: *first fixation duration* (FFD, the duration of the first fixation on a region, regardless of how many fixations it received during first-pass reading), *single fixation duration* (SFD, the fixation duration when only one first-pass fixation was made on the region), *gaze duration* (GD, the sum of all first-pass fixations on a region before moving to another region), *go-past time* (Go-past, the sum of all fixations on a region from the eyes first encountering the region until they leaving it to the right, including the time spent rereading earlier regions and time spent rereading the region itself), *total fixation duration* (TFD, the sum of all fixations on a region), and *skipping probability* (SP, the proportion of times a region is not fixated during first pass reading). Eye movement measures across all regions for each category of target string and preview are shown in Table 2.

To analyze the data, we conducted linear mixed models (LMMs) using the lme4 package (Version 1.1-12) in R (R Development Core Team, 2014). As fixed factors, we included the Target String and Preview conditions and their interaction. To examine differences between target string conditions, successive contrasts were conducted, comparing the word with the MCU and the MCU with the phrase. Participants and items were entered as crossed random effects. We started with running a model with the maximal random effects structure (Barr et al., 2013), but trimmed this down if the maximum random model did not converge. Fixation times were analyzed using log-transformed data to increase the normality, though analyses for untransformed and log-transformed durations yielded similar pattern of effects. SP was analyzed using logistic generalized linear mixed models, given the binary nature of the variable. Fixed effect estimations for the eye movement measures across all regions are shown in Table 3. All data files and analysis scripts are available at: https://osf.io/wk3pj/.

The Pretarget Word (n)

There was a reliable preview effect in GD and SP (but no other measures) such that readers fixated the pretarget word for less time and skipped it more often when they had a pseudocharacter preview than an identical preview (all $|t|$ or $|z| > 2.10$). This was probably due to an incorrect preview attracting the eyes to it more rapidly than an identical preview. The interaction between the preview type and the difference between words and MCUs approached, but did not achieve, significance in Go-past ($t = -1.90$, $p = .06$) but was reliable in TFD ($t = -1.97$). However, comparative analyses of different conditions showed that despite the reliable interaction, preview effects for words and MCUs on the pretarget word were not reliable (all $|t| < 1.48$). Additionally, note that TFD is a late measure of processing that includes second pass fixations that occur after a word has been initially processed. To be sure that we did not miss any pretarget effects, we also considered whether there were any differences between experimental conditions for fixations made on the preboundary character (i.e., fixations closest to the target prior to the eyes crossing the boundary). These analyses produced nonsignificant differences (all $|t| < 1.13$). These findings indicate that the preview manipulation of the second character of the target region did not exert a reliable influence over processing of the pretarget words.

The First Character (n + 1)

For the first character analyses, there was a reliable effect of the preview on all fixation times (all $t > 3.15$) and a marginal effect on SP ($z = 1.79$, $p = .07$), with shorter fixations and slightly more skipping for the identical second character preview than the pseudocharacter preview. The difference between words and MCUs was not significant, and this difference did not interact with preview condition across all eye movement measures (all $|t|$ or $|z| < 1.12$). However, the difference between MCUs and phrases was reliable in TFD ($t = 3.48$), and more interestingly, it interacted with the preview condition significantly in GD ($t = -1.99$), and marginally in FFD ($t = -1.69$, $p = .09$) and TFD ($t = -1.73$, $p = .09$). The planned contrasts showed that the preview effect was reliable for MCUs (FFD: $b = 0.07$, $SE = 0.02$, $t = 3.66$; GD: $b = 0.09$, $SE =$

Table 2
Eye Movement Measures Across All Regions for Each Category of Target String and Preview

Analysis region	Phrase type	Preview	FFD	SFD	GD	TFD	Go-past	SP
The pretarget word (*n*)	Word	Identity	227 (56)	226 (55)	243 (77)	307 (139)	269 (115)	0.26 (0.21)
		Pseudocharacter	226 (55)	225 (56)	243 (76)	320 (150)	274 (116)	0.28 (0.21)
	MCU	Identity	228 (57)	227 (55)	249 (82)	328 (154)	282 (122)	0.26 (0.21)
		Pseudocharacter	222 (52)	220 (51)	241 (79)	315 (148)	267 (109)	0.28 (0.21)
	Phrase	Identity	224 (56)	222 (53)	245 (80)	325 (156)	276 (121)	0.26 (0.22)
		Pseudocharacter	220 (53)	220 (52)	235 (71)	328 (157)	263 (109)	0.29 (0.20)
The first character (*n* + 1)	Word	Identity	246 (56)	247 (56)	248 (57)	291 (106)	280 (94)	0.59 (0.19)
		Pseudocharacter	267 (76)	268 (77)	276 (83)	309 (124)	315 (121)	0.56 (0.24)
	MCU	Identity	250 (64)	250 (64)	252 (66)	280 (103)	287 (108)	0.58 (0.21)
		Pseudocharacter	271 (77)	271 (77)	281 (88)	307 (122)	319 (127)	0.57 (0.21)
	Phrase	Identity	259 (65)	259 (67)	265 (73)	315 (139)	302 (114)	0.57 (0.21)
		Pseudocharacter	271 (74)	272 (74)	281 (81)	322 (136)	319 (124)	0.55 (0.21)
The second character (*n* + 2)	Word	Identity	254 (68)	254 (68)	262 (77)	303 (124)	298 (121)	0.52 (0.21)
		Pseudocharacter	276 (86)	279 (87)	287 (94)	322 (139)	356 (142)	0.46 (0.22)
	MCU	Identity	259 (73)	259 (74)	266 (81)	303 (122)	314 (129)	0.53 (0.20)
		Pseudocharacter	277 (82)	279 (83)	286 (88)	319 (131)	344 (132)	0.45 (0.21)
	Phrase	Identity	277 (85)	279 (87)	294 (100)	349 (160)	357 (159)	0.47 (0.21)
		Pseudocharacter	291 (93)	294 (92)	309 (105)	374 (157)	390 (169)	0.43 (0.23)
The whole region	Word	Identity	252 (68)	254 (67)	287 (102)	374 (182)	328 (153)	0.21 (0.20)
		Pseudocharacter	277 (81)	290 (81)	340 (120)	429 (187)	395 (177)	0.18 (0.21)
	MCU	Identity	257 (70)	261 (72)	295 (105)	364 (172)	338 (153)	0.22 (0.19)
		Pseudocharacter	277 (80)	283 (78)	343 (123)	430 (194)	387 (161)	0.17 (0.18)
	Phrase	Identity	267 (78)	271 (80)	328 (135)	462 (233)	386 (197)	0.19 (0.19)
		Pseudocharacter	280 (82)	293 (82)	367 (140)	504 (214)	430 (196)	0.17 (0.20)

Note. Standard deviations 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; MCU = multiconstituent unit.

0.02, $t = 4.49$; TFD: $b = 0.06$, $SE = 0.03$, $t = 2.28$) but not for phrases (FFD: $b = 0.03$, $SE = 0.02$, $t = 1.32$; GD: $b = 0.04$, $SE = 0.03$, $t = 1.56$; TFD: $b = 0.01$, $SE = 0.03$, $t = 0.38$, see Figure 2). It appears that during the period that the first constituent of the target was fixated, the linguistic category associated with the two-constituent string as a whole affected the extent to which the second constituent was preprocessed. Clearly, robust effects of the preview occurred when the second constituent alongside the first formed a word or a MCU, but not when it formed a phrase. This result indicates that MCUs, like words, are processed parafoveally as a single unit during reading. This in turn suggests that MCUs may be lexicalized.

*The Second Character (***n + 2***)*

There was a reliable preview effect in all eye movement measures such that readers fixated the second constituent of the target string for less time and skipped it more often when the preview was identical compared with when it was a pseudocharacter (all $|t|$ or $|z| > 4.54$). Furthermore, the difference between words and MCUs was not significant, and it did not interact with the preview conditions across all measures (all $|t|$ or $|z| < 1.38$). However, the difference between MCUs and phrases was reliable across all measures (all $|t|$ or $|z| > 2.66$) such that readers spent less time processing the second constituent when it, along with the first character of the target, formed a MCU compared to a phrase. There was a nonreliable interaction between MCUs versus phrases and preview conditions in SP ($z = 1.83$, $p = .07$). The planned contrasts showed a reliable preview effect for MCUs ($b = -0.35$, $SE = 0.08$, $z = -4.08$) but not for phrases ($b = -0.13$, $SE = 0.09$, $z = -1.50$). Once again, the numerical pattern associated with these results is consistent with the suggestion that the constituents comprising the MCUs were as easy to process as those comprising words, and both of these were easier to process than when the constituents formed a phrase. Again, these results lend support to the MCU hypothesis.

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

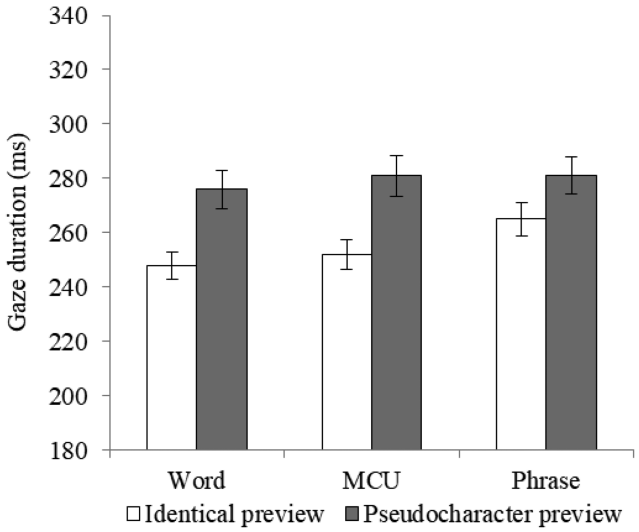
Again, there was a reliable preview effect in all eye movement measures such that readers fixated the two-character string for less time and skipped it more often when they had an identical preview rather than a pseudocharacter preview (all $|t|$ or $|z| > 4.28$). The difference between words and MCUs was not significant, and it did not interact with the preview condition across all measures (all $|t|$ or $|z| < 1.47$) other than an interaction that missed significance which occurred in Go-past ($t = 1.87$, $p = .06$). The planned contrasts showed reliable preview effects for both words ($b = 0.19$, $SE = 0.02$, $t = 8.02$) and MCUs ($b = 0.14$, $SE = 0.02$, $t = 5.67$), though with slightly larger effects for words than MCUs. Most importantly, the difference between MCUs and phrases was reliable in GD, Go-past, and TFD (all $t > 4.44$) with shorter reading times for MCUs compared to phrases. In addition, there was an interaction with preview condition in TFD ($t = -2.18$) such that the preview effect was greater for MCUs ($b = 0.15$, $SE = 0.02$, $t = 7.52$) than phrases ($b = 0.10$, $SE = 0.02$, $t = 4.59$). These results are similar in pattern to the effects that we observed for each separate character of the target, but they are more robust in the later measures (rather than the earliest measures) due to the larger target region (i.e., this is a coarser measure of processing) and the total time measure capturing the effect across all the fixations that were made on the target string.

Table 3
Fixed Effect Estimations for the Eye Movement Measures Across All Regions

		Measures																	
		FFD			SFD			GD			TFD			Go-past			SP		
Region	Effect	*B*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
The pretarget word (*n*)	Word versus MCUs	−0.00	0.01	−0.50	−0.01	0.01	−0.91	0.01	0.01	0.60	0.00	0.00	0.85	0.01	0.01	0.75	−0.01	0.06	−0.10
	MCUs versus phrase	−0.01	0.01	−1.23	−0.01	0.01	−0.84	−0.02	0.01	−1.53	0.00	0.00	0.61	−0.02	0.01	−1.45	0.03	0.06	0.53
	Preview type	−0.01	0.01	−1.65	−0.01	0.01	−1.65	−0.02	0.01	**−2.11**	−0.00	0.00	−0.04	−0.02	0.01	−1.85	0.12	0.05	**2.28**
	Word versus MCUs × preview	−0.02	0.02	−1.05	−0.02	0.02	−1.31	−0.02	0.02	−1.30	−0.01	0.00	**−1.97**	−0.04	0.02	−1.90	0.00	0.13	0.00
	MCUs versus phrase × preview	0.00	0.02	0.20	0.01	0.02	0.78	−0.00	0.02	−0.17	0.01	0.00	1.42	−0.01	0.02	−0.40	0.07	0.13	0.52
The first character (*n* + 1)	Word versus MCUs	0.01	0.01	0.48	0.01	0.01	0.46	0.01	0.01	0.46	−0.01	0.02	−0.27	0.01	0.02	0.45	0.01	0.06	0.09
	MCUs versus phrase	0.01	0.01	0.83	0.01	0.01	0.78	0.02	0.01	1.17	0.06	0.02	**3.48**	0.03	0.02	1.53	−0.08	0.06	−1.34
	Preview type	0.06	0.01	**4.19**	0.06	0.01	**5.15**	0.07	0.02	**4.81**	0.05	0.02	**3.16**	0.08	0.02	**5.41**	−0.10	0.06	−1.79
	Word versus MCUs × preview	0.01	0.03	0.25	0.00	0.03	0.12	0.01	0.03	0.19	−0.02	0.03	−0.64	−0.04	0.03	−1.11	0.13	0.12	1.10
	MCUs versus phrase × preview	−0.05	0.03	−1.69	−0.04	0.03	−1.60	−0.06	0.03	**−1.99**	−0.05	0.03	−1.73	−0.04	0.04	−1.11	−0.08	0.12	−0.64
The second character (*n* + 2)	Word versus MCUs	0.01	.01	0.84	0.01	0.01	0.81	0.01	0.02	0.84	−0.00	0.02	−0.11	0.02	0.02	1.37	0.00	0.06	0.04
	MCUs versus phrase	0.05	.01	**3.48**	0.06	0.01	**4.29**	0.07	0.01	**4.68**	0.12	0.02	**5.64**	0.10	0.02	**5.81**	−0.15	0.06	**−2.67**
	Preview type	0.06	.01	**4.72**	0.07	0.01	**6.22**	0.07	0.01	**5.53**	0.07	0.02	**4.55**	0.12	0.01	**8.38**	−0.24	0.05	**−4.73**
	Word versus MCUs × preview	−0.01	0.03	−0.36	−0.01	0.03	−0.45	−0.01	0.03	−0.20	0.01	0.03	0.28	−0.01	0.03	−0.22	−0.09	0.12	−0.80
	MCUs versus phrase × preview	−0.02	0.03	−0.88	−0.03	0.03	−1.02	−0.02	0.03	−0.77	0.02	0.03	0.56	−0.04	0.03	−1.18	0.21	0.12	1.83
The whole region	Word versus MCUs	0.01	0.01	0.95	0.01	0.01	0.84	0.02	0.02	1.29	0.00	0.02	0.01	0.01	0.02	0.76	−0.03	0.08	−0.41
	MCUs versus phrase	0.02	0.01	1.55	0.02	0.01	1.56	0.07	0.02	**4.45**	0.18	0.02	**9.03**	0.09	0.02	**5.24**	−0.09	0.08	−1.09
	Preview type	0.07	0.01	**5.79**	0.09	0.01	**7.86**	0.13	0.01	**10.17**	0.13	0.02	**8.23**	0.15	0.01	**10.10**	−0.27	0.06	**−4.29**
	Word versus MCUs × preview	−0.02	0.02	−0.78	−0.03	0.02	−1.46	−0.02	0.02	−0.86	0.01	0.03	0.49	−0.05	0.03	−1.87	−0.13	0.15	−0.83
	MCUs versus phrase × preview	−0.03	0.02	−1.53	−0.02	0.02	−0.75	−0.04	0.03	−1.39	−0.06	0.03	**−2.18**	−0.02	0.03	−0.78	0.25	0.16	1.59

Note. Significant terms are marked in bold and marginal terms are underlined. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration; Go-past = go-past time; SP = skipping probability; *b* = regression coefficient; SE = standard error; MCUs = multiconstituent units.

Figure 2
Preview Effects for the First Constituents of Different Target Strings for Gaze Duration in Experiment 1 (Error Bars Represent Standard Errors of the Mean)



Additional Analysis

Though the predictability of the second constituents on the basis of the context including the first constituents was relatively low across the three target strings, we carried out an additional set of LMM analyses in which predictability was included as a centered continuous covariate to examine the possibility that it could contribute to our effects (see Table A1 in the Appendix A). There was an effect of predictability for the TFD on the second constituent and for SP on the whole region with longer reading times and less skipping when the second constituents were less predictable. The effect of predictability of the second constituent did not appear on the first constituent, that is a clear lack of a parafoveal-on-foveal predictability effect. However, all these analyses from the first character, second character, as well as the whole two-character string produced an identical set of results for our experimental variables, indicating that this variable did not cause our effects. Similarly, when the transitional probability for each category was also included as a covariate in the LMM analyses (see Table A2 in the Appendix A), there was an effect of transitional probability for the first and SFDs only on the second constituents with shorter times when the transitional probability was higher. However, all the results produced exactly the same pattern as those reported. It is very likely that transitional probabilities were too low to exert any influence on our results.

To summarize, our results in Experiment 1 are straightforward and consistent with the MCU hypothesis (Zang, 2019): For the first constituent analyses, the preview effect from the second constituent was reliable across all fixation time measures with shorter fixations for identical preview than pseudocharacter preview. Interestingly, as shown particularly for the GD measure, this effect was robust when the second constituent alongside the first formed a word or a MCU but not a phrase. Furthermore, reading times were shorter for words and MCUs than phrases when the second constituent was fixated. Finally, the analyses for the full two-constituent string showed similar patterns to the results for the first character analyses with increased and more pronounced preview effects for the TFD measure for words and MCUs compared to phrases. It is clear that the linguistic category of the two-constituent string affected the extent to which the second constituent was preprocessed prior to the eyes transgressing the invisible boundary. Specifically, when processing of the first constituent lexically licenses parafoveal processing of the second (i.e., the first constituent signals that a second constituent might likely be part of the entire lexical unit), a parafoveal preview benefit from the second constituent is observed. Further, when the first constituent does not provide such a signal, as when the two constituents form a phrase, then the extent to which the second constituent is parafoveally processed is reduced. The pattern of results across all three regions of analysis demonstrates very clearly that frequently used MCUs, like words, appear to be lexicalized and processed foveally and parafoveally as single units during reading. To extend the findings from Experiment 1, in Experiment 2, we investigated whether "1 + 2" MN idioms are also processed foveally and parafoveally as MCUs.

Experiment 2

Method

Participants

Ninety-two students (mean age 22 = years, $SD = 2$ years; Male = 12, Female = 80) at Tianjin Normal University who did not take part in Experiment 1 were recruited to participate in Experiment 2. They had the same characteristics and participated on the same basis as the participants in Experiment 1.

Apparatus

Sentences were presented on a 19-inch DELL CRT monitor with a refresh rate of 150 Hz and a screen resolution of 1,024 × 768 pixels. All other details were identical to Experiment 1.

Materials and Design

We selected a set of 76 three-character idioms and matched phrases with identical syntactic structure—a one-character modifier (Constituent 1) and a two-character noun (Constituent 2). Idioms were selected from the Chinese Idiom Dictionary (2009) and the Modern Chinese Idiom Standard Dictionary (2001). They were all figurative and clearly defined ($M = 88\%$, $SD = 13\%$) and rated familiar ($M = 4.1$, $SD = 0.5$ on a five-point scale, "1" = *very unfamiliar*, "5" = *very familiar*) in a prescreen rating study involving 16 participants who did not take part in the eye-tracking study. Each set of idioms and phrases shared the first constituent (i.e., the one-character modifier) and differed only in their second constituents (i.e., the two-character noun), with these being controlled for stroke complexity and word frequency (Fs < 2, Cai et al., 2010), see Table 4.

As in Experiment 1, each set of target strings was embedded in a corresponding sentence frame. The context preceding the targeting strings was identical and neutral (see Figure 3). All sentences were pre-screened for naturalness and predictability. The mean sentence naturalness assessed by a separate group of 32 participants (16 for each of the target strings) was 4.0 ($SD = 0.4$), with no difference between idioms and phrases ($F < 1$). An additional group of 32

Table 4
Statistical Properties for the Idioms and Phrases in Experiment 2

Preview	Property of Constituent 2 (the two-character noun)	Idioms	Phrases
Identity	The first character's stroke number	8.3 (3.5)	8.2 (3.3)
	The second character's stroke number	6.8 (3.7)	7.1 (3.3)
	The Constituent 2's stroke number	15.1 (4.1)	15.3 (4.0)
	The Constituent 2's frequency (per million)	17.4 (33.7)	17.3 (36.6)
Pseudocharacter	The first character's stroke number	8.5 (2.8)	8.5 (2.8)
	The second character's stroke number	6.9 (3.9)	6.9 (3.9)
	The Constituent 2's stroke number	15.4 (3.7)	15.4 (3.7)

Note. Standard deviations are provided in parentheses.

participants was required to assess the predictability of the target strings given the preceding context up to the targets (16 participants) and the predictability of the second constituent given the preceding sentence up to and including the first constituent (16 participants). The target strings were unpredictable from sentence context ($M = 0.01$, $SD = 0.09$), whereas the second constituents of idioms ($M = 0.10$, $SD = 0.19$) were more predictable than those of phrases ($M = 0.001$, $SD = 0.007$, $F = 22$). As in Experiment 1, the transitional probability of the second constituent given the first for idioms and phrases was 0.006 ($SD = 0.041$) and 0.00 ($SD = 0.00$), respectively, with no difference between the two target strings ($F = 1.8$).

The preview of the second constituent (the two-character noun) of idioms and phrases was manipulated to be an identity or a pseudocharacter using the boundary paradigm (Rayner, 1975). Hence, Experiment 2 was a 2 (Phrase Type: idiom or phrase) × 2 (Preview of the Second Constituent: identity or pseudocharacter) within-participant design. The invisible boundary was directly located prior to the target string (the phrase or the idiom), and the identity or a pseudocharacter preview was replaced by the target once the eyes crossed the boundary (see Figure 3). In addition, the stroke complexity of identical and pseudocharacter previews was controlled across different conditions (all $F < 2$). We constructed four files, with each file containing 76 experimental sentences, 38 filler sentences (without display changes), and eight practice sentences presented prior to the formal experiment. One-third of the sentences were followed by yes/no questions. Conditions were rotated across files according to a Latin Square design, and each participant read experimental sentences presented randomly from one of the four files.

Procedure

The procedure was identical to Experiment 1.

Power Analysis

As in Experiment 1, the power of our current sample size in Experiment 2 (92 participants and 76 sets of target string triplets in total) is estimated to be between 0.816 ($d = 0.33$) and 0.974 ($d = 0.45$), indicating that we have sufficient power to establish an effect of average size in our study.

Results and Discussion

The mean comprehension accuracy was high ($M = 96\%$), indicating that all participants understood the sentences. The same data exclusion criteria were used as in Experiment 1. All fixations shorter than 80 ms or longer than 1,200 ms were excluded from the analyses. Trials were removed if (a) track loss occurred or fewer than three fixations were made (0.2%); (b) blinks occurred during display changes or during a fixation on the target word, or a display change triggered early or late (18.6%); and (c) measures were above or below three standard deviations from each participant's mean (1.4%).

As in Experiment 1, we carried out analyses for the same measures on the pretarget word (n), the first constituent (the one-character modifier, $n + 1$), the second constituent (the two-character noun, $n + 2$), and the whole target string. The means and SDs are shown in Table 5. LMMs were again conducted to analyze the data using the lme4 package (Version 1.1-21) in R. As fixed factors, we included the phrase type, preview type, and their interaction. As random factors, we included participants and items. For all measures, models with maximum random effects structure were conducted, allowing both random intercepts and random slopes for

Figure 3
An Example of Sentences Used in Experiment 2

Phrase type	Preview type	Sentence
Three-character idiom	Identity	张丽丽非常需要|**铁饭碗**来维持稳定的生活。
	Pseudocharacter	张丽丽非常需要|**铁邮戛**来维持稳定的生活。
Three-character phrase	Identity	张丽丽非常需要|**铁衣架**来晾晒这件厚衣服。
	Pseudocharacter	张丽丽非常需要|**铁邮戛**来晾晒这件厚衣服。

Note. The target strings are in bold but were presented normally in the experiment. The vertical line represents the position of the invisible boundary. Once the eyes crossed the boundary, the preview of the second constituents changed to the target characters. The English translation for the sentence is "Lili Zhang really needs a *secure job* (literally meaning *an iron rice bowl*) to maintain a stable life/Lili Zhang really needs an *iron clothes hanger* to dry the thick clothes."

Table 5
Eye Movement Measures for All Regions Across the Two Experimental Conditions

Analysis region	Phrase type	Preview	FFD	SFD	GD	TFD	Go-past	SP
The pretarget word (*n*)	Idiom	Identity	210 (35)	210 (35)	228 (47)	324 (99)	287 (93)	0.36 (0.17)
		Pseudocharacter	213 (37)	213 (37)	227 (44)	334 (103)	282 (91)	0.36 (0.15)
	Phrase	Identity	214 (34)	212 (34)	228 (45)	331 (109)	281 (91)	0.37 (0.17)
		Pseudocharacter	212 (33)	211 (33)	226 (47)	346 (114)	282 (94)	0.37 (0.17)
The first constituent (*n* + 1)	Idiom	Identity	227 (41)	226 (40)	230 (43)	274 (63)	306 (113)	0.61 (0.17)
		Pseudocharacter	243 (48)	244 (52)	247 (55)	289 (67)	312 (119)	0.60 (0.18)
	Phrase	Identity	237 (40)	238 (39)	239 (41)	300 (79)	294 (102)	0.59 (0.17)
		Pseudocharacter	252 (55)	251 (54)	255 (58)	312 (80)	317 (107)	0.60 (0.17)
The second constituent (*n* + 2)	Idiom	Identity	226 (34)	227 (36)	246 (46)	339 (96)	310 (87)	0.22 (0.14)
		Pseudocharacter	266 (48)	265 (49)	321 (68)	435 (134)	420 (138)	0.15 (0.13)
	Phrase	Identity	253 (41)	252 (48)	303 (73)	471 (160)	402 (132)	0.17 (0.14)
		Pseudocharacter	278 (50)	280 (52)	341 (82)	536 (177)	483 (148)	0.11 (0.14)
The whole region	Idiom	Identity	229 (30)	230 (31)	294 (70)	445 (138)	376 (114)	0.05 (0.08)
		Pseudocharacter	269 (46)	285 (61)	400 (100)	566 (185)	508 (165)	0.04 (0.08)
	Phrase	Identity	247 (35)	249 (45)	367 (100)	614 (231)	478 (170)	0.04 (0.09)
		Pseudocharacter	277 (47)	297 (63)	446 (114)	706 (233)	587 (184)	0.04 (0.10)

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

both participants and items. The trimming procedure was the same as in Experiment 1. Fixed effect estimations for the eye movement measures across all regions are shown in Table 6.

The Pretarget Word (n)

There were reliable effects of phrase type and preview type on the pretarget word *n* in TFD, with shorter total times for idioms than phrases, and for identity than pseudocharacter previews. These effects are relatively late with respect to their time course, and given that none of the earlier measures showed robust effects, they very likely reflect processing associated with the integration of the word into sentential meaning. Presumably, integration is easier for well-known idioms than for less frequently occurring phrases, and presumably, words with an inconsistent relative to a consistent preview will receive more return fixations in order to verify and confirm lexical identification. No other effects were reliable.

The First Constituent (n + 1)

There was a reliable effect of phrase type in all fixation time measures except Go-past, such that readers spent less time processing idioms than phrases (all $t > 3.52$), replicating a processing advantage for idioms over matched phrases (Yu et al., 2016; Zang et al., 2021). Relative to the identity preview, readers spent significantly longer processing the first constituent $n + 1$ when the preview of $n + 2$ was a pseudocharacter rather than the target identity (all $t > 2.51$). This effect suggests a sensitivity to the orthographic characteristics of the preview. No other effects were reliable.

The Second Constituent (n + 2)

Both effects of phrase type and preview were reliable in all eye movement measures with less time and increased skipping for idioms than phrases (all $|t|$ or $|z| > 4.94$), and for identity than pseudocharacter previews (all $|t|$ or $|z| > 6.75$). More importantly, preview type reliably interacted with phrase type across all fixation time measures (all $|t| > 2.45$). Further analyses showed that the preview effect was more robust for idioms (FFD: $b = 0.13$, $SE = 0.01$, $t = 9.51$; SFD: $b = 0.14$, $SE = 0.02$, $t = 9.01$; GD: $b = 0.23$, $SE = 0.02$, $t = 13.00$; TFD: $b = 0.23$, $SE = 0.02$, $t = 11.08$; Go-past: $b = 0.28$, $SE = 0.02$, $t = 12.58$) than phrases (FFD: $b = 0.09$, $SE = 0.01$, $t = 6.37$; SFD: $b = 0.08$, $SE = 0.02$, $t = 5.34$; GD: $b = 0.12$, $SE = 0.02$, $t = 6.71$; TFD: $b = 0.15$, $SE = 0.02$, $t = 7.08$; Go-past: $b = 0.19$, $SE = 0.02$, $t = 8.82$, see Figure 4). These results indicate that readers parafoveally process the second constituent of idioms to a greater extent than phrases. This suggestion is clearly consistent with the MCU hypothesis, according to which highly recognizable strings such as idioms are represented as single lexical units in contrast with two-word phrases for which there are two separate lexical entries. Thus, the second constituent of an idiom is preprocessed substantially more than the counterpart constituent of a matched phrase with an identical first constituent. We note that readers do also obtain parafoveal preview benefit from the second constituent of phrases. However, the important point to note is that this effect is significantly reduced, indicating that readers processed the critical string as a whole to a far greater extent when it was an idiom compared to a phrase. To reiterate, in line with the MCU hypothesis, these results very consistently align with the results from Experiment 1 and Zang et al. (2021), such that the idioms with a one-character modifier and a two-character noun structure are processed parafoveally as a single unit, rather like single words, during Chinese reading.

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

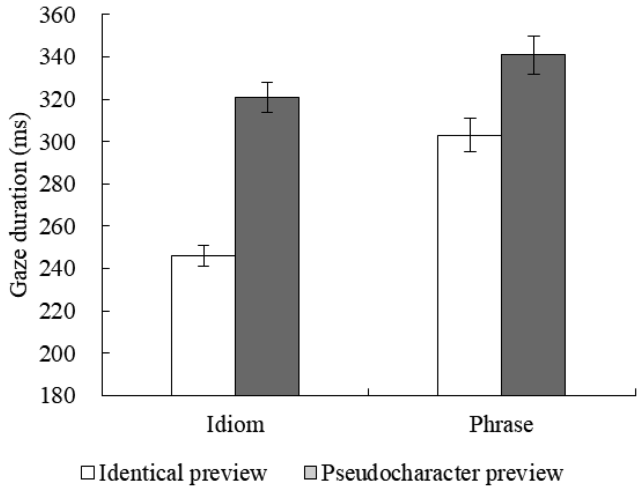
The whole target region was composed of both the first and second constituents. There was a reliable effect of phrase type in all fixation time measures such that readers spent less time processing idioms than phrases (all $t > 4.77$), replicating the effects observed from the individual first and second constituent analyses. These results demonstrate a processing advantage of idioms over matched phrases. Again, there was a reliable effect of preview type in all eye movement measures such that readers spent less time on the whole

Table 6
LMM Analyses for All Measures Across All Regions

		Measures																	
		FFD			SFD			GD			TFD			Go-past			SP		
Region	Effect	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
The pretarget word (*n*)	Phrase type	0.01	0.01	0.87	0.00	0.01	0.15	0.00	0.01	0.10	0.03	0.01	**2.39**	0.00	0.01	0.09	0.06	0.06	1.03
	Preview type	0.00	0.01	0.11	0.00	0.01	0.31	−0.01	0.01	−0.55	0.04	0.01	**2.68**	−0.01	0.01	−0.46	0.01	0.06	0.13
	Phrase type × preview type	−0.02	0.02	−1.17	−0.02	0.02	−1.20	−0.00	0.02	−0.17	0.01	0.03	0.23	0.02	0.03	0.89	0.05	0.11	0.44
The first constituent (*n* + 1)	Phrase type	0.05	0.01	**3.69**	0.05	0.01	**3.64**	0.05	0.01	**3.53**	0.08	0.02	**5.15**	0.00	0.02	0.19	−0.02	0.06	−0.27
	Preview type	0.07	0.01	**5.21**	0.07	0.01	**5.03**	0.07	0.01	**4.88**	0.05	0.02	**2.52**	0.07	0.02	**4.05**	−0.01	0.06	−0.23
	Phrase type × preview type	−0.01	0.03	−0.43	−0.02	0.03	−0.94	−0.02	0.03	−0.69	−0.00	0.03	−0.02	0.02	0.04	0.61	0.07	0.11	0.60
The second constituent (*n* + 2)	Phrase type	0.07	0.01	**7.25**	0.07	0.01	**6.73**	0.11	0.01	**7.82**	0.25	0.02	**14.04**	0.18	0.02	**9.30**	−0.37	0.08	**−4.95**
	Preview type	0.11	0.01	**10.13**	0.11	0.01	**10.00**	0.17	0.01	**12.27**	0.19	0.02	**10.80**	0.24	0.02	**13.64**	−0.65	0.10	**−6.76**
	Phrase type × preview type	−0.05	0.02	**−2.68**	−0.05	0.02	**−2.46**	−0.12	0.02	**−5.05**	−0.10	0.03	**−3.56**	−0.11	0.03	**−3.45**	−0.05	0.15	−0.34
The whole region	Phrase type	0.05	0.01	**5.81**	0.06	0.01	**4.78**	0.15	0.01	**12.25**	0.26	0.03	**10.22**	0.17	0.01	**12.48**	−0.26	0.24	−1.08
	Preview type	0.12	0.01	**11.16**	0.17	0.02	**11.04**	0.25	0.01	**17.00**	0.22	0.02	**11.22**	0.28	0.02	**16.54**	−0.28	0.14	**−2.06**
	Phrase type × preview type	−0.04	0.02	**−2.31**	−0.04	0.02	−1.65	−0.10	0.02	**−4.21**	−0.08	0.03	**−3.21**	−0.09	0.03	**−3.35**	0.24	0.28	0.86

Note. Significant terms are featured in bold. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration; Go-past = go-past time; SP = skipping probability; *b* = regression coefficient; SE = standard error.

Figure 4
Preview Effects for the Second Constituents of Idioms and Phrases for Gaze Duration in Experiment 2 (Error Bars Represent Standard Errors of the Mean)



target region and were more likely to skip the target string when the preview was an identity string rather than a pseudocharacter ($|t|$ or $|z| > 2.05$). The interaction between phrase type and preview type was also reliable in FFD, GD, TFD, and Go-past ($|t| > 2.30$). Similar to the second constituent analysis, the planned contrasts showed more pronounced preview effects for idioms (FFD: $b = 0.14$, $SE = 0.01$, $t = 11.73$; GD: $b = 0.30$, $SE = 0.02$, $t = 17.81$; TFD: $b = 0.26$, $SE = 0.02$, $t = 13.81$; Go-past: $b = 0.33$, $SE = 0.02$, $t = 13.92$) than phrases (FFD: $b = 0.10$, $SE = 0.01$, $t = 8.58$; GD; $b = 0.21$, $SE = 0.02$, $t = 11.99$; TFD: $b = 0.18$, $SE = 0.02$, $t = 9.62$; Go-past: $b = 0.24$, $SE = 0.02$, $t = 10.63$). These results are again consistent with the MCU hypothesis, suggesting that idioms are processed parafoveally (and foveally) as a single, whole, representation during Chinese reading.

Additional Analysis

As in Experiment 1, we undertook a further set of analyses in which predictability of the second constituent given the preceding context, including the first constituent, was included as a covariate in the LMMs (see Table A3 in the Appendix A). Again, there was an effect of predictability for reading time measures (except for the FFD) on the second constituent and the GD, TFD, and Go-past on the whole region with longer reading times when the second constituents were less predictable. However, all these analyses produced an identical set of results for our experimental variables, indicating that this variable did not cause our effects.

General Discussion

A central question regarding models of eye movement control in reading is whether multiple words are lexically processed serially or in parallel. In the present study, we employed eye tracking methodology to investigate whether frequently occurring MCUs composed of more than a single word might be lexically processed as single representations during reading (Zang, 2019). Evidence in support

of this hypothesis might offer an explanation for why lexical processing appears to operate serially on some occasions, but in parallel on others during reading. Specifically, in Experiment 1, we manipulated the linguistic category of two-constituent Chinese character strings (word, frequently used MCU, or a phrase), and the preview of its second constituent (identical or pseudocharacter) using the boundary paradigm with the boundary located prior to the target string. Similarly, in Experiment 2, we manipulated the linguistic category of two-constituent, but three-character Chinese strings ("1 + 2" MN idiom or "1 + 2" MN phrase), and the preview of its second constituent (identical or pseudocharacter). In line with the MCU hypothesis (Zang, 2019), we predicted that if frequently used two-constituent MCUs or idioms are processed as lexical units, then increased preview effects should be observed for the second constituent of the MCUs or idioms compared with the second constituent of otherwise matched phrases. Our results from both experiments are very straightforward and entirely consistent with our prediction.

The results of both of these experiments are also completely consistent with, and provide an important extension of, the findings of Cutter et al. (2014) who showed that English spaced compounds (e.g., *teddy bear*) operate as MCUs in reading. For such linguistic units, preview effects associated with the second constituent (e.g., *bear*) only occurred when the first constituent (e.g., *teddy*) was parafoveally available to license processing of the spaced compound (i.e., MCU) as a single unit. Note, though, as discussed in the Introduction, in the Cutter et al. study, the whole compound was fairly predictable from the preceding context up to the pretarget word, and the second constituent was very predictable given the preceding context including the first word of the spaced compound. It was argued that the high predictability of a second constituent on the basis of the preceding context including the first, might contribute, or even may be required, for early parafoveal processing of a second constituent as part of a lexicalized MCU. However, in the present study, we controlled for these potential effects of predictability and transitional probability (an alternative type of predictability index) experimentally and statistically. Recall that all of our Chinese target strings in both experiments were unpredictable from the preceding context, and also that the predictability of the second constituent given the preceding context, including the first, was slightly higher in MCUs or idioms relative to the counterpart phrases. However, we undertook analyses in which we accounted for variance associated with the predictability of the second constituent given the first and the inclusion of covariates in our LMM analyses did not change the nature of the effects (see Tables A1–A3 in the Appendix A). Thus, the results demonstrate influences beyond local predictability relations with respect to the degree to which nonadjacent parafoveal constituents are processed. To be clear, it appears that lexical processing in Chinese can be operationalized over linguistic units that are larger than an individual word, that is MCUs, and that the operationalization of such processing is not licensed (or at least not entirely licensed) on the basis of predictability (for more discussions, see Zang et al., 2021).

It is important to note that there has been other relevant research examining parafoveal processing across different lexical constituents in reading of Chinese monomorphemic words (e.g., 玫瑰 meaning *rose*), compound words (e.g., 灯塔 meaning *beacon*) and phrases (e.g., 斜塔 meaning *leaning tower*) using the boundary paradigm (Cui et al., 2013). Note both monomorphemic words and compound words were defined as words but not phrases in this study. Cui et al. manipulated the preview of the second character of a two-character target string, and placed the invisible boundary between the two characters (rather than before the whole target string as per our study). Their basic finding was that fixation durations were longer on the first character when a pseudocharacter relative to an identity preview of the second character was present and that this effect only occurred for monomorphemic words but not for compound words or phrases. At first sight, this result is only partially consistent with the present results and those of Cutter et al., and therefore, it is not entirely supportive of the MCU hypothesis that we advocate. While increased preview effects for the second character of monomorphemic words would be expected, it is also the case, according to the MCU hypothesis, that such effects should occur for the compound words. Compound words are very likely MCU candidates and should therefore be lexicalized (while matched phrases are not MCUs and should not be lexicalized). However, if we take a closer look at the stimuli from the Cui et al. experiment, it becomes clear that some of their characteristics may explain why their results patterned as they did. First, their compound words were less frequently used relative to those used in the present study, with an average occurrence of only 3.42 per million. Given the infrequency of occurrence, it is likely that such stimuli would not be accessed directly from the mental lexicon, but instead be processed in a nonunitary manner, with the constituents being processed serially and sequentially. Second, the familiarity of the compound words and the phrases to the participant population was not assessed. Again, on the assumption that the stimuli were less familiar to participants than were the stimuli adopted in the present study (for which very high familiarity ratings were obtained), then it is likely that they would not be represented as MCUs. Again, the frequency data for the Cui et al. stimuli are in line with this suggestion. Third, Cui et al. did not employ an a priori word segmentation pre-screen procedure to assess the extent to which participants considered character strings to form a single unit or to be composed of multiple separable units. Furthermore, the only post-screen assessment of participants' judgments showed that linguistic characterization of their stimuli was very ambiguous, with participants rating compounds as words 74% of the time and even phrases as words 45% of the time. Finally, and importantly, the second character's predictability and its transitional probability on the basis of the first were neither experimentally nor statistically controlled. Given the looseness with respect to the defining characteristics of the stimuli in the different experimental conditions in the study by Cui et al., it is perhaps not surprising that the compound words and phrases were processed similarly during reading.

In contrast, in the present study, we were very careful to take account of these variables and thus select our target stimuli strictly on the basis of their linguistic categorization and participants' assessments of them. We also captured extraneous transitional probability and predictability variance in our data with our statistical analyses. We ensured that our MCUs in Experiment 1 and our idioms in Experiment 2 were frequently used as common terms and we matched them with words with respect to their familiarity. When we undertook these procedures, in both our eye movement experiments, we obtained clear and robust results showing effects that patterned similarly for MCUs and words, while dissimilarly for matched phrases (Experiment 1), and greater second constituent preview effects for "1 + 2" MN idioms relative to matched phrases (Experiment 2). There is one important and slightly discrepant

point to note about the results of Experiment 1 relative to those from Experiment 2. This concerns the time-course over which the effects appeared and maintained in the two experiments. In Experiment 1, the interactive effects appeared quite rapidly in early fixations on the first constituent (marginal for FFD and robust for GD), were quite short-lived (marginal in TFD), and appeared most prominently for reading time measures associated with the first constituent. In contrast, In Experiment 2, interactive effects appeared first for reading times on the second constituent across early and late measures with a longer lasting time course. To be clear, the effects in Experiment 1 appeared earlier, were less substantive, and had a shorter time course than the effects in Experiment 2, which appeared later, were more substantive and were of increased duration. It is very likely that these alternative patterns of effects arose due to differences in the nature of the target stimuli in the two experiments. Of course, in Experiment 1, the stimuli were words and matched phrases, whereas in Experiment 2, the stimuli were idioms and matched phrases. However, more importantly from our perspective, the stimuli in Experiment 1 were composed of two characters, whereas those in Experiment 2 were composed of three characters. It is very likely that the requirement to process three compared with two constituent characters led to the delayed and more substantive effects that were observed in Experiment 2 relative to Experiment 1. This suggestion aligns with arguments put forward by He et al. (2021) and Zang et al. (2018) who each showed reading time costs associated with processing an increased number of linguistic constituents in character strings that were otherwise comparable. Nonetheless, it is certainly the case that further research is required to better understand the time course of reading time effects (i.e., how they emerge, their extent, and when they terminate) over linguistic constituents of different types that may also differ in length during Chinese reading.

Taking both experiments together, we consider that our results provide direct evidence in support of the MCU hypothesis. Furthermore, note that the two-constituent MCUs were composed of two single-character words (Experiment 1), and the "1 + 2" MN idioms were composed of a single-character word followed by a two-character word (Experiment 2). Since the boundary was positioned prior to the MCU or the idiom, then the preview effects in relation to the second constituent when fixations were made on the first constituent are very comparable to and quite consistent with the word $n + 2$ preview effects that have been reported in English (e.g., Cutter et al., 2014) and in Chinese reading (e.g., Yu et al., 2016; Zang et al., 2021).

Currently implemented eye movement control models do not straightforwardly account for our findings. According to E–Z Reader (e.g., Reichle et al., 1998), word identification occurs serially and sequentially, that is, the upcoming words are lexically processed only after the preceding words have been fully identified. Clearly, in the current study, when the two constituents formed a MCU, they were processed simultaneously, a finding that is inconsistent with the serial processing specification of the E–Z Reader as currently implemented. Of course, a modification to the model such that MCUs can be represented and processed lexically as single unified elements would result in a ready explanation for the effects reported here.

The processing of words in parallel does not appear to be an issue of difficulty for the SWIFT model of eye movement control (e.g., Engbert et al., 2002). Here, two words can be, and are argued to be, processed in parallel. However, the prior literature with regard to $n + 2$ preview effects demonstrates that ordinarily this occurs only when word $n + 1$ is a high frequency or function word (e.g., Yan et al., 2010; Yang et al., 2009). Note though that the first constituent of the target strings in our study was identical across the three (Experiment 1) or two (Experiment 2) experimental conditions, however, preview effects from the second constituent were more pronounced for MCUs (Experiment 1) or idioms (Experiment 2) compared with matched phrases, indicating that the linguistic category of the whole target string modulated serial versus parallel processing. To be specific, the determinant of whether constituents were processed serially, or in parallel, was whether the upcoming string formed a lexicalized unit. Thus, the results are inconsistent with a standard parallel account as specified by a model such as SWIFT.

Finally, it is worth noting that while the recently proposed CRM (Li & Pollatsek, 2020) is not currently implemented to explain the effects reported here, rather like the E–Z Reader model, if it was modified to allow for MCUs (as well as words) to be recognized as lexicalized units, then it could readily account for the current findings. It is not clear how the CRM model defines what a word is in Chinese. If it considers frequently used phrases and idioms to be long words (as corpuses do for segmentation convenience), then in our view, it appears that this account can explain MCU findings of the type we have obtained in Chinese reading. The critical issue in relation to this theory, in our view, relates to how the model determines whether a character string is composed of a single word or multiple words, and on our understanding, this will be determined by whether the string is or is not represented as a "word" in the mental lexicon.

In adopting the MCU perspective in relation to the operationalization of foveal and parafoveal processing and the computation of oculomotor commitments, the issue of serialism versus parallelism of lexical processing comes less to the fore. The theoretical issue of contention is no longer whether readers lexically identify one word or multiple words simultaneously during any particular fixation, but instead, how the lexical processing system treats upcoming constituents in respect of their lexical status. The word or words that are processed foveally and parafoveally during a fixation will be determined by whether those constituents are treated as individual, separate lexical elements, or instead as lexicalized MCUs. And saccadic computations will be made such that they are consistent with any such commitments. From this perspective, serial sequential processing of information within a sentence remains critical to incremental interpretation and the construction of a well-formed sentential interpretation. However, since constituents that occur early in a MCU license relatively immediate processing of subsequent constituents, then despite sequentiality and serialism of process, lexical identification can operate over two or more words simultaneously when those words comprise a lexicalized unit. According to this perspective, a key issue now becomes how a reader works out whether the next few upcoming words in a Chinese sentence are represented lexically, separately, and individually, or instead as a single lexicalized MCU (cf., the discussion of the CRM above). That is to say, a critical theoretical question concerns the factors that cause a character string to attain MCU status within the lexicon. Existing theoretical frameworks on language use and processing have provided some important pointers. For example, the Usage-Based theory (Bybee, 2006) proposes that if a word sequence is encountered sufficiently frequently, then it will develop lexical status and be lexically processed as a single unit. The Exemplar-Based theory (Bod, 2006) posits that whether a word

sequence is represented as a unit lexically is determined entirely by linguistic experience. Therefore, the frequency of occurrence within the language appears to be a central factor in relation to the question of lexicalization. And, again, we reiterate that several researchers have argued previously that frequently occurring multiple word units (or formulaic sequences, see Conklin & Schmitt, 2008, 2012; Shaoul & Westbury, 2011; Siyanova-Chanturia et al., 2011) can be lexicalized alongside individual words in the mental lexicon. To reiterate, however, the novel theoretical contribution we offer here is consideration of this possibility in relation to accounts of lexical processing and oculomotor control decision-making during natural reading. To us, it is increasingly apparent that models of eye movement control need to take into account the possibility that some linguistic units are composed of multiple words that may be processed and identified via a single lexical representation and that online saccadic computations will be made on this basis.

To summarize, results from two experiments showed that when the second constituent of a two-constituent Chinese character string is part of a MCU rather than a phrase, readers parafoveally preprocess it to a greater degree. These results support the hypothesis that Chinese readers lexically process highly familiar, recognizable MCUs foveally and parafoveally during reading. Earlier constituents seem to license processing of later constituents within those units (though not on the basis of predictability or transitional probability relations). Critically, our findings and theorizing offer potential for reconciling the impasse between the serial and parallel processing accounts of eye movement control during natural sentence reading.

References

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*(5), 1277–1287. https://doi.org/10.1037/0096-1523.34.5.1277

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*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, *23*(3), 291–320. https://doi.org/10.1515/TLR.2006.012

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language*, *96*, 9–22. https://doi.org/10.1016/j.jml.2017.04.001

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

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

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89. https://doi.org/10.1093/applin/amm022

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45–61. https://doi.org/10.1017/S0267190512000074

Cui, L., Drieghe, D., Yan, G., Bai, X., Chi, H., & Liversedge, S. P. (2013). Parafoveal processing across different lexical constituents in Chinese reading. *The Quarterly Journal of Experimental Psychology*, *66*(2), 403–416. https://doi.org/10.1080/17470218.2012.720265

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1778–1786. https://doi.org/10.1037/xlm0000013

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 839–856). 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*(5), 621–636. https://doi.org/10.1016/S0042-6989(01)00301-7

Fan, X. (1981). How to distinguish words and phrases of modern Chinese. *Dongyue Tribune* [in Chinese], *4*, 104–111.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 862–877. https://doi.org/10.1037/0278-7393.31.5.862

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*(3), 662–689. https://doi.org/10.1111/bjop.12482

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). North-Holland.

Huang, B. (2009). *Chinese Idiom Dictionary* [Hanyu Guanyongyu Cidian]. The Commercial Press.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*(2), 153–168. https://doi.org/10.1016/j.visres.2004.07.037

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Li, X. (2001). *Modern Chinese Idiom Standard Dictionary* [Xiandai Hanyu Guanyongyu Cidian]. Changchun Press.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, *143*(2), 895–913. https://doi.org/10.1037/a0033580

Li, X., Gu, J., Liu, P., & Rayner, K. (2013). The advantage of word-based processing in Chinese reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 879–889. https://doi.org/10.1037/a0030337

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, *127*(6), 1139–1162. https://doi.org/10.1037/rev0000248

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology*, *58*(4), 525–552. https://doi.org/10.1016/j.cogpsych.2009.02.003

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). Oxford University Press.

Liu, P., Li, W., Lin, N., & Li, X. (2013). Do Chinese readers follow the national standard rules for word segmentation during reading? *PLoS ONE*, *8*(2), Article e55440. https://doi.org/10.1371/journal.pone.0055440

Mcclelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. https://doi.org/10.3758/BF03203972

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*(6), 648–652. https://doi.org/10.1046/j.0956-7976.2003.psci_1480.x

Packard, J. L. (2003). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Rayner, K. (1975). The perceptual span and peripheral cues during reading. *Cognitive Psychology*, *7*(1), 65–81. https://doi.org/10.1016/0010-0285(75)90005-5

R Core Team. (2014). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, *13*(3), 115–119. https://doi.org/10.1016/j.tics.2008.12.002

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

Schotter, E. R., & Payne, B. R. (2019). Eye movements and comprehension are important to reading. *Trends in Cognitive Sciences*, *23*(10), 811–812. https://doi.org/10.1016/j.tics.2019.06.005

Schotter, E. R., Reichle, E. D., & Rayner, K. (2014). Re-thinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and $N + 2$ preview effects. *Visual Cognition*, *22*(3–4), 309–333. https://doi.org/10.1080/13506285.2013.873508

Shaoul, C., & Westbury, C. F. (2011). Formulaic sequences: Do they exist and do they matter? *The Mental Lexicon*, *6*(1), 171–196. https://doi.org/10.1075/ml.6.1.07sha

Siyanova, A. (2010). *On-line processing of multi-word sequences in a first and second language: Evidence from eye-tracking and ERP* [Ph.D. Thesis]. University of Nottingham.

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*(3), 776–784. https://doi.org/10.1037/a0022531

Snell, J., Declerck, M., & Grainger, J. (2018). Parallel semantic processing in reading revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition and Neuroscience*, *33*(5), 563–574. https://doi.org/10.1080/23273798.2017.1392583

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, *23*(7), 537–546. https://doi.org/10.1016/j.tics.2019.04.006

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

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. https://doi.org/10.1037/rev0000119

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, *31*(12), 1655–1674. https://doi.org/10.1016/S0378-2166(99)00008-9

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. https://doi.org/10.3758/s13423-016-1147-x

Wang, T. (1995). Ways to distinguish words and phrases. *Chinese Knowledge* [in Chinese], *5*, 12–14.

Westfall, J. (2015). *PANGEA: Power ANalysis for GEneral Anova designs* [Unpublished manuscript]. Retrieved from http://jakewestfall.org/publications/pangea.pdf

White, A. L., Boynton, G. M., & Yeatman, J. D. (2019). You can't recognize two words simultaneously. *Trends in Cognitive Sciences*, *23*(10), 812–814. https://doi.org/10.1016/j.tics.2019.07.001

Wray, A. (2002). *Formulaic language and the lexicon*. 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*(3), 371–376. https://doi.org/10.1075/ml.9.3.001int

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*(6), 1669–1676. https://doi.org/10.1037/a0019329

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*(4), 1192–1204. https://doi.org/10.1037/a0013554

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*(9), 1130–1149. https://doi.org/10.1080/23273798.2016.1197954

Zang, C. (2019). New perspectives on serialism and parallelism in oculomotor control during reading: The multi-constituent unit hypothesis. *Vision*, *3*(4), Article 50. https://doi.org/10.3390/vision3040050

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. https://doi.org/10.1037/xlm0000738

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2018). Investigating word length effects in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(12), 1831–1841. https://doi.org/10.1037/xhp0000589

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. https://doi.org/10.1016/j.jml.2021.104243

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). Oxford University Press.

(*Appendices follow*)

Appendix

Table A1

Fixed Effect Estimations for the Eye Movement Measures When Predictability of the Second Constituents Given the Preceding Context Including the First Constituents Was Included as a Covariate in Experiment 1

		Measures																	
		FFD			SFD			GD			TFD			Go-past			SP		
Region	Effect	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
The first constituent ($n + 1$)	Word versus MCUs	0.01	0.01	0.51	0.01	0.01	0.57	0.01	0.01	0.49	−0.00	0.02	−0.05	0.01	0.02	0.47	−0.00	0.06	−0.00
	MCUs versus phrase	0.01	0.01	0.84	0.01	0.01	0.87	0.02	0.01	1.18	0.07	0.02	**3.58**	0.02	0.02	1.50	−0.08	0.06	−1.33
	Preview type	0.06	0.01	**5.18**	0.06	0.01	**5.15**	0.07	0.02	**4.81**	0.05	0.01	**3.45**	0.08	0.02	**5.44**	−0.11	0.05	**−2.26**
	Predictability	0.01	0.05	0.25	0.03	0.05	0.65	0.01	0.05	0.18	0.07	0.07	1.06	0.00	0.06	0.03	−0.06	0.20	−0.27
	Word versus MCUs × preview type	0.01	0.03	0.19	0.00	0.03	0.13	0.01	0.03	0.19	−0.02	0.03	−0.63	−0.04	0.03	−1.11	0.13	0.12	1.09
	MCUs versus phrase × preview type	−0.05	0.03	−1.67	−0.04	0.03	−1.60	−0.06	0.03	**−1.99**	−0.05	0.03	−1.73	−0.04	0.04	−1.11	−0.07	0.12	−0.63
The second constituent ($n + 2$)	Word versus MCUs	0.01	0.01	0.68	0.01	0.01	0.57	0.01	0.01	0.57	−0.01	0.02	−0.40	0.02	0.02	0.87	0.00	0.06	0.08
	MCUs versus phrase	0.05	0.01	**3.48**	0.05	0.01	**4.06**	0.06	0.01	**4.46**	0.12	0.02	**5.34**	0.09	0.02	**4.33**	−0.15	0.06	**−2.63**
	Preview type	0.06	0.01	**4.69**	0.07	0.01	**6.20**	0.07	0.01	**6.26**	0.07	0.01	**4.88**	0.12	0.02	**7.41**	−0.24	0.05	**−5.11**
	Predictability	−0.05	0.04	−1.07	−0.05	0.04	−1.20	−0.08	0.05	−1.71	−0.13	0.07	**−2.05**	−0.11	0.06	−1.91	0.04	0.19	0.19
	Word versus MCUs × preview type	−0.01	0.03	−0.36	−0.01	0.03	−0.45	−0.00	0.03	−0.09	0.01	0.03	0.25	−0.01	0.03	−0.34	−0.09	0.12	−0.80
	MCUs versus phrase × preview type	−0.03	0.03	−0.93	−0.03	0.03	−1.01	−0.03	0.03	−1.15	0.02	0.03	0.58	−0.03	0.04	−0.93	0.22	0.12	1.88
The whole region	Word versus MCUs	0.01	0.01	0.91	0.01	0.01	0.80	0.02	0.01	1.55	−0.01	0.02	−0.30	0.01	0.02	0.49	−0.06	0.08	−0.81
	MCUs versus phrase	0.02	0.01	1.51	0.02	0.01	1.53	0.06	0.01	**4.49**	0.18	0.02	**8.75**	0.09	0.02	**5.02**	−0.11	0.08	−1.43
	Preview type	0.07	0.01	**5.79**	0.09	0.01	**6.54**	0.13	0.01	**10.16**	0.13	0.02	**8.22**	0.15	0.01	**10.08**	−0.27	0.06	**−4.28**
	Predictability	−0.00	0.03	−0.12	−0.00	0.04	−0.02	−0.06	0.04	−1.38	−0.14	0.08	−1.76	−0.09	0.05	−1.65	−0.60	0.28	**−2.16**
	Word versus MCUs × preview type	−0.02	0.02	−0.78	−0.03	0.02	−1.50	−0.02	0.02	−0.75	0.01	0.03	0.49	−0.05	0.03	−1.88	−0.13	0.15	−0.84
	MCUs versus phrase × preview type	−0.03	0.02	−1.53	−0.02	0.02	−0.79	−0.04	0.02	−1.57	−0.06	0.03	**−2.17**	−0.02	0.03	−0.77	0.25	0.16	1.59

Note. Significant terms are marked in bold. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration; Go-past = go-past time; SP = skipping probability; *b* = regression coefficient; SE = standard error; MCUs = multiconstituent units.

(*Appendices continue*)

Table A2

Fixed Effect Estimations for the Eye Movement Measures When Transitional Probability Was Included as a Covariate in Experiment 1

		Measures																	
		FFD			SFD			GD			TFD			Go-past			SP		
Region	Effect	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
The first constituent ($n + 1$)	Word versus MCUs	0.01	0.01	0.54	0.01	0.01	0.54	0.01	0.01	0.54	−0.00	0.02	−0.22	0.01	0.02	0.55	0.00	0.06	0.02
	MCUs versus phrase	0.02	0.01	1.05	0.02	0.02	1.05	0.02	0.01	1.44	0.07	0.02	**3.43**	0.03	0.02	1.46	−0.08	0.06	−1.35
	Preview type	0.06	0.01	**4.20**	0.06	0.01	**4.27**	0.07	0.02	**4.84**	0.05	0.02	**3.17**	0.09	0.01	**5.91**	−0.11	0.05	**−2.26**
	Transitional probability	0.32	0.38	0.83	0.38	0.38	0.99	0.40	0.38	1.05	0.27	0.54	0.51	0.20	0.47	0.42	−0.63	1.75	−0.36
	Word versus MCUs × preview type	0.01	0.03	0.25	0.00	0.03	0.17	0.01	0.03	0.19	−0.02	0.03	−0.64	−0.04	0.03	−1.24	0.13	0.12	1.09
	MCUs versus phrase × preview type	−0.05	0.03	−1.70	−0.04	0.03	−1.63	−0.06	0.03	**−2.00**	−0.05	0.03	−1.73	−0.04	0.03	−1.26	−0.07	0.12	−0.62
The second constituent ($n + 2$)	Word versus MCUs	0.01	0.01	0.67	0.01	0.01	0.42	0.01	0.01	0.68	−0.00	0.02	−0.13	0.02	0.02	1.18	−0.00	0.06	−0.05
	MCUs versus phrase	0.04	0.01	**2.88**	0.05	0.01	**3.32**	0.06	0.01	**4.02**	0.12	0.02	**5.21**	0.09	0.02	**4.94**	−0.17	0.06	**−2.85**
	Preview type	0.06	0.01	**4.68**	0.07	0.01	**6.14**	0.07	0.01	**6.27**	0.07	0.01	**4.91**	0.12	0.01	**8.37**	−0.24	0.05	**−5.12**
	Transitional probability	−0.78	0.35	**−2.19**	−0.08	0.04	**−2.19**	−0.72	0.37	−1.93	−0.48	0.53	−0.90	−0.89	0.47	−1.89	−1.50	1.55	−0.97
	Word versus MCUs × preview type	−0.01	0.03	−0.37	−0.01	0.03	−0.42	−0.00	0.03	−0.11	0.01	0.03	0.25	−0.01	0.03	−0.21	−0.09	0.12	−0.81
	MCUs versus phrase × preview type	−0.03	.003	−0.92	−0.03	0.03	−0.99	−0.03	0.03	−1.15	0.02	0.03	0.57	−0.04	0.03	−1.16	0.22	0.12	1.89
The whole region	Word versus MCUs	0.01	0.01	0.94	0.01	0.01	0.82	0.02	0.01	1.83	−0.00	0.02	−0.04	0.01	0.02	0.68	−0.04	0.08	−0.54
	MCUs versus phrase	0.02	0.01	1.48	0.02	0.01	1.50	0.07	0.01	**4.53**	0.18	0.02	**8.43**	0.09	0.02	**4.84**	−0.11	0.08	−1.37
	Preview type	0.07	0.01	**5.79**	0.09	0.01	**6.54**	0.13	0.01	**10.17**	0.13	0.02	**8.23**	0.15	0.01	**10.10**	−0.27	0.06	**−4.29**
	Transitional probability	−0.01	0.28	−0.02	0.02	0.31	0.06	−0.02	0.34	−0.06	−0.31	0.63	−0.49	−0.38	0.46	−0.82	−2.56	2.43	−1.05
	Word versus MCUs × preview type	−0.02	0.02	−0.78	−0.03	0.02	−1.50	−0.02	0.02	−0.75	0.01	0.03	0.49	−0.05	0.03	−1.88	−0.13	0.15	−0.83
	MCUs versus phrase × preview type	−0.03	0.02	−1.53	−0.02	0.02	−0.80	−0.04	0.02	−1.58	−0.06	0.03	**−2.17**	−0.02	0.03	−0.77	0.25	0.16	1.59

Note. Significant terms are marked in bold, and marginal terms are underlined. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration; Go-past = go-past time; SP = skipping probability; *b* = regression coefficient; SE = standard error; MCUs = multiconstituent units.

(*Appendices continue*)

Table A3

Fixed Effect Estimations for the Eye Movement Measures When Predictability of the Second Constituents Given the Preceding Context Including the First Constituents Was Included as a Covariate in Experiment 2

		Measures																	
		FFD			SFD			GD			TFD			Go-past			SP		
Region	Effect	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
The first constituent ($n + 1$)	Phrase type	0.04	0.01	**2.85**	0.04	0.01	**2.68**	0.04	0.01	**2.60**	0.08	0.02	**4.39**	0.00	0.02	0.06	−0.03	0.06	−0.46
	Preview type	0.07	0.01	**4.96**	0.07	0.01	**5.02**	0.07	0.01	**4.89**	0.05	0.02	**2.53**	0.07	0.02	**4.06**	−0.01	0.06	−0.23
	Predictability	−0.06	0.06	−1.14	−0.08	0.06	−1.40	−0.08	0.06	−1.33	−0.04	0.07	−0.50	−0.02	0.08	−0.24	−0.14	0.27	−0.51
	Phrase type × preview type	−0.01	0.03	−0.46	−0.02	0.03	−0.94	−0.02	0.03	−0.70	0.00	0.03	−0.03	0.02	0.04	0.61	−0.03	0.06	−0.46
The second constituent ($n + 2$)	Phrase type	0.06	0.01	**5.97**	0.06	0.01	**5.26**	0.10	0.02	**6.51**	0.21	0.02	**11.19**	0.16	0.02	**9.54**	−0.35	0.08	**−4.28**
	Preview type	0.11	0.01	**10.12**	0.11	0.01	**9.98**	0.17	0.01	**12.26**	0.19	0.01	**13.28**	0.24	0.02	**13.56**	−0.65	0.10	**−6.75**
	Predictability	−0.08	0.04	−1.78	−0.11	0.05	**−2.35**	−0.13	0.06	**−2.29**	−0.36	0.07	**−5.06**	−0.19	0.07	**−2.52**	0.25	0.30	0.84
	Phrase type × preview type	−0.05	0.02	**−2.68**	−0.05	0.02	**−2.44**	−0.12	0.02	**−5.06**	−0.10	0.03	**−3.57**	−0.11	0.03	**−3.45**	−0.05	0.15	−0.33
The whole region	Phrase type	0.05	0.01	**4.10**	0.05	0.01	**4.04**	0.13	0.01	**8.79**	0.23	0.01	**15.74**	0.15	0.02	**8.80**	−0.15	0.15	−1.02
	Preview type	0.12	0.01	**11.29**	0.17	0.02	**11.01**	0.25	0.01	**17.02**	0.22	0.02	**12.84**	0.28	0.02	**16.57**	−0.29	0.14	**−2.08**
	Predictability	−0.03	0.04	−0.68	−0.06	0.05	−1.17	−0.15	0.06	**−2.60**	−0.35	0.07	**−5.27**	−0.19	0.07	**−2.83**	−1.06	0.69	−1.53
	Phrase type × preview type	−0.04	0.02	**−2.18**	−0.04	0.02	−1.61	−0.10	0.02	**−4.26**	−0.08	0.03	**−3.25**	−0.09	0.03	**−3.39**	0.24	0.27	0.87

Note. Significant terms are marked in bold. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration; Go-past = go-past time; SP = skipping probability; *b* = regression coefficient; SE = standard error.

Received February 11, 2022
Revision received January 13, 2023
Accepted January 18, 2023 ■