Word n+ 2 preview effects in three-character Chinese idioms and phrases

Citation for published version:
Yu, L, Cutter, MG, Yan, G, Bai, X, Fu, Y, Drieghe, D & Liversedge, SP 2016, 'Word n+ 2 preview effects in three-character Chinese idioms and phrases', *Language, Cognition and Neuroscience*, vol. 31, no. 9, pp. 1130-1149. https://doi.org/10.1080/23273798.2016.1197954

Digital Object Identifier (DOI):
10.1080/23273798.2016.1197954

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Language, Cognition and Neuroscience

Word *n*+2 Preview Effects in Three-Character Chinese Idioms and Phrases

Lili Yu [1,2], Michael G. Cutter [2], Guoli Yan [1], Xuejun Bai [1], Yu Fu [1], Denis Drieghe [2], and Simon P. Liversedge[2]

[1] Tianjin Normal University

[2] University of Southampton

Author Note

Lili Yu, Guoli Yan, Xuejun Bai, and Yu Fu, Academy of Psychology and Behaviour , Tianjin Normal University, 241 Weijing Road, Tianjin, China. Michael G. Cutter, Denis Drieghe, and Simon P. Liversedge, School of Psychology, University of Southampton, Southampton, England, SO17 1BJ.

Correspondence regarding this article should be addressed to Simon P. Liversedge, School of Psychology, Shackleton Building, University of Southampton, Highfield Campus, Southampton SO17 1BJ. E-mail: S.P.Liversedge@soton.ac.uk.

Word Count: 13,593

Abstract

Prior research using the boundary paradigm suggests that Chinese readers only process word *n*+2 in the parafovea when word *n*+1 is a single character, high-frequency word. We attempted to replicate these findings (Experiment 1), and investigated whether greater *n*+2 preview effects are observed when word *n*+1 and *n*+2 form an idiom rather than a phrase (Experiment 2). Experiment 1 replicated prior findings, although additional analyses of word *n*+1 and *n*+2 as a single region revealed significant preview effects regardless of word *n*+1 frequency. In Experiment 2 there was a main effect of phrase type, such that idioms were read more quickly than phrases, and significant *n*+2 preview effects. There was no interaction between these variables, suggesting that idioms are not parafoveally processed to a greater extent than phrases. These results suggest that *n*+2 preview effects in Chinese occur under several circumstances. Factors influencing the observation of these effects are discussed.

Keywords: Parafoveal processing, eye movements, word *n*+2 preview effect, multi-word units, Chinese reading

During a single fixation in reading, processing is not confined to the fixated word. Rather, information is also extracted from words in the parafovea. In an alphabetic language such as English readers extract up to 3–4 characters to the left of fixation, and 14–15 characters to the right of fixation (McConkie & Rayner 1975; 1976). In a logographic, character based language such as Chinese readers extract up to 1 character to the left of fixation, and 3 characters to the right of fixation (Inhoff & Liu, 1998). This is known as the perceptual span. In both languages the rightward span is roughly equivalent to three words, including the currently fixated word. However, this represents the maximum amount of information that can be processed during a fixation, and does not account for the factors that influence the extent of parafoveal processing from one fixation to the next. The current article investigates some of these factors. Specifically, we investigate whether readers are more likely to process a word further into the parafovea when it and a preceding word together form a multi-word unit, which contains more than one word, but may have a unified meaning and therefore a single lexical entry.

Before proceeding it is important to outline two theoretical approaches to eye-movement control, and how they account for parafoveal processing. The first approach takes a serial view to lexical processing, and is embodied in the E-Z Reader model (Reichle, Rayner, & Pollatsek, 2003). In this model lexical processing is confined to one word, and attention does not shift to subsequent words until prior words have been identified. The amount of time it takes to identify a word is determined by its processing difficulty. Once a word is identified, attention shifts to the next word. Typically, there is a lag between attention shifting and a saccade being made, such that attention tends to arrive at the next word before the eye, resulting in parafoveal processing of this word. When this parafoveal word can be identified quickly, processing may also begin on the next word before a saccade is made. The second approach takes the view that multiple words are processed simultaneously, and is embodied in the SWIFT model (Schad & Engbert, 2012). All words within the perceptual span are processed in parallel, with the extent of the perceptual span (in characters) being determined by foveal processing difficulty. Thus, the number of words that are processed in a single fixation is determined by the length of each word due to visual acuity limitations, and the processing load of the foveal word. Essentially, the main controversies between the two approaches to eye movement control are whether lexical processing is carried out one word at a time, and whether the factors that influence the spatial extent of parafoveal processing are lexical or perceptual. Thus, investigating parafoveal processing of multi-word units—which technically consist of more than one word, but may have a unified lexical representation—has important implications for these contrasting theoretical approaches. To be clear, if the processing of parafoveal information was found to be influenced by whether upcoming words form a unit (i.e. have a corresponding lexical representation), then this might imply that multiple words may be processed simultaneously in the parafovea even though lexical processing progresses in a serial and sequential manner.

One way in which parafoveal processing has been investigated is using the boundary paradigm (Rayner, 1975). In this paradigm an invisible boundary precedes a target word in a sentence. Prior to the point of fixation crossing this boundary either a correct or incorrect parafoveal preview of the target word will be available. As the eye crosses the boundary this preview is replaced by the target word. Reduced fixation times on this word after receiving a correct, as opposed to incorrect preview, is referred to as *preview benefit.* Prior research has demonstrated that when the preview is of the word directly after the boundary (i.e. word *n*+1) then a preview benefit is reliably obtained (see Cutter, Drieghe, & Liversedge, 2015 for a review).

Studies have also investigated whether a preview benefit is obtained from a word two words to the right of an invisible boundary (i.e. word *n*+2). Typically, in alphabetic languages effects of a word *n*+2 preview are only observed when word *n*+1 is three letters long (Angele & Rayner, 2011; Kliegl, Risse, & Laubrock, 2007; Radach, Inhoff, Glover, Vorstius, 2013; Risse & Kliegl, 2011; 2012), but not when it is longer (Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Rayner, Juhasz, & Brown, 2007). This is in line with both E-Z Reader and SWIFT. In E-Z Reader it is easier to identify a short rather than long word *n*+1, allowing processing to begin on word *n*+2 prior to a saccade being made across the boundary. In SWIFT a short word *n*+1 will allow word *n*+2 to remain in the perceptual span, allowing it to be processed from word *n*. Thus, one factor determining the spatial extent of parafoveal pre-processing is the length of word *n*+1.

More recently Cutter, Drieghe, and Liversedge (2014) demonstrated that word *n*+2 can be parafoveally processed, despite being preceded by a long word *n*+1, when these two words form a single lexical unit. In this study words *n*+1 and *n*+2 after the boundary formed a specific type of multi-word unit—a spaced compound (e.g. *teddy bear*)—and each word had their previews manipulated to be either correct or a non-word. Crucially, word *n*+1 was on average 5.65 characters long, meaning that no *n*+2 preview effects were expected on the basis of prior literature. Nevertheless, when there was a correct preview of the first constituent of the spaced compound readers obtained a substantial *n*+2 preview benefit. Cutter et al. concluded that this was due to the first constituent licensing the processing of the second constituent as part of a lexically represented multi-word unit. Thus, a second variable that determines whether word *n*+2 is parafoveally processed in alphabetic languages is whether it forms a single unit with word *n*+1. In the current study we investigated whether a similar phenomenon occurs in Chinese reading; namely, that when word *n*+2 forms a multi-word unit with word *n*+1, parafoveal processing of word *n*+2 may occur to a larger extent than previously observed.

There are a number of differences between alphabetic languages and Chinese that are relevant to the study of parafoveal processing. Chinese is a logographic script, in that it consists of several thousand visually complex characters, rather than strings of letters. Each character occupies the same amount of space. A single character can be a word by itself approximately 20% of the time, combine with a second character to form a longer word about 70% of the time, or form an even longer word with more characters 10% of the time (Modern Chinese Frequency Dictionary, 1986). Thus, the space that a word takes up in Chinese is typically smaller than in an alphabetic language. Another difference is that there are no word spaces in Chinese, which further increases the density of Chinese text (Bai, Yan, Liversedge, Zang, & Rayner, 2008). Due to the greater spatial density of Chinese, words to the right of fixation are typically closer to the fovea, and are thus more available for parafoveal processing.

Despite the factors discussed above, prior research suggests that word *n*+2 is only rarely parafoveally processed in Chinese. Yang, Wang, Xu, and Rayner (2009) manipulated the previews of the second and third character after a boundary, which together formed a single word. This word was preceded by a single character high-frequency word, and significant preview benefits were observed. However, in a follow-up study in which word *n*+1 was controlled to be low-frequency, Yang, Rayner, Li, and Wang (2012) did not observe an *n*+2 preview effect. Yan, Kliegl, Shu, Pan, and Zhou (2010) observed a similar pattern of effects in a single experiment. These studies suggest that in Chinese, word *n*+2 is only parafoveally processed when word *n*+1 is high frequency. However, it may be that in Chinese, much like in English, words *n*+1 and *n*+2 can form a multi-word unit together. In this case word *n*+2 may be parafoveally processed regardless of the frequency of word *n*+1. One such possible multi-word unit that occurs in Chinese is an idiom, the processing of which is investigated in the current study.

In Chinese, an idiom is defined as a fixed phrase or short sentence that consists of several words, and has a unified meaning (Huang & Liao, 2007). Similar to the spaced compounds used in Cutter et al.'s study, idioms are made up of more than one word in a linguistic sense; however, all of the constituent words together express a single concept, and as such may have a single lexical entry. Specifically, an idiom has a figurative meaning, which is different from the literal meaning of its constituent words. For example, a literal reading of *kick the bucket* refers to somebody forcefully hitting a bucket with their foot. However, when used figuratively this idiom refers to somebody dying. It is unlikely that this figurative meaning could be computed by identifying and combining the individual words. Consequently, an idiom may also be stored and processed as a multi-word unit, and the meaning retrieved rather than computed. Several theories have addressed this possibility.

According to the lexical representation hypothesis (Swinney & Cutler, 1979) idioms are stored in the mental lexicon as long words. Upon encountering the first word of an idiom a reader will begin activating the whole unit's lexical entry. The configuration hypothesis (Cacciari & Tabossi, 1988) proposes that rather than being stored as single lexical units, idioms are stored as a configuration of their constituent words at a higher level of memory. As an idiom is encountered each word is lexically identified, and feeds activation up to this configuration. Once sufficient activation has built up, an idiomatic key is reached, leading to the retrieval of the idiom's figurative meaning. In this theory idioms are treated as larger units once the idiomatic key has been reached. Hybrid theories (Libben & Titone, 2008; Titone & Connine, 1999) propose that when idioms are encountered the figurative meaning is retrieved as a lexicalised unit, while at the same time each constituent word is also identified. The immediacy with which the whole idiom's lexical entry is retrieved is influenced by a number of factors, such as the frequency of the idiom's initial constituent word (Libben & Titone, 2008). Despite their differing assumptions, all three theories predict that, at least at some point during the identification of an idiom, it will be processed as a larger unit. Consequently, an idiom's constituent words may be processed in parallel to a greater extent than individual words that do not form an idiom.

A large amount of research in alphabetic languages has investigated the processing of idioms (e.g., *break the ice*) relative to novel phrases (e.g., *break the cup*). Findings suggest that idioms are processed faster than phrases, across several measures including response times in a phrase classification task (Swinney & Cutler, 1979), self-paced reading times (Conklin & Schmitt, 2008), and fixation times on final words (Underwood, Schmitt, & Galpin, 2004) and across the whole idioms (Siyanova-Chanturia, Conklin, & Schmitt, 2011). This processing advantage has been viewed as evidence that idioms are lexically represented. The logic behind this is that idioms are processed faster due to their meaning being directly retrieved, whereas a phrase's meaning has to be computed.

There are several reasons why idioms may be processed as a single unit in Chinese. As mentioned above, there are no word boundaries in Chinese text, and consequently readers must segment text into appropriate processing units online. However, the rules of how Chinese words are constructed are complex; a single Chinese character may be a word by itself or often combine with other characters to form a longer word, and thus ambiguity regularly exists amongst Chinese readers as to where word boundaries lie (e.g., Liu, Li, Lin & Li, 2013). Therefore, it is possible that since idioms convey a figurative meaning they might be segmented into and processed as a single word, despite being comprised of more than one word.

One prior study has used eye-tracking methodology to examine the processing of Chinese idioms. Carrol and Conklin (2015) presented four-character Chinese idioms, which were translated into English, to bilingual native Chinese speakers. It was found that the final word of these translated idioms received shorter fixations than the final word of a matched phrase. While this study suggests a processing advantage for the final parts of a Chinese idiom, it remains unclear whether such effects occur during the reading of non-translated Chinese idioms, and how the characteristics of Chinese text discussed earlier may affect this process. Also, the focus of the current study was on the processing of three- rather than four-character Chinese idioms. Furthermore, prior studies using tasks other than sentence reading tasks have also found that three-character idioms are processed more quickly than normal phrases. Gu and Liao (1995) asked participants to judge whether a linguistic unit was meaningful, and found that response times were faster for three-character idioms than for matched phrases. Similar results were observed in a naming task, such that idioms were named more quickly than matched phrases (Zhang & Shi, 2009). If we assume that this processing advantage is due to idioms being retrieved as multi-word units from the lexicon we may, arguably, expect their constituent words to be processed in parallel during reading. As such, the final word of an idiom may be processed to a greater extent in the parafovea than the same word in a novel phrase. If this were the case, we may observe similar preview effects in Chinese idioms as those observed by Cutter et al. (2014) in English spaced compounds.

However, it is worth noting that many of the aforementioned studies reporting these effects, in both English and Chinese, have not controlled the relative predictability of the final words of the idioms and the phrases. This is important since the models of eye-movement control discussed earlier attribute a large proportion of variance in reading times to a word's predictability in context. As such, the differences between the idioms and phrases observed in previous studies may be due to the high degree of predictability in idioms, rather than the idioms having a special representational status. Thus, a further aim of our study was to establish the idiom processing advantage in natural Chinese reading using eye-tracking methodology, while taking care to control the predictability of the later parts of our target items. To be clear, if we still observe an idiom processing advantage with the predictability of the idiom from the prior context controlled, the processing advantage has to be due to the special representational status of idioms; if, on the other hand, the idiom processing advantage is missing from the current study, it may indicate that within idiom predictability plays an important role in idiom processing.

Before proceeding, we feel it is necessary to explain why we chose to investigate idioms in the current study, in contrast to the spaced compounds used by Cutter et al. While we consider both idioms and spaced compounds to fall under the broader category of multi-word units, they are not without their differences. We will save an in-depth discussion of these differences, and the way in which they may have affected the results of the current study, until the General Discussion. However, for now it is important to note that the idioms used in the current study may differ from spaced compounds in several respects, such as the level of similarity between the meaning constructed from their constituent words and whole unit meaning (i.e. the level of decomposability[1]) and the amount of ambiguity this may cause during processing. Given that these differences exist, the question may arise as to why we used idioms rather than spaced compounds in the current study. At a practical level the answer is obvious; there is no such thing as a spaced compound in Chinese on account of the language being entirely unspaced. While the most similar Chinese item we could have chosen to test our suggestions about multi-word processing of would have been unspaced compounds, this would have been fairly uninteresting; compounds are generally considered as ‘words’ within Chinese rather than multi-word units (Huang & Liao, 2007), and as such would not be considered a compelling test of our hypothesis. Thus, we settled on idioms as the most appropriate linguistic unit in Chinese. In addition, at a theoretical level, we are of the opinion that the way in which different types of multi-word units have often been considered quite separately in the literature may well be to the detriment of overall theoretical understanding of the way in which the human cognitive system processes them. For the most part, separate literatures and theories have developed around the processing of idioms and compound words, despite similar issues being relevant to the processing of both. For example, the issue of decomposability of meaning has been investigated for both compounds and idioms, in addition to models being developed for both which allow for meaning retrieval by both a word-by-word and whole unit route (see Pollatsek, Hyönä, & Bertram, 2000 and Kuperman, Schreuder, Bertram & Baayen, 2010 for compounds, and Libben & Titone, 2008 for idioms). It does not seem overly bold to advance a position whereby we should expect to see these different units being processed in a similar manner, and in cases where this does not occur, consider differences in relation to overarching accounts of multi-word unit processing, rather than specific idiosyncratic theoretical accounts. As such, while we acknowledge that idioms are different to spaced compounds, we consider it worthwhile to examine whether they are processed in a similar manner, and, if not, to consider the reasons why.

To summarise, the primary aim of the current study was to investigate whether an *n*+2 preview benefit can be observed in Chinese when words *n*+1 and *n*+2 form an idiom together, despite the low-frequency of the first word. Furthermore, we were interested in examining whether there was a more general processing advantage for idioms over matched phrases when predictability is controlled. We placed three-character idioms and matched phrases directly after an invisible boundary. In both item types the first character formed a low-frequency word (and hence was word *n*+1 after the boundary) and the second and third characters together formed a word (and hence together were word *n*+2 after the boundary).

The preview of word *n*+2 was manipulated. Given that prior studies have not observed *n*+2 preview effects in Chinese when word *n*+1 was low-frequency, no main effects of our preview manipulation were expected. However, we did hypothesise an interaction, such that an *n*+2 preview effect would be observed in the idioms, but not the phrases. We made this hypothesis on the basis that idioms may have multi-word unit representational status, such that the first character of the idiom may licence the processing of the entire unit (idiom), thus leading to the early extraction of information from word *n*+2. In contrast the first character of a phrase should not licence the processing of a larger lexical unit (the phrase) and therefore the early extraction of information from word *n*+2. In this way the second constituent of an idiom would be processed in the parafovea despite the processing difficulty of the first constituent, in a similar manner to the second constituent of a spaced compound in Cutter et al.'s (2014) study. If this did occur, it would be strong evidence for the idioms being parafoveally processed as a multi-word unit. Finally, we hypothesised that the idioms should be read faster than the matched phrases, even with the predictability of word *n*+2 being controlled. In addition to investigating whether idioms are processed as single lexical units during reading, we also attempted to replicate the findings of Yang et al. (2009) and Yang et al. (2012) (i.e. word *n*+2 preview effects are only observed in Chinese text when word *n*+1 is highly frequent) in the same testing session as our main study. The primary reason for doing this was to establish that our participants were showing similar preview effects to those observed in prior studies.

Experiment 1

Experiment 1 was an attempt to replicate findings from previous studies on *n*+2 preview benefit in Chinese reading, using the same stimuli as Yang et al. (2009) and Yang et al. (2012). The reason for including this replication of "benchmark" *n*+2 preview effects was to be certain that parafoveal processing was occurring in the way that we would anticipate based on previous findings. For example, if we obtained null *n*+2 preview effects in both the idioms and matched phrases we wanted to be sure that it was due to the particular way the idioms were being processed rather than our participants simply not extracting information from word *n*+2 regardless of how favourable the circumstances were. Replicating Yang et al.'s (2009) effect would allow us to reject the latter explanation. Conversely, if we observed *n*+2 preview effects in both the idioms and phrases we wanted to ensure that it was due to some characteristic of our stimuli that differed from those used by Yang et al. (2012), rather than our participants extracting a greater deal of parafoveal information than those tested in prior studies. Replicating Yang et al.'s (2012) null effect would allow us to reject the latter explanation.

Method

Participants. 40 (20 males and 20 females) native Chinese speakers who were graduate students studying in England with an average age of 25 and 40 (10 males and 30 females) native Chinese speakers who were undergraduate students studying in China with an average age of 21 years with normal or corrected to normal vision participated.[2] Eighteen additional participants were tested, but excluded from the analysis due to noticing more than three display changes.

Apparatus. Participants from England were tested with an SR Research Eyelink 1000 Desktop mounted eye tracker with a sample rate of 1000 Hz. Head rests were used to reduce head movements. Participants from China were tested with an SR Research Eyelink 2000 Tower mounted eye tracker with a sample rate of 1000 Hz. The display monitors were 21" CRT monitors with a refresh rate of 120 Hz. The distance between the monitor and participant was 75cm. Sentences were presented on a black background, and each character occupied 0.9 degrees of visual angle. Participants' right eyes were tracked, although viewing was binocular. The setting of the monitors, the distance between the monitor and participant, and the stimuli display were exactly the same in both laboratories.

Materials and Design. 20 items with a high-frequency single-character first word and two-character second word from Yang et al.'s (2009) Experiment 2, and 20 items with a low-frequency single-character first word and two-character second word from Yang et al.'s (2012) experiment were used in the current experiment. These were interspersed with the 40 items from Experiment 2 and 20 filler items, and all items were preceded by 6 practice sentences.

The boundary paradigm (Rayner, 1975) was used. The boundary was located before the high- or low-frequency single-character word, and either a correct or incorrect preview of the two-character word *n*+2 was presented. The incorrect previews were pseudo-words, consisting of two real characters, with both characters different to the characters in the correct previews (see Figure 1). Although the stimuli in this experiment included both items with a high-frequency and low-frequency word *n*+1, these two subsets of stimuli were treated as two independent experiments, and were not manipulated as an independent variable. Therefore, analyses were carried out separately for items with high- and low-frequency word *n*+1. The characteristics of the stimuli are listed in Table 1.

INSERT TABLE 1 AND FIGURE 1 ABOUT HERE

Procedure. An information sheet and a consent form were presented to the participants prior to beginning the experiment. A three-point horizontal calibration grid was used, with an acceptance criterion of a maximum error below 0.40 degrees and an average error below 0.20 degrees. Prior to each sentence, there was a drift check in the center of the screen, followed by a gaze-contingent target that occupied half of a character on the first character of each sentence. The participant needed to gaze at the target for 20ms for the trial to begin. After the target was triggered, the sentence would appear on the screen, with the first character beginning in the same position as the gaze contingent target. If the participants failed to trigger the gaze contingent target (0.45 degrees of visual angle), or the participants' level of error on the drift correct increased beyond 0.9 degrees, they were recalibrated. Participants were recalibrated regularly, with each participant being recalibrated a minimum of twice.

For sentences presented with the boundary paradigm, before participants' eyes crossed the boundary, either a correct or incorrect word *n*+2 was displayed as a preview, and as soon as participants' eyes crossed the boundary, the incorrect word *n*+2 was replaced by the correct target in the text. The boundary change was programmed using SR Research Experiment Builder. The boundary triggered after a single sample was detected beyond it, and the delay for the display change to occur was on average 5.8 and 6.4ms in England and China, respectively. The difference between the two laboratories in terms of the time needed to implement the display change after the boundary was crossed was negligible (0.6ms) even though it was statistically significant ($t(8478) = 11.07$, $p < .001$).

Participants were instructed to read for comprehension. After reading each sentence the participant pressed a button to move on. One third of the sentences were followed by a yes/no comprehension question. The experiment took approximately 30 minutes.

Results

Across all participants an average of 94% of the questions were answered correctly. Linear mixed-effect models constructed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R (R Development Core Team, 2013) were used to analyse the data. Preview type was a fixed factor in the model, and contrasts were set up using the contr.sdif function in the MASS library (Venables & Ripley, 2002). This function sets up contrasts so that the output from the models represents main effects and the intercept corresponds to the grand mean. Items with a high-frequency word *n*+1 and items with a low-frequency word *n*+1 were analysed separately. Participants and items were treated as random factors, with both random intercepts and random slopes included in the models. Since two different groups of participants took part in the experiment, we also constructed models in which participant group was included as (i) a main effect, or (ii) both as a main effect, and allowed to interact with *n*+2 preview type. This was done to ensure no confounding effects of this factor. The fit of each of these models to the data was compared with the simplest model (i.e. the one including the main effect of word *n*+2 preview type alone). If one of the models including participant group had a significantly better fit to the data, this model is reported. Otherwise, the simplest model is reported.

Trials in which participants blinked in any of the critical regions (14%), or in which the display change triggered early (9%) or completed more than 10ms after fixation onset (4%) were excluded. Altogether, 27% of the data were excluded.

Analyses were carried out on four regions: word *n* (the word prior to the boundary); word *n*+1; word *n*+2; and a composite region consisting of both word *n*+1 and word *n*+2. The composite region was used due to the high skipping rates in the current study. Word *n*+1 was skipped 51% of the time in the current study, and on 49% of trials the composite region received a single, first-pass fixation. It has been proposed (e.g. Li, Liu, & Rayner, 2011) that Chinese readers may target saccades beyond the end of the currently fixated word, and upon landing attempt to identify as many characters as possible. If this is correct then it is likely that the effect of a preview manipulation on word *n*+2 would affect fixations across the entire three character composite region, rather than just fixations on word *n*+2. [3] Furthermore, the analysis across this region will retain greater statistical power than either individual region, due to less data being lost due to skipping. Due to these issues we consider this composite region to provide the most informative measure of our manipulation. It is worth noting that Yang et al. did not analyse a composite region, and so any effects observed in this region represent an extension, rather than replication, of their work.

For each region, five measures were computed. First fixation duration (FFD; the duration of the first fixation in an interest area); single fixation duration (SFD; the duration of a fixation when it was the only fixation in the interest area); gaze duration (GD; the sum of all fixations in a region from the first fixation until a saccade leaving the region); go-past time (GP; the sum of all fixations from the first fixation in a region until a saccade was made to the right of the region); skipping probability (the proportion of times that a region was skipped). The means and standard errors are shown in Table 2. The output from the models is displayed in Table 3. To increase the normality of the data, the fixation time measures were log-transformed. The skipping probability data were analyzed using logistic models due to the variable's binary nature.

INSERT TABLES 2 AND 3 ABOUT HERE

Word *n*, and Word *n*+1. There were no significant effects of the preview manipulation on fixation times on the pre-target word or word *n*+1, for either type of stimuli.

Word *n*+2. There were no significant effects of the *n*+2 preview when word *n*+1 was low-frequency. However, when word *n*+1 was high-frequency, a 22ms *n*+2 preview effect was observed in gaze durations. Also, when word *n*+1 was low-frequency, there was a significant effect of whether participants were tested in England or China in FFD, SFD and GD (Participants tested in China's means = 307, 311, 394ms for each measure; participants tested in England's means =290, 284, 342ms). This effect did not interact with the preview type, which means that regardless of whether fixation times were longer or shorter for a particular group of participants, word *n*+2 preview effects were observed when *n*+1 was high frequency, but not when *n*+1 was low frequency. Given the lack of interaction, we did not further analyse these two data sets separately.

Composite region. On the composite region significant effects of the preview type were observed regardless of word *n*+1 frequency. In the case of a high-frequency word *n*+1 this effect was reliable across all fixation time measures. When word *n*+1 was low-frequency this effect was only reliable in gaze duration. Furthermore, a main effect of participant group appeared in go-past time. Again, there was no interaction between group and word *n*+2 preview benefit.

Discussion

Experiment 1 was an attempt to replicate Yang et al.'s findings, and the analyses on the regions examined by Yang et al. (i.e., words *n*, *n*+1, and *n*+2) provided a full replication. Consistent with their findings, we observed *n*+2 preview effects of 22ms on word *n*+2 in gaze duration when word *n*+1 was high-frequency. Yang et al. (2009) found a marginal (*p*=0.051) effect in gaze duration of 22ms, and no effect in other measures. Furthermore, Yang et al. (2012) did not find any effect of the word *n*+2 preview when a low-frequency word *n*+1 was used. Thus, the pattern of effects was replicated in the current experiment, and our results are consistent with Yang et al.'s findings.

However, we also performed analyses on a composite region consisting of both word *n*+1 and word *n*+2, which Yang et al. did not examine. In this region we observed an *n*+2 preview benefit regardless of the frequency of word *n*+1. When word *n*+1 was high-frequency, there was an *n*+2 preview effect of 17ms, 31ms, 30ms, and 25ms in FFD, SFD, GD, and GP. When word *n*+1 was a low-frequency word, word *n*+2 preview benefit was only observed in gaze duration, the size of which was 24ms. The finding that *n*+2 preview effects are obtained even when word *n*+1 is a low-frequency word might be considered to be inconsistent with previous findings. However, it should perhaps not come as a surprise that we observed an effect of our preview manipulation in our composite region, and that these effects were more stable than in the individual word regions. As discussed above it is likely that the effect of a preview manipulation on word *n*+2 would affect fixations across the entire three character composite region, rather than just fixations on word *n*+2. Furthermore, the analysis across this region retained greater statistical power than either individual region.

In summary, the results of the current experiment replicated Yang et al.'s studies. Word *n*+2 preview effects were observed during fixations on word *n*+2 only when word *n*+1 was high-frequency. However, in a further analysis which Yang et al. did not conduct *n*+2 preview effects were observed for both types of stimuli. This suggests that while word *n*+2 preview effects were not entirely prevented by a low-frequency word *n*+1, they were certainly delayed and decreased relative to when word *n*+1 was high frequency. The *n*+2 effect only appeared in gaze duration in the analysis of the composite region when word *n*+1 was a low-frequency word, whereas it appeared across multiple measures and regions when word *n*+1 was high frequency. As such, the issue of whether word *n*+2 is processed differently when it is part of an idiom relative to novel phrase remains interesting. To be clear, it may be that in novel phrases similar *n*+2 preview effects will be observed as in the low-frequency items, while in idioms we may expect to see a similar pattern of effects to the high-frequency items.

Experiment 2

Experiment 2 was carried out to investigate parafoveal processing of Chinese idioms relative to matched phrases. We hypothesised that the reading times in idioms would be shorter than phrases. We also predicted that the parafoveal processing of an idiom's constituent words (word *n*+1 and *n*+2 after an invisible boundary) may be increased relative to parafoveal processing of a matched phrase's constituent words. To recapitulate, these idioms and phrases were selected in order to be as similar as possible to the items with a low-frequency word *n*+1 from Experiment 1 (254 per million in Experiment 1, 188 per million for idioms and 162 per million for phrases in Experiment 2), such that *n*+2 preview effects should be relatively modest in the phrases. However, due to the lack of visual word boundaries and the necessity of online word segmentation in Chinese, as well as the characteristics of Chinese idioms, it may be that such multi-word units are segmented and processed as single words. Was this the case, even with a low-frequency word *n*+1, *n*+2 preview effects should be observed in the idioms, but not the phrases.

Method

Participants. The same participants were tested as in Experiment 1 in the same testing session.

Apparatus. The same apparatus was used as in Experiment 1.

Materials and design. 40 verb-noun structure idioms and matched phrases with a low-frequency word *n*+1 were used in the present experiment. The syntactic structure, and the frequencies of word *n*+1 and word *n*+2 of the idioms and phrases were matched with the phrases used in Yang et al.'s (2012) study. The idioms were selected from the *Chinese Idiom Dictionary* (2009). 16 participants who did not take part in the eye-tracking study were asked to define the figurative meaning of the idiom and rate their familiarity on a 6-point scale. Idioms which were correctly defined by over 88% of the participants were selected, and the average familiarity of the selected idioms was 4.95. According to the *Modern Chinese Words Dictionary* (5th edition, 2005), word *n*+1 was always a single-character verb by itself, and the second and the third character within the idiom or phrase formed a two-character noun, in which the second character after the boundary was not typically processed as an individual single-character word.[4] Thus, word *n*+1 and *n*+2 after the boundary should have been processed as such.

As shown in Figure 2 and Table 4, the idioms and phrases only differed in their first character, with this character being controlled for stroke complexity and frequency. The preceding sentential context was the same for both idioms and phrases, with the predictability of the whole idioms and phrases from the sentence context being controlled. Moreover, the predictability of word *n*+2 between the idioms and phrases from the context including word *n*+1 was matched, as was the whole sentence naturalness. Each of these factors was rated by a separate group of 16 participants, none of who participated in the eye tracking experiment. Finally, the stroke complexity and frequency of the characters of word *n*+2 was matched between the correct and incorrect condition. The characteristics of these stimuli are listed in Table 4.

The invisible boundary was directly before the first character of an idiom or phrase, and either a correct or incorrect preview of word *n*+2 was presented prior to participants crossing the boundary. The incorrect previews were of the same kind as those used in Experiment 1.

INSERT TABLE 4 AND FIGURE 2 ABOUT HERE

Procedure. The procedure was identical to Experiment 1.

Results

Linear mixed-effect models constructed using the lme4 package in R were used to analyse the data. Word *n*+2 preview type and phrase type were treated as fixed factors, and were allowed to interact. Contrasts were set up using the contr.sdif function from the MASS library. These successive differences contrasts were set up for both the two-level variables such that the intercept corresponds to the grand mean and the fixed factor estimates to the main effects. The interaction between the two factors was also included in the model. Participants and items were treated as random factors in the analysis model, with random slopes for each fixed factor. We also constructed models in which the participant group was included as only a fixed factor, or was allowed to interact with other factors. Again, the contr.sdif function was used to set up the successive differences contrasts on this variable. When a more complex model significantly improved the fit of the model, this model is reported. If the more complex models did not significantly improve the fit of the data the simplest model is reported. In the case of any significant interactions between our variables we set up custom contrasts to test the underlying simple effects, and constructed a new linear mixed-effects model with these contrasts.

The same data exclusion criteria were used as in Experiment 1. 19% of the trials were lost due to participants blinking in any of the critical regions, 11% and 6% of the trials were lost because the display change triggered early or completed more than 10ms after fixation onset, respectively. In total, 35% of the data were excluded. Analyses were carried out on the same regions and for the same measures as in Experiment 1. The means and standard errors are shown in Table 5, and the model outputs in Table 6.

INSERT TABLE 5 AND 6 ABOUT HERE

Word *n*. There were no significant effects of the phrase type, the preview manipulation, or the interaction between the phrase type and preview. There was however a significant interaction between the *n*+2 preview and participant group in the skipping of the pre-target word. Simple effects contrasts revealed that this was due to a significant effect of the *n*+2 preview for participants tested in England (b = -0.42, z = -2.953, p = 0.003), but not for those tested in China (b = 0.23, z = 1.541, p = 0.120). The participants tested in England skipped the pre-target word 37% of the time given a valid *n*+2 preview, but this increased to 45% when there was an invalid *n*+2 preview. This suggests that the non-word previews attracted the point of fixation towards them, in a manner similar to low-frequency words in alphabetic languages (e.g. Hyönä & Bertram, 2004).

Word *n*+1. There was a main effect of preview type in gaze durations of 23ms, in addition to an interaction between preview type and phrase type in first fixation durations. Simple effects contrasts were performed in order to further investigate this interaction. There were significant *n*+2 preview effects in the phrases of 36ms, while no *n*+2 preview effects appeared in the idioms (b = -0.02, t = -0.61).

Furthermore, in FFD, SFD, and GD, there was a significant interaction between the word *n*+2 preview effect and participant group. The participants tested in England showed a 28ms, 32ms, and 37ms *n*+2 preview effect in FFD (b = -0.10, t = -3.11), SFD (b = -0.22, t = -2.99) and GD (b = -0.12, t = -3.36), respectively. Participants tested in China did not show any preview effects (b = -0.01, t = -0.22; b = -0.20, t = -0.45; b = -0.00, t = -0.02).

Word *n*+2. There was a significant main effect of the *n*+2 preview type in FFD (15ms), SFD (21ms) and GD (24ms). While this effect was not significant in GP, there was a trend of a 15ms preview benefit. There was a significant effect of phrase type in FFD (18ms) and GD (30ms), such that a target word was fixated for less time when it was the final word of an idiom rather than phrase. This trend was also present in SFD (12ms) and GP (20ms). No interaction between the two factors was observed. There was a main effect of participant group in gaze duration. However, this did not interact with any of our variables of interest.

Composite region. In the composite region there were significant word *n*+2 preview effects across all fixation time measures. This effect was 22ms, 53ms, 45ms, and 38ms in FFD, GD, GP, and SFD, respectively. There was a 48ms significant effect of phrase type in

GD, with shorter gaze durations for idioms than for phrases. There was no significant interaction between the two factors. [5]

Discussion

A main effect of our preview manipulation on word *n*+2 was observed in first, single, and gaze durations on word *n*+2, and in all fixation time measures on the composite region. Furthermore, there were significant effects of phrase type on first fixation and gaze durations on word *n*+2, and gaze durations across the entire unit, such that idioms were processed faster than the matched phrases. Finally, there was no interaction between our two variables on word *n*+2 or the composite region, and an interaction in the opposite direction to what was predicted during fixations on word *n*+1.

This pattern of results is somewhat surprising. We hypothesised that we would observe an interaction between our two variables, such that participants would obtain an *n*+2 preview effect only when word *n*+2 was the second constituent of an idiom. We did not observe this interaction. Had this lack of interaction occurred alongside a null effect of our preview manipulation it would not have been surprising. This would simply have suggested that the idioms were not processed as multi-word units in the parafovea, meaning that word *n*+2 was not processed to a greater extent in idioms than in other contexts where null effects have been observed. However, instead we observed a main effect of our preview manipulation, with substantial effects being observed in idioms and matched phrases. These effects were larger and more stable than those observed in prior studies (Yan et al., 2010; Yang et al., 2009, 2012) and Experiment 1 of the current paper. Clearly, something else about our stimuli drove these effects. We will return to this issue in the General Discussion.

The direction of the interaction between our two variables on word *n*+1 was also surprising. In this region we observed significant effects of the *n*+2 preview in phrases, but not idioms. It is not unusual for *n*+2 preview effects to appear on word *n*+1 (see Risse & Kliegl, 2012). However, what was puzzling was the fact that the effect was only observed within the phrases. This suggests an earlier integration of information extracted from word *n*+2 in these items. We have no explanation as to why this effect may have occurred, and will not discuss it further. However, it is worth noting that this effect was not present in our analysis of a composite region. We consider this to be a more representative measure of the processing of our items, on the basis that it consists of all instances in which the target region was fixated. In contrast, the fixation time data on word *n*+1 represents a sub-set of about 50% of trials.

While we did not observe the expected interaction, we did observe main effects of phrase type—even though the predictability of the final word of the idiom was controlled—with idioms being read more quickly than phrases, which indicated an idiom processing advantage in Chinese over phrases. These results are in line with the findings from English idioms. For example, Siyanova-Chanturia et al. (2011) found that English idioms were fixated for a shorter period of time than matched phrases. The results of the current experiment are also in agreement with findings from decision and naming tasks in three-character Chinese idioms, in which idioms were responded to faster than matched phrases (Gu & Liao, 1995; Zhang & Shi, 2009). Furthermore, our results are in line with findings by Carrol and Conklin (2015), showing that the final word of a translated four-character Chinese idiom received shorter fixations than the final word of a matched phrase. Our findings show that this advantage is also present for three-character idioms in their non-translated form.

Interestingly there was an interaction between participant group and preview type on word *n*+1. This interaction took the shape of participants in England gaining an *n*+2 preview benefit, and those in China not. This interaction was not observed during fixations on word *n*+2 or the composite target region. Thus, it seems that readers from both groups extracted information from word *n*+2, but that participants from England integrated this on intermediate fixations (i.e. on word *n*+1), whereas participants from China did not integrate it until fixating word *n*+2. This could be due to the fact that participants tested in England were primarily graduate students studying abroad with more experience of reading, whereas participants tested in China were all freshmen in the university. Veldre and Andrews (2014) investigated how reading skill was associated with readers' eye movement behaviour by using the moving window paradigm. Higher skilled readers were more disrupted when reading text presented in a small window size (i.e. 3 characters), indicating a greater sensitivity to parafoveal information for these highly skilled readers. In the current study participants tested in England may have had a greater efficiency in reading, and as such more substantial preview effects (However, see Whitford & Titone, 2015).

General Discussion

The current study aimed to investigate the parafoveal processing of idioms and matched phrases in Chinese reading, alongside replicating prior research in this area. Specifically, we were interested in the extent to which word *n*+2 was processed parafoveally when it was the second constituent of an idiom, rather than part of a phrase. We investigated this across two experiments. In Experiment 1 we tested the effect of word *n*+1 frequency on word *n*+2 preview effects, with the expectation of replicating prior research. We undertook this replication in order to ensure that the participants tested in our study were processing parafoveal information to an equivalent extent to those tested by Yang et al. (2009; 2012). This replication was successful, with us observing near identical effects to both Yang et al. studies when restricting our analysis to the same interest areas used in these studies. Thus, we can be sure that any additional effects we observed in either experiment were due to theoretically interesting reasons, as opposed to any extraneous factors to do with our participants or experimental setup. Through the analysis of a further interest area we also observed additional effects in this experiment, suggesting that the modulation of *n*+2 preview effects by word *n*+1 frequency is not as simple as previously thought. In Experiment 2 we investigated word *n*+2 preview effects in three-character Chinese idioms and matched phrases with a low-frequency word *n*+1 in the same population of participants. We hypothesised that if idioms are processed as multi-word units then word *n*+2 preview effects would be observed in these items, but not the matched phrases. We also set out to investigate how idioms are foveally processed, with prior research suggesting that they would be fixated for less time than matched phrases. We will first address our findings in relation to *n*+2 preview effects, and then discuss our results concerning idiom processing.

N+2 Preview Effects

While Experiment 1 showed a modulating effect of *n*+1 frequency on *n*+2 preview effects, this was less clear than in prior studies. When word *n*+1 was high-frequency, *n*+2 preview effects were observed across several measures and in both a composite region and on word *n*+2 itself. In contrast, when word *n*+1 was low-frequency this effect only appeared in gaze durations across the composite region. While these effects of an *n*+2 preview given a low-frequency word *n*+1 were not stable across measures, they are more than has been observed in prior studies. Furthermore, the results of Experiment 2 showed that word *n*+2 preview effects were surprisingly stable and large in both idioms and phrases compared to either type of item in Experiment 1, despite word *n*+1 always being low frequency in this experiment. Thus, it seems that Chinese readers extract information from word *n*+2 more reliably than has previously been observed.

The fact that the *n*+2 preview effects in Experiment 2 were more stable than in Experiment 1 when word *n*+1 was a low-frequency single-character word is surprising, since the two subsets of stimuli were tested during the same testing session using the same participants, and were well matched on many factors (see Tables 1 and 4), such as the frequency of word *n*+1 ($t(27) = 1.23$, $p = .229$) and the frequency of word *n*+2 ($t(56) = 0.23$, $p = .814$). The predictability of word *n*+2 was also controlled to a similar level, with the predictability being 4.0% for the phrases used in Yang et al.'s study and 0.3% in the current study.[6] It seems that factors other than word *n*+1 frequency and those we controlled for may modulate the extent to which parafoveal information is processed during Chinese reading. Potential factors that might explain the differences in *n*+2 preview benefits between the two subsets of phrases with a low-frequency word *n*+1 were carefully examined using independent t-tests. The only difference that was both significant and substantial between the phrases with a low frequency word *n*+1 in Experiment 1 and the phrases in Experiment 2 was the frequency of the characters within word *n*+2 (first character of word *n*+2, $t(46) = 2.52$, $p = .007$; second character of word *n*+2, $t(42) = 4.22$, $p < .001$). The frequency of the first and second character in word *n*+2 was 965 and 712 per million in Experiment 2, and only 295 and 130 per million in Experiment 1. Thus, it seems that the most likely factor driving our effect was the character frequency of word *n*+2.

Prior studies have shown evidence of character frequency within words influencing word recognition in Chinese. Zhang and Peng (1992) found effects in a lexical decision task, by varying character frequency while controlling word frequency. Words with high-frequency characters were identified more quickly and accurately than words with low-frequency characters. Yan, Tian, Bai, and Rayner (2006) independently varied the frequency of a target word and the characters composing it during a sentence reading task. Results showed that both word frequency and character frequency influenced the processing of the fixated word in first-pass reading times. Crucially, fixations were longer on words with low-frequency characters than words with high-frequency characters. In a similar manner, character frequency might also influence the processing of words located in the parafovea.

Ma, Li, & Rayner (2015) investigated how the properties of a parafoveal word influence the processing of the currently fixated word (a parafoveal-on-foveal effect); they manipulated both the frequency of the first character of and the whole of a two-character word in the parafovea. They found that the first character frequency—but not the whole word frequency—of the parafoveal word affected fixation times on the currently fixated word. Thus, their results suggest that the characteristics of Chinese characters might be processed in parallel in the parafovea independently of the words they form.

Our findings of word *n*+2 preview benefit in Chinese are compatible with parallel models of eye movement control due to the assumption that all words within the perceptual span are processed in parallel. Since all of the words within the foveally determined perceptual span are processed in parallel, a high-frequency character will simply be processed faster than a low-frequency character, leading to a larger preview benefit from a high-frequency parafoveal character than a low-frequency parafoveal character (Inhoff & Rayner, 1986). However, how the activation of the characters (within words) and the word's activation interact in Chinese reading, is not yet clear in this kind of model.

The *n*+2 preview benefit observed in this study may not be entirely compatible with a serial model of eye movement control, such as E-Z Reader. In this model, for word *n*+2 to be parafoveally processed, word *n*+1 must be identified whilst fixation remains on word *n*. In other words, *n+2* preview effects should only be obtained when word *n*+1 can be easily identified. In our experiment, word *n*+2 preview effects were still observed in gaze durations within the composite region despite the low-frequency of word *n*+1. However, it is not necessarily impossible that word *n*+1 was often identified from word *n* despite its low-frequency. Indeed, with a skipping rate of almost 52% it seems reasonable to assume that this word was often identified parafoveally. In cases when word *n*+1 can be identified quickly, and an attention shift occurs to word *n*+2, it is reasonable to assume that character frequency influences the processing of word *n*+2, leading to greater activation of word *n*+2 in the parafovea when its characters are high frequency. Furthermore, it is also possible that the effect occurred due to the necessity of online word segmentation in Chinese, as opposed to being the result of attention being deployed to word *n*+2.

Current models of Chinese word segmentation (e.g. Li, Rayner, & Cave, 2009) propose that characters from multiple upcoming words may be processed in parallel, with these characters feeding activation up to congruent word representations. As such, it may be that due to the ambiguity of word boundaries in Chinese, information is extracted from characters from several words in the parafovea, and that the frequency of these characters will influence the extent of this activation. Note that if this is true, our results would not necessarily violate an assumption of serial lexical processing, due to the fact that the effects were driven by the properties of sub-lexical units (i.e. the characters) rather than the lexical unit (i.e. the words). Further research is required to establish concretely that *n*+2 preview effects in Chinese are modulated by character frequency.

Before moving on, we also feel it is necessary to briefly discuss whether the lack of an enhanced *n*+2 preview effect in idioms relative to phrases is problematic for a serial processing account. Recall, our original hypothesis was that if parafoveal processing is limited by lexical as well as perceptual factors, as is the case in a serial approach, then idioms should be processed to a greater extent in the parafovea than phrases. As such, our results may seem problematic for a model such as E-Z Reader. However, we believe that our findings with regard to the processing of idioms and phrases may tell us more about the conditions required for multi-word unit processing than serial versus parallel lexical processing in general. This will become clear in the following section.

Chinese Idiom Processing

Our hypothesis that idioms may be stored and processed as larger multi-word units was only partially supported by the current results. In terms of parafoveal processing, our results showed no obvious effects of the phrase type, with our only effect being an interaction on word *n*+1 in the opposite direction to that which was hypothesised. Thus, it does not seem as though Chinese idioms, at least those with a low predictability of the final words, are immediately segmented into larger units in the parafovea and processed as a single word.[7]

Importantly, however, there were significant main effects of phrase type for foveal inspection times of idioms. First of all, the terminal word of our items received shorter fixations when they were part of an idiom as opposed to part of a phrase. This replicates an effect that has previously been observed in English idioms (see Underwood, Schmitt, & Galpin, 2004) and translated Chinese idioms (see Carrol & Conklin, 2015). Furthermore, idioms as a whole received shorter gaze durations than phrases, an effect that has also previously been observed in English (Siyanova-Chanturia, Conklin, & Schmitt, 2011). More generally, this pattern of effects supports findings from a range of experimental paradigms showing that idioms are processed more quickly than phrases. It is worth noting that we observed these effects despite controlling the relative predictability of the later parts of our idioms and phrases, which suggests idioms have a special representational status, and this is not determined by the within idiom predictability. To the best of our knowledge this is the first study that has shown the idiom processing advantage in natural Chinese reading. Furthermore, it is also the first to show such effects whilst controlling for potential predictability effects.

Our results provided little support for the lexical representation hypothesis (Swinney & Cutler, 1979), which proposes that upon encountering the initial word of an idiom, readers will activate lexical representations of the individual constituent words and whole idiom. Upon beginning to parafoveally process the first constituent of an idiom (i.e. word *n*+1), the representation of the entire idiom should have become activated, and thus word *n*+2 should have been pre-processed to a greater extent than in the matched phrases regardless of the predictability of the final parts of the idioms. This did not occur, with the pre-processing of word *n*+2 in an idiom not differing from the pre-processing of word *n*+2 in a matched phrase. The configuration hypothesis (Cacciari & Tabossi, 1988) and hybrid theories (e.g. Libben & Titone, 2008) provide better explanations for our findings. In both of these theories the processing of an idiom begins with the identification of each individual word, and the idiom is only processed as a whole once a certain level of activation has been reached. This level of activation may not have been reached prior to readers crossing the invisible boundary in our study, hence explaining why the idioms were not pre-processed in the parafovea to a greater extent than the phrases. However, sufficient activation would have been reached once the idioms were directly fixated, leading to the overall processing advantage over phrases.

It is worth considering the characteristics of our idioms which may have led to them not being processed as multi-word units in the parafovea, in contrast to the spaced compounds investigated by Cutter et al. (2014). They proposed a mechanism that would allow for the increased parafoveal processing of multi-word units in the context of an interactive activation approach to word identification. Essentially, their mechanism worked on the assumption that if a multi-word unit is stored in the lexicon, then the first constituent would activate the lexical entry for the whole multi-word unit, and that this activation would lead to the extension of processing further into the parafovea. The findings of the current study suggested that while the idiom was processed as a single lexical unit once foveal processing commenced—indicated by the standard idiom processing advantage—the activation of the whole unit by the first character was not adequate enough to lead to greater parafoveal processing. One obvious difference between the spaced compounds used in Cutter et al.'s study and the idioms used in the current study as multi-word units, which may influence the speed at which the first constituent activates a larger unit, is how likely this first constituent is to appear as part of the multi-word unit, rather than in other contexts. In Cutter et al.’s study the first constituent could be followed by very few words, with a predictability norming study showing that the second constituent was 97% predictable given the first, and a corpus search showing that the first constituent would be followed by the second constituent 42% of the time. In contrast, the second constituent of the idioms in the current study was only 1.3% predictable given the first constituent, with a search of the CCL (Center for Chinese Linguistics PKU) corpus showing that the first constituent would be followed by the second a mere 0.42% of the time. Within Cutter et al.’s processing mechanism, it can be assumed that when the first constituent of a multi-word unit is first encountered in the parafovea, the lexical entry of this word alone will be activated, as well as any multi-word units starting with this constituent. The level of activation of these various competing lexical representations will be strongly influenced by their relative frequency. In cases when the first constituent does not tend to appear outside of the context of the multi-word unit (e.g. the spaced compounds in Cutter et al.’s study), then multi-word unit activation will be high, causing the rapid expansion of processing across this whole unit. In cases when the first constituent regularly appears outside of the context of the multi-word unit (e.g. the idioms in the current study), then activation of the multi-word unit will be relatively modest compared to activation of the word alone, leading to the continued processing of this constituent as an isolated word. In this case, sufficient activation of the multi-word unit would not occur until it is directly fixated, and further excitation of this multi-word unit is caused by the latter constituents. Future studies on this topic are required to test this possibility, perhaps by manipulating the predictability of the later part of multi-word units, or the transitional probability between the first constituent and the rest of the unit.

There may, of course, be other differences between the idioms used in the current study and the spaced compounds used by Cutter et al. As discussed above, idioms have both a literal compositional meaning (i.e. *kick the bucket* referring to hitting a bucket forcefully with a foot) and a figurative idiomatic meaning (i.e. *kick the bucket* referring to dying). One could argue that the extra processing resources required to disambiguate the dual meanings of an idiom somehow led to the suppression of an *n*+2 preview effect in our idioms[8], perhaps through a constriction of the perceptual span. In contrast the spaced compounds used by Cutter at al. may have had fairly decomposable meanings, allowing for a larger perceptual span and the processing of word *n*+2. However, we do not consider this position to be supported by either our data, or the larger literature. It is worth reiterating that we observed very little difference between our idioms and phrases in terms of preview benefit, while also finding that the idioms were fixated for less time, regardless of preview type. This pattern of effects would seem to be contrary to an interpretation reliant on processing problems caused by meaning ambiguity. Presumably, if participants had struggled with resolving the conflict between the literal and figurative meaning of the idioms, then we should have seen evidence of this such that participants would remain fixated on the ambiguous idioms for longer than the unambiguous phrases. This is the opposite of what we found. Furthermore, it seems reasonable to argue that even if there was an effect of meaning ambiguity, this should not have adversely affected the parafoveal lexical activation of the idioms (or the individual words of which they were comprised), since at this point in processing there should not yet be two conflicting meanings. A further point of note is that there is not a dichotomous relationship in terms of spaced compounds always having a decomposable meaning, and idioms always having an entirely non-decomposable meaning unrelated to their constituent words. Rather, both of these classes of multi-word units vary on their level of decomposability. For example, while some items from Cutter et al.'s study were highly decomposable (e.g. *olive oil* being oil extracted from olives), it is hard to see how somebody could derive the meaning of terms such as *tikka masala*, *helter skelter*, *ferris wheel, hula hoop,* and *cocker spaniel* by simply combining the meaning of the two constituent words. Idioms also vary in terms of their decomposability. For example, while there is little relation between literally kicking a bucket and dying (i.e. a non-decomposable idiom), in the idiom *pop the question*, meaning to ask someone to marry you, there is a relatively transparent link between the idea of *the question* and a marriage proposal. Thus, it is not entirely clear that idioms and spaced compounds do fundamentally differ in terms of the level of ambiguity between their constituent-based meaning and whole phrase meaning. In the current study there was a mixture of decomposable and non-decomposable idioms; for example, the non-decomposable idiom "炒鱿鱼" literally reads as *stir-fry squid* with an idiomatic meaning of *fire someone*, while the decomposable idiom "敲警钟" literally reads as *toll the alarm bell* with an idiomatic meaning of *give a warning*. Presumably if the results of the current study were influenced by a conflict between the literal and figurative interpretation of idioms, then *n*+2 preview effects would vary between these two different types. Prior research has actually investigated the differential processing of these two types of idioms, with results being inconclusive. While a comprehensive review of this literature is beyond the scope of the current article, there is one paper worth highlighting. Libben and Titone (2008) investigated a large number of factors influencing the processing of idioms using regression analysis. Critically, in the task most resembling natural reading, no effect of decomposability was observed on reading times. In contrast, the frequency of the initial verb of the idiom did significantly influence reading times; idioms with infrequent initial verbs were read faster than idioms with highly frequent initial verbs. Furthermore, in offline predictability ratings they found that a low frequency initial verb led to a higher predictability rating for the rest of the idiom than a high frequency initial verb. Taken together, these findings suggest that the frequency of the initial verb—and the related issue of how well this predicts the rest of the idiom—is more important to the lexical retrieval of idioms than decomposability. For this reason, we consider that the way in which the initial constituents of our idioms may activate (or rather, fail to activate) the latter parts at a lexical level to be the more likely explanation of the discrepancy between the current study and the results of Cutter et al.

To close, the current study investigated factors influencing the parafoveal processing of word *n*+2 in Chinese. In Experiment 1 we observed that while *n*+2 preview effects are modulated by the frequency of word *n*+1, this effect is not as strong as has previously been proposed. Our findings also seem to suggest that *n*+2 preview effects in Chinese are modulated by the character frequency of word *n*+2. This latter finding seems quite novel, and may warrant further investigation under other controlled conditions. While idioms were not processed to a greater extent than phrases in the parafovea, there were clear differences in fixation times on the terminal word and across the whole unit. Thus, it seems that idioms are processed differently to compositional phrases, though the extent to which this occurs during parafoveal as well as foveal processing remains unclear. Further investigation of factors influencing the time course of idiom processing should shed light on this aspect of Chinese reading.

Acknowledgements

This project was initiated while the first author was visiting the University of Southampton on State Scholarship Fund 201208120023 from the China Scholarship Council. We are grateful for support from the Recruitment Program of Global Experts (1000 Talents Award from Tianjin). This project was also supported by the third author’s Grant: Ministry of Education Project of Key Research Institute of Humanities and Social Sciences in Universities 15JJD190003.We would like to thank Jinmian Yang and Keith Rayner for supplying the stimuli used in Experiment 1.

References

Angele, B., & Rayner, K. (2011). Parafoveal processing of word *n*+2 during reading: Do the preceding words matter? *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1210-1220. doi: 10.1037/a0023096

Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating *n*+1 and *n*+2 previews simultaneously. *Visual Cognition, 16*, 697-707. doi:10.1080/13506280802009704

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* New York, NY: Cambridge University Press. doi:10.1017/CBO9780511801686

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1277-1287. doi: 10.1037/0096-1523.34.5.1277

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4 classes* (R Package Version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4

Carrol, G., & Conklin, K. (2015). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea ‘has legs’. *Bilingualism: Language and Cognition.* Advance online publication. doi:10.1017/S1366728915000103

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27*, 668-683. doi:10.1016/0749-596X(88)90014-9

Chinese Character Information Dictionary (Hanzi Xinxi Zidian). (1988). Beijing, China: Science Press.

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

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

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

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2015). How is information integrated across fixations in reading? In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading.* (pp. 245-260). NY: Oxford University Press. doi:10.1093/oxfordhb/9780199324576.013.32

Gu, B., & Miao, X. (1995). A study on the comprehension of Chinese idioms [in Chinese]. *Acta Psychologica Sinica*, *27*, 15-19.

Huang, B., & Liao, X. (Eds.). (2007). *Modern Chinese Language (Updated 3rd edition) [Xiandai Hanyu (Zengding di san ban)]*. Beijing, China: Advanced Education press.

Hyönä, J, & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology, 16,* 104-127. doi: 10.1080/09541440340000132

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 20-34. doi: 10.1037/0096-1523.24.1.20

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40,* 431-439. doi: 10.3758/BF03208203

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773-795. doi: 10.1080/01621459.1995.10476572

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word *n*+2. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1250-1255. doi: 10.1037/0096-1523.33.5.1250

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 876-895. doi:10.1037/a0013484

Li, X., Liu, P., & Rayner, K. (2011). Eye movement guidance in Chinese reading: Is there a preferred viewing location? *Vision Research, 51,* 1146-1156. doi: 10.1016/j.visres.2011.03.004

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

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition, 36,* 1103-1121. doi: 10.3758/MC.36.6.1103

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

Ma, G., Li, X., & Rayner, K. (2015). Readers extract character frequency information from nonfixated-target word at long pretarget fixations during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 41,* 1409-1419. doi:10.1037/xhp0000072

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

McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society, 8,* 365-368. doi: 10.3758/BF03335168

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: The location of initial eye fixations on words. *Vision Research, 28,* 1107-1118. doi: 10.1016/0042-6989(88)90137-X

Modern Chinese Frequency Dictionary (Xiandai Hanyu Pinlv Cidian). (1986). Beijing: Beijing Language and Culture University Press.

Modern Chinese Words Dictionary (5th edition) [Xiandai Hanyu Cidian (Di wu ban)]. (2005). Beijing, China: The Commercial Press.

Morey, R. D., Rouder, J. N., & Jamil, X. (2015). *BayesFactors: Computation of Bayes factors for common designs. Version 0.9.12.* Retrieved from http://bayesfactorpcl.r-forge.r-project.org/

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 820-833. doi: t0.1037//0096-1523.26.2.820

R Development Core Team. (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.Rproject.org/

Radach, R., Inhoff, A. W., Glover, L., & Vorstius, C. (2013). Contextual constraint and *n*+2 preview effects in reading. *The Quarterly Journal of Experimental Psychology, 66,* 619-633. doi: 10.1080/17470218.2012.761256

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

Rayner, K., Juhasz, B. J., & Brown, S. J. (2007). Do readers obtain preview benefit from word *n*+2? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 230-245. doi: 10.1037/0096-1523.33.1.230

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445-476. doi: 10.1017/S0140525X03000104

Risse, S., & Kliegl, R. (2011). Adult age differences in the perceptual span during reading. *Psychology and Aging, 26,* 451-460. doi: 10.1037/a0021616

Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word *n*+2 in reading. *Journal of experimental psychology: Human perception and performance, 38,* 1026-1042. doi: 10.1037/a0027735

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 365-374. doi:10.1016/j.jmp.2012.08.001

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20,* 391-421. doi: 10.1080/13506285.2012.670143

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27,* 251-272. doi: 10.1177/0267658310382068

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior, 18,* 523-534. doi: 10.1016/S0022-5371(79)90284-6

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

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In, N. Schmitt (Eds.) *Formulaic Sequences* (pp. 155-172). Amsterdam, John Benjamins.

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology, 67,* 703-727. doi: 10.1080/17470218.2013.826258

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S.* New York: Springer.

Whitford, V., & Titone, D. (2014). Second-language experience modulates eye movements during first- and second-language sentence reading: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1118-1129. doi: 10.1037/xlm0000093

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology, 97,* 259-268. doi: 10.1348/000712605X70066

Yan, M., Kliegl, R., Shu, H., Pan, J., & Zhou, X. (2010). Parafoveal load of word *n*+1 modulates preprocessing effectiveness of word *n*+2 in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1669-1676. doi: 10.1037/a0019329

Yang, J., Rayner, K., Li., & Wang, S. (2012). Is preview benefit from word *n*+2 a common effect in reading Chinese? Evidence from eye movements. *Read Writ, 25,* 1079-1091. doi: 10.1007/s11145-010-9282-7

Yang, J., Wang, S., Xu, Y., & Rayner, K. (2009). Do chinese readers obtain preview benefit from word *n*+2? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1192-1204. doi: 10.1037/a0013554

Zang, C., Liversedge, S. P., Bai, X., & Yan, G. (2011). Eye movements during Chinese reading. In, Gilchrist, Iain and Everling, Stefan (Eds.) *The Oxford Handbook of Eye Movements* (pp. 961-978). Oxford, GB, Oxford University Press.

Zhang, B., & Peng, D. (1992). Decomposed storage in the Chinese lexicon. In H. C. Chen & O. J. L. Tzeng. (Eds.), *Language Processing in Chinese* (pp. 131-149), North-Holland.

Zhang, J., & Shi, Y. (2009). The production mechanism of Chinese idioms [in Chinese]. *Acta Psychologica Sinica*, *41,* 659-675.

Footnotes

[1] Terminology differs between the idiom and compound literature for this factor. Given the focus of this paper we adopt the terminology traditionally used for idioms (i.e. decomposable vs. non-decomposable) rather than compounds (i.e. transparent vs. opaque).

[2] We were forced to include different participant groups due to where the first author was based at the different times of testing, rather than for any theoretical reason. It is worth noting that all participants were exposed to all experimental conditions, making this a within subject design.

[3] It should be noted that effects appearing on word *n*+1 and the composite region could, of course, also be due to delayed parafoveal-on-foveal effects (Risse & Kliegl, 2012), as well as a preview benefit appearing in mislocated fixations. Regardless of the exact reason why effects may be distributed over a larger region, at a theoretical level this analysis still demonstrates an effect of upcoming parafoveal information having been processed prior to the eyes crossing the invisible boundary.

[4] The probability of the second character after the boundary forming a word with other characters, as opposed to a single character word, was 0.78. This was calculated by dividing the frequency that this character appeared as part of a multi-character word by the total frequency of the character.

[5] While no significant interactions were observed on word *n*+2 or in our composite region, there were clearly numerical differences in the size of preview benefit effects for idioms and phrases, albeit of inconsistent directions across measures. In order to affirm that we had obtained a null effect, rather than failed to obtain an effect due to a lack of statistical power, we undertook Bayes factor analysis. Bayes factors provide a value indicating the extent to which a dataset provides support for one of two competing hypotheses. Decreasing

Bayes values (below 1) provide increasing support for one hypothesis, while increasing Bayes values (above 1) provide support for the alternative hypothesis (see Kass & Raftery, 1995 and Rouder, Morey, Speckman, & Province, 2012 for a more in-depth discussion). We calculated Bayes factors using the BayesFactor package (Morey, Rouder, & Jamil, 2015) in R to ascertain whether our results for word *n*+2, and for our composite region were more adequately reflected by a model including an additive, or interactive relationship between phrase type and *n*+2 preview type. For all reading time measures on word *n*+2 Bayes factors were between 0.1 and 0.15, representing substantial support for an additive model. For first fixation, gaze duration, go past time, and single fixation duration on the composite region, Bayes factors were 0.07, 0.07, 0.12, and 0.38. Three of these values also represent at least substantial support for the additive effect over the interactive effect, while one represents anecdotal support. These values suggest that the lack of a significant interaction in our main analyses were due to the null effect being supported, rather than a lack of statistical power.

[6] Although the predictability from the text prior to word *n*+2 was slightly higher in the phrases used in Yang et al.’s (2012) study than the phrases in the current Experiment 2 ($t(19) = -1.79$, $p = 0.089$), the difference favored Yang et al.’s stimuli producing larger word *n*+2 preview effects, which was obviously not the case in the current results. Therefore, in the current study the predictability of word *n*+2 could not be the cause of larger word *n*+2 preview effect.

[7] It could be argued that a ceiling effect might exist for the word *n*+2 preview effect observed in the current study, and this ceiling effect might prevent a larger word *n*+2 preview effect being observed in idioms. Although we do not reject this possibility, we do consider it unlikely. Indeed, our initial question was whether there would be any *n*+2 preview effects at all given the frequency of word *n*+1.

[8] We thank an anonymous reviewer for this insight.

Table 1

Characteristics of Stimuli from Two Subsets of Stimuli in Experiment 1

	Type	Low-frequency word *n*+1	High-frequency word *n*+1
Word *n*+1	Number of strokes	8.80 (3.14)	6.65 (3.82)
	Word frequency	254 (306)	6070 (8149)
Word *n*+2 Correct preview	Word frequency	6.23 (5.80)	8.19 (8.31)
	Number of strokes of the first character	9.65 (2.70)	9.40 (3.20)
	Frequency of the first character	295 (357)	262 (314)
	Number of strokes of the second character	9.50 (2.46)	9.70 (2.42)
	Frequency of the second character	130 (121)	194 (267)
Word *n*+2 Incorrect preview	Number of strokes of the first character	9.65 (2.70)	9.25 (3.08)
	Frequency of the first character	294 (353)	265 (320)
	Number of strokes of the second character	9.50 (2.46)	9.65 (3.13)
	Frequency of the second character	130 (123)	165 (248)

Note. Word frequency was taken from the People's Daily Corpus (1998), and character frequency was taken from the Chinese Character Information Dictionary (1988). The frequencies are per million.

Table 2

Means (Standard Errors) for Fixation Time and Fixation Probability Measure for All Target Regions in Experiment 1

Frequency of word *n*+1	Analysis region	Preview type	Measures				
			FFD	SFD	GD	GP	SP
Low-frequency	Word *n*	Correct	290 (8)	296 (10)	296 (9)	305 (9)	0.53 (0.02)
		Incorrect	283 (6)	273 (8)	301 (9)	313 (10)	0.55 (0.02)
	Word *n*+1	Correct	319 (8)	327 (10)	329 (9)	443 (10)	0.47 (0.02)
		Incorrect	316 (7)	308 (8)	339 (9)	444 (10)	0.43 (0.02)
	Word *n*+2	Correct	296 (5)	292 (7)	375 (10)	479 (10)	0.14 (0.02)
		Incorrect	301 (5)	301 (8)	360 (9)	472 (9)	0.11 (0.01)
	Composite region	Correct	315 (6)	318 (12)	531 (14)	616 (16)	0.03 (0.01)
		Incorrect	320 (6)	341 (15)	555 (15)	632 (16)	0.03 (0.01)
High-frequency	Word *n*	Correct	270 (5)	273 (7)	276 (6)	287 (6)	0.47 (0.02)
		Incorrect	273 (5)	278 (6)	280 (6)	294 (6)	0.51 (0.02)
	Word *n*+1	Correct	257 (5)	258 (6)	261 (6)	369 (8)	0.55 (0.02)
		Incorrect	266 (5)	274 (7)	270 (6)	371 (7)	0.54 (0.02)
	Word *n*+2	Correct	268 (4)	275 (6)	319 (6)	434 (7)	0.14 (0.01)
		Incorrect	280 (4)	286 (6)	340 (7)	456 (7)	0.12 (0.01)
	Composite region	Correct	262 (4)	269 (8)	404 (9)	449 (10)	0.04 (0.01)
		Incorrect	279 (4)	300 (8)	434 (8)	474 (9)	0.03 (0.01)

Note. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; GP = go past time; SP = skipping probability

Table 3

Fixed Effects Estimates from the LME Models for All Measures Across All Target Regions in Experiment 1

Region	Effect	Measure														
		FFD			SFD			GD			GP			SP		
		Word *n*+1 High Frequency														
		b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
Word *n*	Preview	-0.02	0.03	-0.75	-0.02	0.03	-0.65	-0.02	0.03	-0.62	-0.02	0.03	-0.7	-0.19	0.12	-1.63
Word *n* + 1	Preview	-0.03	0.03	-1.34	-0.04	0.04	-0.99	-0.03	0.03	-0.95	-0.01	0.03	-0.21	0.06	0.11	0.57
Word *n* + 2	Preview	-0.04	0.02	-1.80	-0.05	0.03	-1.53	**-0.06**	**0.03**	**-2.07**	-0.04	0.02	-1.70	0.15	0.17	0.85
Composite region	Preview	**-0.06**	**0.02**	**-2.80**	**-0.12**	**0.04**	**-2.81**	**-0.09**	**0.03**	**-3.37**	**-0.09**	**0.03**	**-3.41**	0.37	0.32	1.16
		Word *n*+2 Low Frequency														
Word *n*	Preview	-0.02	0.04	-0.45	0.04	0.05	0.87	-0.03	0.04	-0.69	-0.03	0.04	-0.73	-0.09	0.13	-0.71
Word *n* + 1	Preview	-0.01	0.03	-0.28	0.02	0.04	0.46	-0.04	0.03	-1.13	-0.02	0.03	-0.45	0.20	0.13	1.47
Word *n* + 2	Preview	-0.02	0.02	-0.77	-0.02	0.04	-0.59	0.02	0.03	0.57	-0.01	0.05	-0.17	0.30	0.20	0.13
	Group	**-0.07**	**0.03**	**-2.03**	**-0.12**	**-0.12**	**-2.71**	**-0.14**	**0.04**	**-3.38**	-	-	-	-	-	-
Composite region	Preview	-0.02	0.03	-0.98	-0.09	0.07	-1.38	**-0.07**	**0.03**	**-2.19**	-0.06	0.03	-1.94	0.15	0.39	0.38
	Group	-	-	-	-	-	-	-	-	-	**-0.15**	**0.06**	**-2.55**	-	-	-

Note. Significant terms are presented in bold. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; GP = go past time; SP = skipping probability

Table 4

Mean (Standard Deviation) for Characteristics of both Idioms and Phrases

		Type	Idioms	Phrases
Word *n*+1		Number of strokes	9.32 (3.96)	9.15 (3.03)
		Word frequency	188 (230)	162 (191)
Word *n*+2		Word frequency	6.86 (10.90)	6.86 (10.90)
	Correct preview	Number of strokes of the first character	7.22 (3.90)	7.22 (3.90)
		Frequency of the first character	965 (1603)	965 (1603)
		Number of strokes of the second character	8.15 (3.68)	8.15 (3.68)
		Frequency of the second character	712 (854)	712 (854)
	Incorrect preview	Number of strokes of the first character	7.25 (3.88)	7.25 (3.88)
		Frequency of the first character	878 (1353)	878 (1353)
		Number of strokes of the second character	8.12 (3.65)	8.12 (3.65)
		Frequency of the second character	687 (830)	687 (830)
Sentence		Naturalness	4.75 (0.48)	4.56 (0.55)
		Predictability from text prior word *n*+1	0.000 (0.000)	0.000 (0.000)
		Predictability from text prior word *n*+2	0.013 (0.051)	0.003 (0.014)

Table 5

Mean (Standard Error) for Fixation Time and Fixation Probability Measure for All Target Regions in Experiment 2

Analysis region	Phrase type	Preview type	Measures				
			FFD	SFD	GD	GP	SP
Word *n*	Idiom	Correct	249 (5)	246 (5)	265 (7)	278 (8)	0.41 (0.02)
		Incorrect	255 (5)	254 (6)	281 (8)	295 (9)	0.42 (0.02)
	Phrase	Correct	252 (5)	254 (6)	269 (7)	285 (8)	0.42 (0.02)
		Incorrect	255 (5)	257 (7)	284 (9)	294 (9)	0.43 (0.02)
Word *n* + 1	Idiom	Correct	294 (7)	301 (9)	301 (7)	417 (9)	0.49 (0.02)
		Incorrect	294 (6)	295 (8)	306 (8)	408 (8)	0.52 (0.02)
	Phrase	Correct	270 (6)	272 (8)	282 (8)	394 (10)	0.54 (0.02)
		Incorrect	306 (8)	308 (11)	322 (9)	426 (10)	0.51 (0.02)
Word *n* + 2	Idiom	Correct	276 (5)	276 (7)	330 (9)	442 (9)	0.18 (0.02)
		Incorrect	289 (5)	299 (8)	358 (9)	458 (9)	0.16 (0.02)
	Phrase	Correct	292 (5)	289 (8)	364 (11)	452 (10)	0.16 (0.02)
		Incorrect	308 (6)	309 (10)	384 (11)	487 (10)	0.13 (0.02)
Composite region	Idiom	Correct	284 (5)	284 (9)	435 (11)	500 (13)	0.05 (0.01)
		Incorrect	302 (5)	333 (12)	484 (12)	528 (12)	0.04 (0.01)
	Phrase	Correct	283 (4)	294 (10)	478 (15)	527 (16)	0.04 (0.01)
		Incorrect	310 (5)	318 (15)	537 (15)	590 (16)	0.03 (0.01)

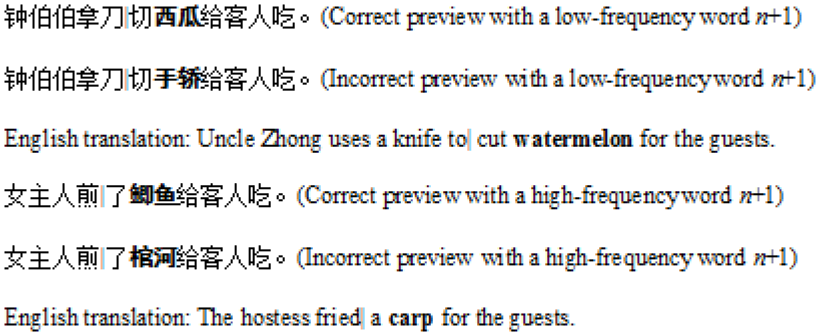
Note. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; GP = go past time; SP = skipping probability

Table 6

Fixed Effects Estimates from the LME Models for All Measures Across All Target Regions of Experiment 2

Region	Effect	Measure														
		FFD			SFD			GD			GP			SP		
		b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
Word *n*	Phrase	-0.01	0.02	-0.35	-0.00	0.02	-0.11	-0.00	0.02	0.02	0.00	0.02	0.08	0.03	0.10	0.31
	Preview	-0.01	0.02	-0.40	-0.02	0.02	-0.71	-0.03	0.02	-1.39	-0.02	0.02	-1.03	-0.10	0.10	-0.94
	Phrase*Preview	0.01	0.03	0.46	0.05	0.04	1.19	0.02	0.04	0.56	0.04	0.04	1.05	-0.01	0.20	-0.06
	Group	-	-	-	-	-	-	-	-	-	-	-	-	-0.04	0.21	-0.17
	Preview*Group	-	-	-	-	-	-	-	-	-	-	-	-	**-0.65**	**0.21**	**-3.12**
Word *n* + 1	Phrase	-0.03	0.02	-1.27	-0.04	0.03	-1.34	-0.02	0.03	-0.89	-0.03	0.03	-1.18	0.06	0.09	0.69
	Preview	**0.06**	**0.02**	**-2.55**	-0.05	0.03	-1.55	**-0.06**	**0.02**	**-2.50**	-0.04	0.02	-1.57	-0.01	0.09	-0.08
	Phrase*Preview	**-0.09**	**0.04**	**-2.06**	-0.08	0.06	-1.39	-0.09	0.05	-1.90	-0.09	0.05	-1.92	0.27	0.19	1.44
	Group	-0.04	0.04	-1.20	-0.01	0.04	-0.34	-0.05	0.04	-1.23	-	-	-	-	-	-
	Preview*Group	**-0.09**	**0.04**	**-2.06**	**-0.16**	**0.06**	**-2.75**	**-0.11**	**0.05**	**-2.33**	-	-	-	-	-	-
Word *n* + 2	Phrase	**-0.05**	**0.02**	**2.8**	0.02	0.03	0.81	**0.05**	**0.02**	**2.21**	0.02	0.03	0.70	**-0.25**	**0.13**	**-1.97**
	Preview	**-0.04**	**0.02**	**-2.3**	**-0.06**	**0.03**	**-2.45**	**-0.07**	**0.02**	**-2.88**	-0.05	0.03	-1.90	0.14	0.13	1.12
	Phrase*Preview	0.01	0.04	0.2	0.02	0.05	0.40	0.03	0.04	0.80	-0.05	0.05	-1.10	0.20	0.25	0.77
	Group	-	-	-	-	-	-	0.10	0.04	-2.53	-	-	-	-	-	-
Composite region	Phrase	0.01	0.02	0.50	-0.00	0.04	-0.06	**0.06**	**0.02**	**2.37**	0.04	0.03	1.56	-0.25	0.24	-1.03
	Preview	**-0.07**	**0.02**	**-4.30**	**-0.12**	**0.04**	**-2.76**	**-0.12**	**0.03**	**-4.69**	**-0.11**	**0.02**	**-4.54**	0.15	0.24	0.62
	Phrase*Preview	-0.01	0.03	-0.40	0.08	0.07	1.19	-0.01	0.04	-0.04	-0.05	0.04	-1.18	0.33	0.48	0.68

Note. Significant terms are presented in bold. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; GP = go past time; SP = skipping probability



钟伯伯拿刀|切**西瓜**给客人吃。(Correct preview with a low-frequency word *n*+1)

钟伯伯拿刀|切**手轿**给客人吃。(Incorrect preview with a low-frequency word *n*+1)

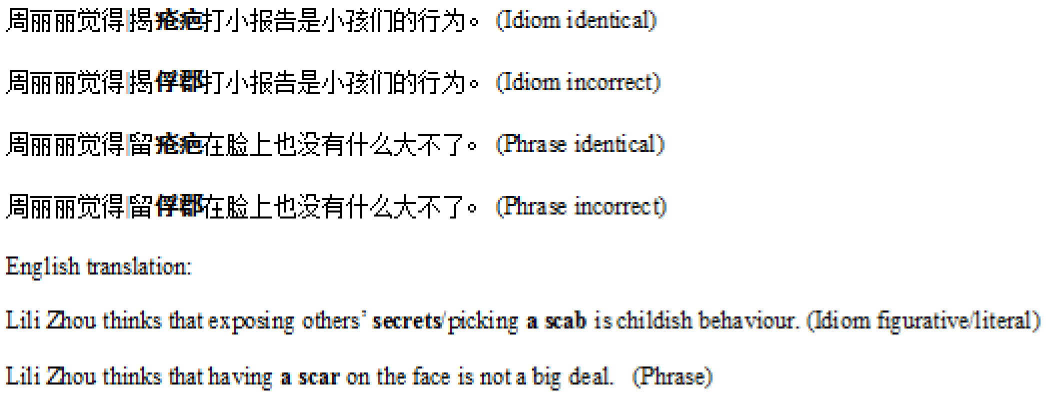
English translation: Uncle Zhong uses a knife to| cut **watermelon** for the guests.

女主人煎|了**鲫鱼**给客人吃。(Correct preview with a high-frequency word *n*+1)

女主人煎|了**棺河**给客人吃。(Incorrect preview with a high-frequency word *n*+1)

English translation: The hostess fried| a **carp** for the guests.

Figure 1. An Example of Stimuli in Experiment 1. The two characters in bold comprise word *n*+2. They were not in bold and in same font as the other characters in the experiments. The character after the boundary "|" and prior to word *n*+1 is word *n*+1.



周丽丽觉得|揭**疮疤**打小报告是小孩们的行为。 (Idiom identical)

周丽丽觉得|揭**俘郡**打小报告是小孩们的行为。 (Idiom incorrect)

周丽丽觉得|留**疮疤**在脸上也没有什么大不了。 (Phrase identical)

周丽丽觉得|留**俘郡**在脸上也没有什么大不了。 (Phrase incorrect)

English translation:

Lili Zhou thinks that exposing others' **secrets**/picking **a scab** is childish behaviour. (Idiom figurative/literal)

Lili Zhou thinks that having **a scar** on the face is not a big deal. (Phrase)

Figure 2. An Example of Stimuli in Experiment 2. The two characters in bold comprise word *n*+2. They were not in bold and in the same font as the other characters in the experiment. The character after the boundary "|" and prior to word *n*+2 is word *n*+1. Sentence frames were the same for idioms and phrases prior to the boundary.