Journal of Memory and Language 134 (2024) 104474

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Reading compound words in Finnish and Chinese: An eye-tracking study

Jukka Hyönä [a,*,1], Lei Cui [b,*,2], Timo T. Heikkilä [a], Birgitta Paranko [a], Yun Gao [b], Xingzhi Su [b]

[a] *University of Turku, Finland*
[b] *Shandong Normal University, China*

ARTICLE INFO

Keywords:
Word recognition
Compound words
Visual acuity
Reading
Finnish
Chinese

ABSTRACT

Two eye-tracking experiments in alphabetic Finnish and two in logographic Chinese examined the recognition of two-constituent compound words in reading. In Finnish, two-constituent compound words vary greatly in length, whereas in Chinese they are identical in length. According to the visual acuity principle (Bertram & Hyönä, 2003), short Finnish compound words and all two-character Chinese compound words that fit in foveal vision are recognized holistically, whereas long Finnish compound words are recognized via components. Experiment 1 in Finnish provided evidence consistent with the account, whereas the results for long compound words presented in condensed font in Experiment 2 were inconsistent with it. In Chinese, the first-character frequency effect was non-significant even when the compound words were presented in large font. The Finnish results suggest that componential processing is necessary when the compound word entails more than 10 letters. The Chinese results are compatible with the Chinese Reading Model (Li & Pollatsek, 2020) that assumes whole-word representations to overrule the activation of components during compound word recognition.

Introduction

In many languages, word compounding is very productive. Compound words can be formed by gluing together existing words (*birth* + *day* -> *birthday*). Productivity in word compounding may manifest in the majority of words in a dictionary being compound words. This is the case in two typologically and orthographically distinct languages, Finnish and Chinese. Written Finnish is an alphabetic script with one-to-one mapping between graphemes and phonemes. Written Chinese, on the other hand, is a logographic script, where characters and character combinations are mapped onto meanings without a direct recourse to spoken language. Despite significant differences in these two scripts, word compounding is very productive in both languages. In Finnish, more than 60 % of all words are compound words; in Chinese, the figure is even higher (more than 70 %). Thus, in order to understand written word recognition in these two languages (for universal and script-specific reading mechanisms, see Li, Huang, Yao, & Hyönä, 2022; Liversedge, Drieghe, Li, Yan, Bai, & Hyönä, 2016; Liversedge, Olkoniemi, Zang, Li, Bai, Yan, & Hyönä, 2023; Siegelman, Schroeder, & Acartürk et al., 2022), it is of importance to study how compound words are recognized during reading.

In the present study, we examined how two-component compound words (the most typical compound word type in both languages) are recognized during reading of Finnish and Chinese. Prior research suggests that the recognition process may differ between the two languages. In Finnish, morphological decomposition is likely to play a relevant role (for a review of reading compound words in Finnish and other alphabetic languages, see Hyönä, 2015), whereas for Chinese the available evidence favoring decompositional versus holistic processing is mixed (see below for further details).

Compound word recognition in Finnish

Pollatsek, Hyönä and Bertram (2000) put forth a model to describe compound word recognition in Finnish. They proposed two routes to lexically accessing compound words, a decomposition route and a holistic route. The model bears close resemblance to models proposed to describe recognition of all morphologically complex words (Schreuder & Baayen, 1995). The two mechanisms provide alternative routes to word identity. The decomposition route attempts access via the free morphemes comprising the word meaning, whereas the holistic route achieves it by activating the whole-word representation in the reader's

---

* Corresponding authors.
E-mail addresses: hyona@utu.fi (J. Hyönä), cuilei_cn@163.com (L. Cui).
[1] Department of Psychology and Speech-Language Pathology, University of Turku, FI-20014 Turku, Finland.
[2] Department of Psychology, Shandong Normal University, 250014 Jinan, China.

https://doi.org/10.1016/j.jml.2023.104474
Received 1 March 2023; Received in revised form 17 October 2023; Accepted 17 October 2023

mental lexicon. The two routes work in parallel; whichever route completes first, wins the race.

Subsequently, Bertram and Hyönä (2003, see also Bertram & Hyönä, 2013) demonstrated that in Finnish compound word recognition length is a significant determiner in the race. They found that the frequency of the first constituent as a separate word played a significant role in recognizing long but not short compound words, whereas an effect of whole-word frequency emerged early during the recognition process for short compound words but only later for long compound words. Recently, Hyönä, Pollatsek, Koski and Olkoniemi (2020) observed an analogous effect for second constituent frequency: Its effect was robust when reading long compound words but negligible when reading short compound words.

To account for the pattern of results, Bertram and Hyönä (2003) put forth the visual acuity principle. According to this notion, if a printed compound word is short enough that it fits in the fovea where visual acuity is best, the holistic route is the winner. This is because readers are able to perceive all the letters simultaneously, which makes holistic access a viable option. On the other hand, when the compound word is so long that some of the letters fall outside the foveal area, the decomposition route gets a head start. In that case, the reader initially attempts to access the first component followed by the access of the second component and that of the whole word (see Pollatsek et al., 2000).

Juhasz (2008) could not replicate the results of Bertram and Hyönä (2003) for English. Similar to Bertram and Hyönä (2003), Juhasz manipulated first-constituent frequency for both short and long two-constituent compound words. However, a reliable first-constituent frequency effect was observed for short compounds and a tendency for an inverted first-constituent frequency effect for long compounds. Similarly, Kuperman, Schroeder, Bertram and Baayen (2009) did not observe in Dutch a modulation of constituent frequency effect by the length of compound words ranging from 8 to 12 letters (although most of their compound words were 12 letters long). Thus, the first aim of the present study was to test the credibility of the results obtained by Bertram and Hyönä (2003) by replicating the results of their Experiment 1 using a new set of materials and participants.

Compound word recognition in Chinese

The results on reading compound words in Chinese are somewhat mixed. On one hand, there are studies observing a significant character frequency effect with frequent characters leading to faster recognition times (Chen, Song, Lau, Wong, & Tang, 2003; Tse & Yap, 2018; Yan, Tian, Bai, & Rayner, 2006). On the other hand, other studies have failed to do so (Cui, Yan, Bai, Hyönä, Wang, & Liversedge, 2013; Cui, Häikiö, Zhang, Zheng, & Hyönä, 2017; Li, Bicknell, Liu, Wei, & Rayner, 2014; Ma, Li, & Rayner, 2015). Yet, there is evidence for an inverted character frequency effect with frequent characters associated with longer recognition times (Cui, Wang, Zhang, Cong, Zhang, & Hyönä, 2021; Tsang, Huang, Lui, Xue, Chan, Wang, & Chen, 2018; Yu, Liu, & Reichle, 2021). In sum, the available evidence on compound word recognition in Chinese does not consistently favor either holistic or componential processing.

There is some evidence suggesting that compound word frequency may mediate processing of Chinese compound words. Yan et al. (2006) observed reliable first- and second-character frequency effects for infrequent but not for frequent compound words. Moreover, Cui et al. (2021) found an inverted component frequency effect for infrequent compound words: when the first component was frequent, fixation time on the second component was longer than when the first character was infrequent. Cui et al. suggested that the inverted character frequency effect may reflect an effect of morphological family size. Frequent characters combine with many other characters in making up compound words, whereas infrequent characters do so with fewer characters. Thus, the identity of the second character is less constrained preceded by a frequent than an infrequent first character, leading to longer fixation times on the second character. Yu et al. (2021) conducted a sub-analysis of their stimuli for which family size was equated and still found an inverted character frequency effect. Consequently, they propose an alternative hypothesis, according to which the effect may reflect word segmentation. Infrequent characters may stand out from other characters, which in turn may be initially regarded as single-character words subsequently resulting in short fixations on them.

Theoretical models offer alternative viewpoints for compound word recognition in Chinese. The interactive-activation model of Taft and Zhu (1997) postulates four levels of nodes, one for strokes, radicals, characters and words. Importantly in the present context, the model assumes excitatory connections from the character level to the word level. In other words, as frequent characters are activated faster than infrequent characters, compound words comprising frequent characters will also be recognized faster that those comprising infrequent characters.

The Chinese Reading Model (CRM) of Li and Pollatsek (2020) was developed to simulate word recognition and readers' eye movement patterns during continuous reading. One of the model's key aims is to address the question of how readers recognize words when dealing with the Chinese orthography, where word boundaries are not marked by spaces. This feature brings with it the problem of word segmentation. Individual characters may combine in multiple ways with the neighboring characters to form words. According to CRM, all possible word candidates within the reader's perceptual span will be activated. Thus, both the compound word and the embedded words within the compound word (e.g., single-character words) receive activation. However, the compound word will become more activated than single-character words, as it receives activation from the component characters. Thus, the compound word is likely to "win the race"; that is, the compound word rather than the embedded single-character word will be segmented out and recognized. In other words, CRM predicts that two-character Chinese compound words are recognized as holistic units. The model simulations provided strong evidence for this view. However, as embedded words such as single-character words also become activated during the recognition process, the model predicts that infrequent single-character words embedded in a compound word may inhibit the process.

Finally, the Multi-Constituent Unit (MCU) hypothesis (Zang, 2019) is also relevant here. It states that "frequently used linguistic units comprised of more than a single word may be represented, and therefore identified, lexically as single representations" (Zang, Fu, Du, Bai, Yan, & Liversedge, 2023). Zang and colleagues (Zang, Fu, Bai, Yan, & Liversedge, 2021; Zang et al., 2023) have provided strong support for the hypothesis by demonstrating that multiple-constituent units in Chinese (e.g., idioms) are identified as single lexical representations (for English results supporting the hypothesis, see Cutter, Drieghe, & Liversedge, 2014). Thus, the hypothesis posits that also compound words comprised of two single-character words should be identified as single units.

In sum, the model of Taft and Zhu (1997) predicts frequent characters of two-character Chinese compound words to speed up the compound word recognition. On the other hand, the CRM model of Li and Pollatsek (2020) and the MCU hypothesis (Zang, 2019) assumes whole-word representations to prevail over single-character representations. Yet, according to CRM, infrequent single-character words embedded in two-character compound words may delay the process.

The present study

In the present study, we examined reading of two-constituent compound words in two orthographically and typologically distinct languages, Finnish and Chinese. We varied the frequency of the first constituent as a continuous variable by selecting words from different frequency bands (see the Methods, for details). The manipulation of constituent frequency is a litmus test of componential processing (Taft & Forster, 1976). Its effect demonstrates that compound word components become active and are used during compound word recognition. In

Finnish, we manipulated first constituent frequency separately for short and long compound words. To increase compatibility between the Finnish and Chinese target words, we looked for two-constituent compound words with a similar meaning and similar first-constituent frequency in the two languages. Moreover, we kept the whole-word frequency low and ensured that it did not co-vary with first-constituent frequency. The frequency of the second constituent was also matched across the frequency bands of the first constituent as well as across languages.

Four eye-tracking experiments were conducted on reading two-constituent compound words in Finnish and Chinese. In Experiments 1 and 2, short and long Finnish compound words were embedded in sentences and native speakers of Finnish read them for comprehension. Experiment 1 was a modified replication of Experiment 1 of Bertram and Hyönä (2003). Experiment 2 tested further the visual acuity principle by equating the horizontal extent of the short and long compound words by presenting the long compound words in a condensed font. According to the visual acuity principle, when all the letters of long compound words are within foveal reach when fixated, holistic processing will ensue, which should wipe out the first-constituent frequency effect.

Experiments 3 and 4 were conducted in Chinese using materials comparable to the Finnish materials. In Experiment 3, font size was used to fit the compound words in the fovea once fixated. In Experiment 4, for half of the compound words (matched with the long Finnish compound words) the font size was increased so they were similar in horizontal extent to the Finnish long compound words of Experiment 1. In other words, the entire compound word no longer fitted in the fovea. According to the visual acuity principle, a character frequency effect should be observed for this long compound word set.

Experiment 1

In Experiment 1, we aimed to replicate the results of Experiment 1 of Bertram and Hyönä (2003). Thus, we predicted that first-constituent frequency should produce a reliable effect in fixation time on long compound words, but no or little effect in fixation time on short compound words. Experiment 1 differed from that of Bertram and Hyönä in that we manipulated first-constituent as a continuous variable by making use of six frequency quantiles, instead of a dichotomous variable (high vs. low). Moreover, we increased the sample size from 30 to 50 participants.

In the experiment, adult participants read single sentences silently for comprehension while their eye movements were tracked. Each target sentence contained a two-constituent compound word appearing in the middle of the sentence. The frequency of the first constituent as a separate word was manipulated analogously for a set of short and long compound words, while matching for whole-word and second-constituent frequency.

Data availability

The data and the data analyses with analysis code are available for all four experiments in the Open Science Framework (osf.io/9tc3j).

Methods

Participants. Fifty university students (44 female), all native speakers of Finnish, took part in the experiment. They received course credit for participation. All but one reported no reading difficulties. None of them had any neurological disorders. They gave informed consent for participation when signing up for the experiment in the internet. The entire study was approved by the Ethics committee of the University of Turku (#29/2019).

Apparatus. Readers' eye movements were recorded with a desk-mounted Eyelink 1000 (SR Research, Ontario, Canada) eye-tracker using 1000 Hz sampling rate. Movements of the right eye were tracked. The tracker was calibrated using a 9-point calibration grid. The accuracy was better than 0.5 degrees of visual angle. The stimuli were presented on a BenQ XL2411Z monitor (1920x1080 resolution; frame rate 144 Hz). Participants were seated 69 cm from the monitor with their head positioned on a chin rest.

Materials. Fifty-two short (6–9 letters) and 54 long (11–15 letters) Finnish compound words were used as the targets. The average length of the short compounds was 7.94 letters (SD =.54) and that of the long compounds 12.30 letters (SD =.98). The length of the first constituent of short compounds was 3.9 letters (SD =.65) and that of the second constituent 4.0 letters (SD =.71). The corresponding lengths for the long compounds were 6.2 (SD = 1.42) and 6.1 (SD = 1.29) letters, respectively. The average horizontal extent of the short compound words was 2.6 degrees of visual angle and that of the long compound words 4.0 degrees of visual angle.

The frequency of the first constituent as a separate word was similarly manipulated for short and long compound words. We identified six quantiles in the frequency continuum; for each frequency quantile we searched for 10 existing short and long compound words whose first-constituent frequency was similar between short and long compounds. In other words, we manipulated first-constituent frequency as a continuous rather than a dichotomous variable. Logarithmic frequencies were used. Word frequencies were extracted from Parsebank (Luotolahti, Kanerva, Laippala, Pyysalo, & Ginter, 2015), which is a mass-scale Finnish Internet corpus comprising 3.4 billion word tokens. After establishing a list of 120 compound words matched for first-constituent frequency, we eliminated a few words whose whole-word frequency was either very low or high. The final list comprised 54 long and 52 short compound words.

The log frequency of the first constituent varied significantly between the quantiles, $F(5,94) = 132.32$, $p <.001$. However, it did not differ between the short and long compounds, $F < 1$. Whole-word frequency was generally low and it did not vary between the short and long words, $F(1,94) = 1.26$, $p =.27$, or between the quantiles of first-constituent frequency, $F < 1$. Moreover, second-constituent frequency did not systematically vary between the short and long words, $F(1,94) = 1.57$, $p =.21$, or between the quantiles of first-constituent frequency, $F(5,94) = 1.67$, $p =.15$. The frequency values are presented in Table 1.

Within each frequency quantile, a short compound was paired with a long compound and a sentence frame was written for the pair that was identical up to the target word. After excluding a few items (see above), there were two long compound words that were not matched with a short compound word. For most pairs, the word following the target was also identical; if not, it was similar in length and typically started with the same initial letters. The sentence frame leading to the target word was semantically neutral ("I thought that", "According to the forecast", "In the brochure, it was praised that", etc.). After the target word, the sentence continued in a way that was semantically plausible given the sentence beginning. Internet was used to find appropriate scenarios for the target words. Below, an example sentence pair is presented with English translations (the target words, *huvivene* = leisure boat and *synnytysosasto* = maternity ward, are underlined for illustrative purposes).

Short compound: *Oletettavasti upouusi huvivene oli eksynyt reitiltään kartanlukuvirheen takia.*

Long compound: *Oletettavasti upouusi synnytysosasto oli määrä avata jo ennen vuodenvaihdetta.*

Presumably, the brand-new leisure boat had got lost from its course due to a navigation error.

Presumably, the brand-new maternity ward was supposed to open already before the end of the year.

By using neutral contexts, the predictability of the target words was practically zero. This was confirmed by a corpus analysis using Parsebank (Luotolahti et al., 2015) to carry out concordance searches in the 3.4 billion word corpus. Given the sentence beginning, the target word succeeded in the corpus only for 4 out of the 106 targets. Even in those instances, the target word only appeared on average in 0.6 % of the sentence frames (the maximum was 2.3 %). When computing the concordance ratings, the sentence frame could appear in the corpus

Table 1
Logarithmic frequencies (per million values in parentheses) of 1st constituent, 2nd constituent and whole word for short and long Finnish compound words for the six frequency quantiles.

	Quantile 1	Quantile 2	Quantile 3	Quantile 4	Quantile 5	Quantile 6	Total
		1st constituent frequency					
Short	.37 (2.86)	1.11 (13.5)	1.74 (56.4)	2.16 (149)	2.21 (251)	2.74 (870)	1.61 (188)
Long	.34 (2.56)	1.13 (15.5)	1.68 (55.7)	2.00 (105)	2.19 (198)	2.61 (440)	1.72 (135)
		2nd constituent frequency					
Short	1.79 (378)	1.88 (199)	1.38 (86.5)	1.89 (137)	1.60 (63.8)	2.03 (196)	1.75 (184)
Long	1.41 (30.4)	1.72 (102)	1.27 (51.5)	1.82 (353)	1.91 (140)	1.45 (88.2)	1.62 (136)
		Whole-word frequency					
Short	-.11 (1.12)	-.17 (1.82)	-.18 (1.19)	-.11 (0.91)	-.004 (1.39)	.06 (2.03)	-.10 (1.40)
Long	-.67 (0.41)	-.14 (3.73)	-.30 (2.08)	-.16 (1.38)	-.005 (4.05)	-.05 (1.01)	-.19 (2.28)

either in the beginning or middle of the sentence.

Procedure. The participants were asked to read the sentences silently for comprehension at their normal reading speed. After 25 % of the sentences, their comprehension was assessed by a true–false statement. They were instructed to press the blue key when the statement was congruent with the meaning of the last read sentence and to press the red key when it was incongruent. Participants responded correctly to the majority of the statements (the response accuracy was 95 %). The target sentences were preceded by 8 practice sentences, 3 of which were followed by a true–false statement. The target sentences were presented in two blocks; a short pause followed the first block. The order of blocks was counterbalanced across the participants. The testing session lasted an average of about 45 min.

Results

Four fixation time measures were used to index compound word recognition. Single fixation duration reflects the time spent on the target word, when only one fixation was made. First fixation duration is the duration of the initial fixation on the compound word when more than one fixation was made during the first-pass reading. It is worth noting that long compound words were read with a single fixation only in 25 % of the time, whereas for short compound words the single-fixation strategy was more prevalent (59 %). Gaze duration reflects the time spent reading the word before moving on (gaze duration is the summed duration of all fixations made on the word before exiting it to either right or left). First fixation and single fixation duration reflect immediate effects of target word reading, while gaze duration also includes the durations of possible refixations made on the word. Gaze duration is our primary measure of compound word recognition, as it is likely that when the reader decides to move forward in the text, (s)he has identified the compound word.

The fixation duration measures were log-transformed before analysis. The statistical analyses were conducted with linear mixed effects models using the lme4 package (version 1.1–30; Bates, Kliegl, Vasishth, & Baayen, 2015) in R (version 4.1.1; R Core Team, 2021). Before the analyses, the data were trimmed as follows. For the analysis of single and first fixation duration, all fixations shorter than 80 ms and longer than 500 ms were discarded. For the analyses of gaze duration, gaze durations shorter than 80 ms and longer than 1400 ms were excluded. In the analysis of first fixation duration, 3.5 % of the trials were discarded; in the analysis of gaze duration, 1.5 % of the trials were excluded.

In order to further examine the timing of the effect obtained in gaze duration, we also analyzed two fixation probability measures, probability of refixation and probability of making more than two fixations on the target word before fixating away from it. These data were analyzed with generalized mixed linear models using the same lme4 package as with LMMs.

The observed means and their standard deviations for the four dependent measures are presented in Table 2, broken down by word length and first-constituent frequency quantile.

Table 2
Means and standard deviations (in parentheses) in Experiment 1 by participants, computed on all the observations.

		First-constituent frequency quantile					
Measure	Word length	1	2	3	4	5	6
SFD	Short	255 (41)	244 (50)	257 (51)	257 (56)	256 (50)	257 (64)
	Long	223 (63)	232 (56)	257 (80)	248 (51)	249 (62)	246 (65)
FFD	Short	220 (56)	223 (51)	220 (65)	228 (56)	215 (50)	215 (55)
	Long	220 (44)	214 (30)	218 (38)	221 (28)	210 (36)	209 (36)
GD	Short	357 (92)	338 (88)	357 (93)	344 (106)	337 (92)	343 (96)
	Long	506 (136)	494 (148)	476 (141)	414 (102)	397 (120)	443 (140)
Refix. prob.	Short	.41 (.21)	.41 (.22)	.43 (.23)	.35 (.23)	.41 (.27)	.43 (.25)
	Long	.82 (.21)	.77 (.19)	.78 (.23)	.73 (.22)	.66 (.23)	.75 (.23)
> 2 fix.	Short	.09 (.12)	.07 (.10)	.09 (.10)	.07 (.12)	.06 (.10)	.05 (.08)
	Long	.41 (.27)	.35 (.26)	.29 (.22)	.20 (.21)	.18 (.19)	.25 (.24)

Note: SFD = single fixation duration; FFD = duration of first of multiple fixations; GD = gaze duration; Refix. prob. = probability of refixation; > 2 fix. = probability of more than two fixations.

The LMM and GLM models were selected by comparing different random structures with an analysis of variance (ANOVA) and by selecting a parsimonious LMM which was not overparameterized (Bates, Mächler, Bolker, & Walker, 2015). The initial random structures included intercepts for participants and items, all the variables that were included as one-way fixed effects in the model as a random slope for the participant, and a correlation parameter. These models were compared with models in which some or all random effects had been removed, including the correlation parameter. If ANOVA indicated a significant difference between the models, we chose the model which was preferred by at least 2 out of 3 goodness of fit measures (AIC, BIC, Log likelihood). If there was no statistically significant difference between the models, the model with fewer parameters was chosen. The fixed effect estimates were only checked after the model selection; hence, they did not affect the model selection. All continuous fixed variables were centered to the mean.

The results of the linear mixed models for single fixation duration, first fixation duration and gaze duration and those of the generalized linear mixed models (probability measures) are presented in Table 3.

Single and first fixation duration

In the models for single and first fixation duration, no effect approached significance (see Table 3).

Table 3
Linear mixed effect models for single fixation duration, first fixation duration and gaze duration and generalized mixed linear models (probability of refixation and probability of making more than two fixations) in Experiment 1 with word length and first-constituent frequency (both continuous variables) entered as the fixed effects. Statistically significant effects appear in bold.

Measure	Predictor	b	SE	t/z
SFD	Word length (letters)	−0.008	0.005	−1.651
	First-constituent frequency (log)	0.007	0.012	0.615
	Word length × First constituent frequency	0.002	0.005	0.472
FFD	Word length (letters)	−0.001	0.004	−0.134
	First-constituent frequency (log)	−0.007	0.009	−0.801
	Word length × First constituent frequency	−0.001	0.004	−0.330
GD	**Word length (letters)**	**0.061**	**0.006**	**10.768**
	First constituent frequency (log)	**−0.042**	**0.016**	**−2.735**
	Word length × First constituent frequency	**−0.016**	**0.006**	**−2.478**
Refix. prob.	**Word length (letters)**	**0.420**	**0.030**	**13.994**
	First-constituent frequency (log)	−0.120	0.066	−1.799
	Word length × First constituent frequency	**−0.072**	**0.029**	**−2.521**
> 2 fix.	**Word length (letters)**	**0.428**	**0.033**	**13.166**
	First constituent frequency (log)	**−0.342**	**0.096**	**−3.554**
	Word length × First constituent frequency	**−0.078**	**0.039**	**−1.985**

Note: SFD = single fixation duration; FFD = duration of first of multiple fixations; GD = gaze duration; Refix. prob. = probability of refixation; > 2 fix. = probability of making more than two fixations.

Gaze duration

As is evident from Table 3, both main effects and their interaction proved significant. The pattern of results is depicted in Fig. 1. Long and low-frequency first-constituent compound words were read with longer gaze durations than short and high-frequency first-constituent compound words. Most importantly, however, as predicted by the visual acuity principle, the effect of first-constituent frequency was the greater, the longer the compound word was. Fig. 1A plots the model estimates when word length and first-constituent frequency were both considered continuous variables, while Fig. 1B plots the estimates when word length was considered a dichotomous variable. The LME model associated with Fig. 1B is presented in the supplementar y materials in OSF. Also in that model, all three effects proved significant.

In order to analyze the relative timing of the effect in gaze duration, we examined the probability of making at least one refixation on the target word as well as the probability of making more than two fixations on the target.

Probability of refixation

Unsurprisingly, the longer the word, the more probable it was to make a refixation on it. More interestingly, word length interacted with first constituent frequency (see Table 3). As is apparent from Fig. 2A, for longer words the probability of refixation decreased as first constituent frequency increased, which was not the case for shorter compound words. The pattern parallels that obtained for gaze duration.

Probability of making more than two fixations

Readers were more likely to make more than two fixations on long than short words as well as on words with an infrequent than frequent first constituent. Moreover, the reliable interaction between the fixed factors (see Table 3) suggests that the first-constituent frequency effect was only apparent for the long compound words (see Fig. 2B).

Discussion

In Experiment 1, participants read long and short Finnish compound words in a sentence context while their eye movements were tracked. The frequency of the first constituent was manipulated in a continuous fashion and similar to long and short compound words. Whole word frequency was matched (it was generally low) between long and short compound words and across the six quantiles of first-constituent frequency.

The results for gaze duration were perfectly in line with the visual acuity principle (Bertram & Hyönä, 2003). According to this notion, the recognition of long compound words is initiated via the access of the first compound word component, because immediate access to the whole word is not viable due to foveal constraints (the entire word does not fit in the fovea where visual acuity is at its best). On the other hand, short compound words that fit in the fovea are assumed to be recognized holistically with compound word constituents playing no or negligible role. In Experiment 1, the short compound words fitted mostly in the fovea (an area of 2 degrees of visual angle around the fixation point), as they covered on average an area of 2.6 degrees of visual angle. Thus, as predicted, the frequency of first constituent affected gaze durations on long compound words but did not do so for short compound words. Moreover, compound word length modulated the effect of first-constituent frequency in a continuous fashion with the effect linearly increasing as a function of word length.

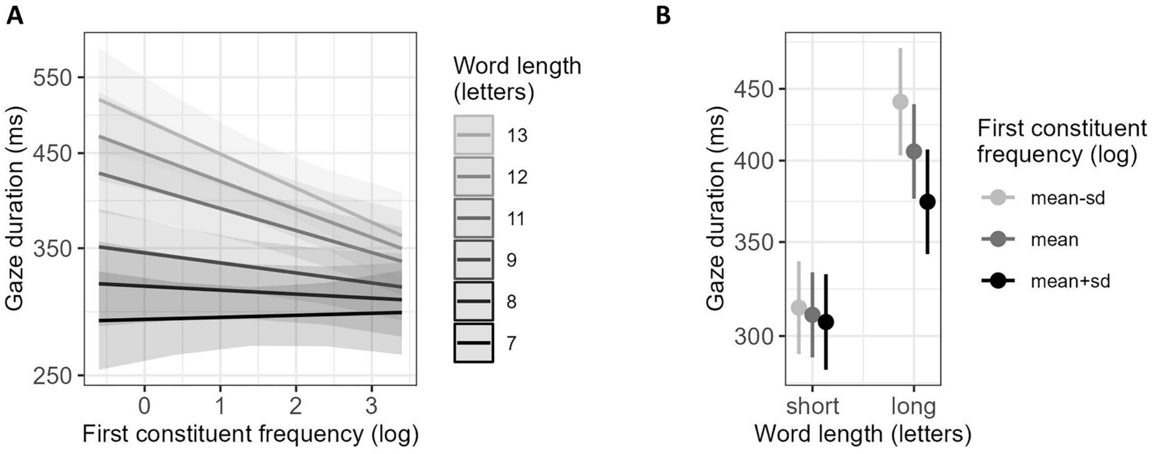

Fig. 1. A: The interaction between word length and first-constituent frequency in gaze duration in Experiment 1 when both independent variables were considered continuous variables. The shaded areas are 95% CIs. B: The interaction between word length and first-constituent frequency in gaze duration when word length was considered a dichotomous variable.

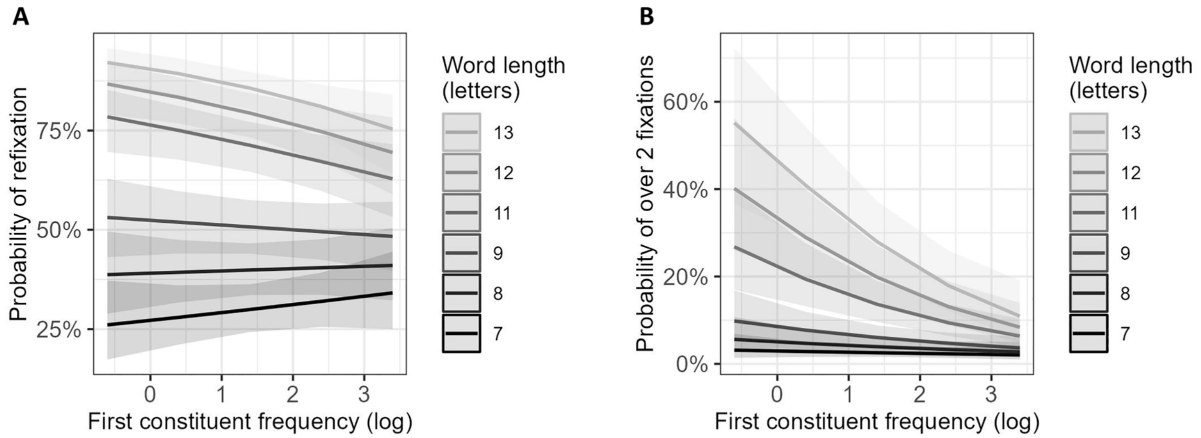

Fig. 2. A: The interaction between word length and first-constituent frequency in the probability of refixation in Experiment 1. B: The interaction between word length and first-constituent frequency in the probability of making more than two fixations in Experiment 1. The shaded areas are 95% CIs.

Although the gaze duration results were perfectly in line with the visual acuity principle, the data on first fixation duration failed to confirm the prediction. We predicted the first-constituent frequency effect to show up early on for long compound words. However, the predicted interaction between word length and first-constituent frequency failed to reach significance in first fixation duration; in fact, the *t* value was practically zero. The same was true for single fixation duration, although one should bear in mind that long compound words were read with a single fixation only in 25 % of the time. Nevertheless, it suggests that sometimes even long compound words can be identified as holistic units. It turned out that most of the long compound words read with a single fixation (69 %) were among the shortest long compound words in our target word set (11 or 12 letters). However, for most long compound words (i.e., 75 %), first constituent frequency exerted an effect that appeared with a slight delay, as it emerged in the probability of making a second fixation on the word. The effect lingered on, as evidenced by the word length × first-constituent frequency interaction being present also in the probability of making more than two fixations on the word.

Experiment 2

Experiment 2 was designed as a further test of the visual acuity principle. More specifically, it predicts that when long compound words are presented in a condensed font so that all letters are within foveal reach, the first-constituent frequency effect should no longer exist. Experiment 2 was analogous to Experiment 1 in all other respects except that now the long compound words were presented in a condensed font so that their horizontal extent was identical to that of the short compound words. In other words, when fixated, all compound words fit in the fovea.

Methods

Participants. Fifty-two university students and other community members, all native speakers of Finnish (38 female), participated in the experiment. The data of one participant had to be discarded due to equipment failure. None of them participated in Experiment 1.

Apparatus. The same eye-tracker was used as in Experiment 1. The stimuli were presented on an Alienware AW2720H monitor (resolution: 1920 × 1080; refresh rate 144 Hz).

Materials. The same materials were used as in Experiment 1, except that now the long compound word sentences were presented in a condensed font to make the horizontal extent of the long compound words comparable to that of the short compound words. The long compound word sentences were presented in Gill Sans MT Condensed 18 font and the short compound words were presented in Lucida Console 15 font. The average horizontal extent of the short compound words was 2.4 degrees of visual angle and that of the long compound words presented in the condensed font 2.5 degrees of visual angle. An example sentence pair is presented in Fig. 3.

Procedure. The experimental procedure was identical to that of Experiment 1 except that the short and long compound word sets were presented as separate blocks. The order of blocks was counterbalanced across participants. The participants responded to the true–false statements with a 94 % accuracy.

Results

An analogous set of analyses were conducted as in Experiment 1. The descriptive statistics of the dependent variables are presented in Table 4. The results of the (generalized) mixed linear models are presented in Table 5. The same exclusion criteria were used as in Experiment 1. In the analysis of first fixation duration, 4.7 % of the trials were discarded; in the gaze duration analysis the percentage was 2.5.

Single and first fixation duration

Participants read the long target words with a single fixation in 48 % of the time, while the short compound words were read with a single fixation in 62 % of the time. No significant effects emerged for single fixation duration. In first fixation duration, the only significant effect was the main effect of word length (see Table 5). It was somewhat longer for long than short compound words.

Gaze duration

The model for gaze duration revealed a main effect of word length and first constituent frequency. Although the first-constituent frequency effect was not reliably modulated by word length ($t = 1.53$), as is apparent from Fig. 4, the pattern looks similar to that observed in Experiment 1. That is, the longer the word (now purely in terms of number of letters), the greater the first-constituent frequency effect (see also the pooled analysis below).

Probability of refixation

As is apparent from Table 5, in the probability of refixation only the main effect of word length proved significant. The more letters the word entailed, the more likely it attracted a refixation.

Short compound: `Oletettavasti upouusi huvivene oli eksynyt reitiltään kartanlukuvirheen takia.`

Long compound: Oletettavasti upouusi synnytysosasto oli määrä avata jo ennen vuodenvaihdetta.

Fig. 3. Graphic depiction of the font manipulation in Experiment 2. The sentences including short compounds were written in a monospaced Lucida Console font, while the sentences including long compound words were written in Gill Sans MT Condensed font. The target word is underlined for illustration.

Table 4
Means and standard deviations (in parentheses) in Experiment 2 by participants, computed on all the observations. Long compound words were presented with condensed font rendering their horizontal extent analogous to that of short compound words.

		First-constituent frequency quantile					
Measure	Word type	1	2	3	4	5	6
SFD	Short	253	245	236	247	243	247
		(58)	(51)	(51)	(61)	(66)	(63)
	Long	251	254	252	268	236	249
		(82)	(67)	(64)	(61)	(53)	(64)
FFD	Short	210	205	212	205	211	226
		(53)	(53)	(61)	(45)	(42)	(46)
	Long	219	213	218	220	224	222
		(57)	(46)	(54)	(53)	(57)	(65)
GD	Short	355	319	350	334	337	339
		(106)	(91)	(109)	(109)	(113)	(114)
	Long	462	416	424	362	373	413
		(166)	(157)	(149)	(110)	(128)	(147)
Refix. prob.	Short	.43	.30	.40	.37	.38	.38
		(.26)	(.21)	(.23)	(.25)	(.26)	(.29)
	Long	.59	.52	.57	.44	.47	.53
		(.27)	(.25)	(.28)	(.26)	(.24)	(.28)
> 2 fix.	Short	.12	.08	.13	.08	.10	.07
		(.17)	(.12)	(.14)	(.13)	(.17)	(.13)
	Long	.26	.19	.19	.08	.11	.18
		(.24)	(.20)	(.22)	(.14)	(.15)	(.24)

Note: SFD = single fixation duration; FFD = duration of first of multiple fixations; GD = gaze duration; Refix. prob. = probability of refixation; > 2 fix. = probability of more than two fixations.

Table 5
Linear mixed effect models for single fixation duration, first fixation duration and gaze duration and generalized mixed linear models for probability of refixation and probability of making more than two fixations in Experiment 2 with word length and first-constituent frequency (both continuous variables) entered as the fixed effects. Statistically significant effects appear in bold.

Measure	Predictor	b	SE	z/t
SFD	Word length (letters)	0.006	0.004	1.419
	First constituent frequency (log)	−0.008	0.011	−0.722
	Word length × First constituent frequency	0.001	0.005	0.285
FFD	**Word length (letters)**	**0.009**	**0.004**	**2.089**
	First constituent frequency (log)	0.010	0.010	1.021
	Word length × First constituent frequency	−0.004	0.004	−0.897
GD	**Word length (letters)**	**0.038**	**0.005**	**6.976**
	First constituent frequency (log)	**−0.033**	**0.015**	**−2.287**
	Word length × First constituent frequency	−0.009	0.006	−1.532
Refix.	**Word length (letters)**	**0.176**	**0.025**	**7.150**
	First constituent frequency (log)	−0.098	0.060	−1.621
	Word length × First constituent frequency	−0.017	0.025	−0.671
> 2 fix.	**Word length (letters)**	**0.187**	**0.033**	**5.687**
	First constituent frequency (log)	**−0.283**	**0.097**	**−2.904**
	Word length × First constituent frequency	−0.037	0.040	−0.936

Note: SFD = single fixation duration; FFD = duration of first of multiple fixations; GD = gaze duration; Refix. = probability of refixation; > 2 fix. = probability of making more than two fixations.

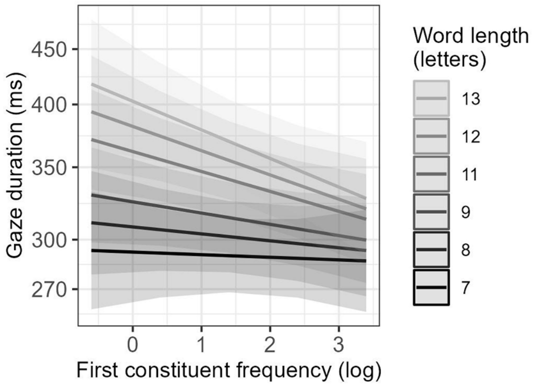

Fig. 4. The interaction between word length and first-constituent frequency in gaze duration of Experiment 2. The shaded areas are 95% CIs.

Probability of making more than two fixations

Both main effects (word length and first constituent frequency) proved significant for the probability of making more than two fixations. The probability increased as a function of word length and decreased as a function of first constituent frequency.

Pooled analyses of gaze durations in Experiment 1 and 2

The Word Length × First Constituent Frequency interaction was present in Experiment 1, but it was non-significant in Experiment 2. In order to further examine the possible effect of the horizontal extent for reading long compound words, we computed pooled analyses of the gaze duration data obtained in Experiment 1 and 2. We computed four different models, one for the complete data set (Model 1), one separately for the long compound words (Model 2), one where possible modulation of initial landing position was examined (Model 3), and one for the short compound words (Model 4). The estimates of the models are reported in Table 6.

Moreover, in order to test the credibility of null results, we also computed analyses of Bayes Factors. They were computed using the lmBF function from the Bayes Factor package in R (Morey & Rouder, 2015). Following Abbott and Staub (2015) and Yao, Staub and Li (2022), we used the default prior for the effect size (Cauchy priors with scale value of 0.5) and ran 100,000 Monte Carlo iterations. In order to use the lmBF function, we removed the random slopes from the models. This may increase the probability of Type I error, but decrease the probability of Type II error (Oberauer, 2022). The decreased probability for Type II error means that we have a lower chance to find support for the null hypothesis; thus, we are less likely to make a mistake in accepting it. According to Wagenmakers, Love, Marsman et al. (2018), Bayes Factors ranging between 0.33 ∼ 1 provide anecdotal evidence to support the null hypothesis, less than 0.33 suggests moderate evidence to support the null hypothesis, and less than 0.1 indicates strong evidence to support the null hypothesis. The detailed analyses of Bayes Factors are available in the OSF.

In Model 1 for the complete data set, the Word Length × First Constituent Frequency interaction proved significant. However, it was not significantly modified by the experiment. In other words, presenting the long compound words in a condensed font did not modulate the effect. The Bayes Factor analysis provided strong evidence (BF = 0.08) against

Table 6

Linear mixed effects models for gaze duration of compound words in Experiment 1 and 2. Word length and first constituent frequency were entered as continuous variables and experiment as a dichotomous variable. In Model 3, initial fixation location was entered as a continuous variable. Statistically significant effects appear in bold.

Measure	Predictor	b	SE	t
GD (Model 1)	**Word length (letters)**	**0.061**	**0.005**	**11.189**
	First constituent frequency (log)	**−0.043**	**0.015**	**−2.898**
	Experiment	−0.080	0.050	−1.600
	Word length × First constituent frequency	**−0.016**	**0.006**	**−2.629**
	Word length × Experiment	**−0.023**	**0.005**	**−4.837**
	First constituent frequency × Experiment	0.010	0.010	0.961
	Word length × First constituent frequency × Experiment	0.007	0.004	1.585
GD (Model 2)	**Word length (letters)**	**0.079**	**0.017**	**4.596**
	First constituent frequency (log)	**−0.071**	**0.027**	**−2.645**
	Experiment	**−0.131**	**0.055**	**−2.410**
	Word length × First constituent frequency	−0.018	0.024	−0.780
	Word length × Experiment	−0.010	0.012	−0.858
	First constituent frequency × Experiment	0.027	0.019	1.425
	Word length × First constituent frequency × Experiment	0.008	0.017	0.514
GD (Model 3)	**Word length (letters)**	**0.061**	**0.025**	**2.455**
	First constituent frequency (log)	**−0.081**	**0.037**	**−2.190**
	Experiment	**−0.165**	**0.058**	**−2.859**
	Initial fixation location	**−0.337**	**0.079**	**−4.254**
	Initial fixation location × Word length	0.068	0.059	1.154
	Initial fixation location × First constituent frequency	0.052	0.087	0.598
	Initial fixation location × Experiment	0.138	0.100	1.382
	First constituent frequency × Word length	−0.039	0.027	−1.454
	First constituent frequency × Experiment	0.048	0.036	1.338
	Word length × Experiment	0.014	0.026	0.545
	Initial fixation location × Word length × First constituent frequency	0.064	0.046	1.376
	Initial fixation location × Word length × Experiment	−0.091	0.072	−1.261
	Initial fixation location × First constituent frequency × Experiment	−0.083	0.103	−0.799
GD (Model 4)	Word length (letters)	0.014	0.030	0.470
	First constituent frequency (log)	−0.012	0.019	−0.644
	Experiment	−0.027	0.048	−0.562
	Word length × First constituent frequency	0.010	0.027	0.384
	Word length × Experiment	0.011	0.022	0.505
	First constituent frequency × Experiment	0.002	0.014	0.131
	Word length × First constituent frequency × Experiment	0.006	0.020	0.307

Note: GD = gaze duration.

an interaction with experiment. On the other hand, experiment modulated the size of the word length effect. Gaze durations increased less steeply in Experiment 2 (long compound words appeared in condensed font) as a function of the number of letters than in Experiment 1.

In Model 2, we tested the modulation of the first-constituent frequency effect by font size more directly by limiting the analysis only to the long compound words. Similar to Model 1, we found no evidence for such modulation (BF = 0.34), only a main effect of first constituent frequency. On the other hand, the font manipulation decreased gaze durations, as indicated by the main effect of experiment. Gaze durations were shorter on long compound words presented in condensed font.

In Model 3, we examined whether the landing position of the initial fixation would modify the effect of first constituent frequency. More precisely, we tested whether the effect would be stronger when the initial fixation is positioned in the word beginning. Such modification may be predicted by the visual acuity principle, which assumes foveal constraints to determine the degree of compositional processing of compound words. The analysis was done for the pooled data of the long compound words. The factors were initial fixation location, first constituent frequency, word length and experiment. Initial fixation location was entered as the relative location (0–1). The model included all the two-way interactions between the predictors and the three-way interactions including initial fixation location. The main effect of initial fixation location was significant; the further into the word the initial fixation was located, the shorter the gaze duration. More importantly, however, initial fixation location did not interact with first constituent frequency ($t < 1$). Moreover, all the other interactions involving initial fixation location were non-significant (see Table 6).

Finally, Model 4 proved that the short compound words of Experiment 1 and 2 showed no significant effects ($ts < 1$). That is, first constituent frequency exerted no effect on reading short compound words (BF = 0.31 for the main effect; BF = 0.09 for the interaction).

Discussion

In Experiment 2, we tested the prediction derived from the visual acuity principle that when long compound words are presented in a condensed font, readers would be able to recognize them holistically, wiping out the first-constituent frequency effect. For the probability of refixation, the font type manipulation exerted an effect in the predicted direction; it showed no reliable effect related to first constituent frequency. It is also noteworthy that the refixation probability was much lower for reading the long compound words in condensed than normal font (.52 vs..75; see Table 2 and 4), indicating that the font manipulation did exert a robust effect on reading. However, in gaze duration the pattern of effects observed for long compound words did not turn out as predicted. Moreover, the pooled analysis of the two Finnish experiments demonstrated no significant modulation of the first-constituent frequency effect by the font size manipulation. Instead, a reliable effect of first constituent frequency was obtained for long compound words, which was absent for short compound words. Nor did the location of the initial fixation on the long compound words modify the effect first-constituent frequency observed in gaze duration.

Similar to Experiment 1, single fixation duration showed no first constituent frequency effect in Experiment 2. In other words, first constituent frequency did not modulate the recognition of long compound words read with a single fixation. Analogously to Experiment 1, the majority of them (67 %) were 11- or 12-letter words (the shortest in our long compound word set).

The visual acuity principle does not appear to provide a completely accurate description of compound word reading in Finnish. It correctly predicts that word length modulates the compound word recognition with the compound word components playing a significant role in recognizing long compound words but not in recognizing short compound words (see the results of Experiment 1). However, Experiment 2 indicates that its explanation for this pattern is likely to be wrong. It appears that it is not the foveal reach that is the key factor, but instead the number of letters the compound word entails. The more letters in a compound word, the more likely it is that its recognition is initiated by accessing the first component.

The above conclusion is generally consistent with the study of McDonald (2006), who scaled the spatial extent of the words (2–9 letters in length) in sentences so that all occupied an identical space horizontally. In an eye-tracking reading experiment, he found a linear increase in fixation time on the words as a function of the number of letters. In other words, the matching of horizontal extent did not wipe out the number-of-letters effect. McDonald suggests that visual crowding is a plausible explanation for the source of the effect. Due to crowding, individual letters may be harder to recognize in longer than shorter words, thus delaying the fixation time on longer words. Similar to McDonald, Hautala, Hyönä and Aro (2011) observed a reliable number-of-letters effect when contrasting gaze durations on four- and six-letter words that were either controlled or not controlled for spatial extent using different fonts. They also found that the number-of-letters effect was not modulated by font, which suggests that number of letters is a more important factor than spatial extent.

We conclude that the compound word length in number of letters modulates the nature of compound word processing in Finnish. Yet, it is noteworthy that our data suggests that it begins to do so when the compound word is rather long, more than 10 letters (see Fig. 1A). It is also noticeable that gaze durations on our short compound words did not vary as a function of number of letters (see Model 4 in Table 6). That is, the short compound words of 6–9 letters were read with similar gaze durations. It compares favorably with the results of a large lexical decision study (New, Ferrand, Pallier, & Brysbaert, 2006; about 800 participants and 40 000 English words); lexical decision times were very similar for words from 5 to 9 letters long. We return to this issue in the General Discussion.

Experiment 3

Experiment 3 was similar to Experiment 1 except that it was conducted on Chinese. A set of two-character compound words were selected that were comparable to the Finnish compound words used in Experiment 1. In choosing the Chinese materials, we sought modifier-head compound word pairs that were similar in meaning across the two languages and had similar frequency characteristics for first and second character as well as for the whole word. This resulted in most Chinese-Finnish word pairs being similar in meaning. However, in order to match them for different frequency characteristics, we needed to choose a few Chinese compound words that did not have a meaning counterpart in Finnish. Matching across the languages was done separately for the word sets that were short and long in Finnish. It is important to note, however, that all two-character Chinese compound words in Experiment 3 were of equal length, as each character occupies the same amount of space in text. Thus, the words in the two sets were equal in length.

In Experiment 3, we used a font size and viewing distance to make the two-character compound words fit in the fovea when fixated. Thus, the visual acuity principle predicts no effect of first character frequency. As summarized in the Introduction, the available evidence regarding character frequency effects in Chinese reading is somewhat mixed, although the recent evidence favors the view that two-character Chinese compound words would be recognized holistically, which is in line with the visual acuity principle. Also the CRM model of Li and Pollatsek (2020) assigns less relevance to individual characters than character combinations in reading multi-character compound words. According to CRM, both the whole-word representation and the embedded words within multi-character words become initially activated. However, the whole-word representation is more strongly activated, as it receives activation from the embedded words, while the activation for the embedded words wanes. It is possible that when compound words are infrequent, as was the case in the present study, whole-word representations become active with more delay, leaving space for embedded words (i.e., single-character words) becoming relatively more activated (Yan et al., 2006). This kind of reasoning may explain the inverted character frequency effects obtained for reading infrequent two-character Chinese words (Cui et al., 2021; Yu et al., 2021).

It is noteworthy that we used as the frequency measure for the first component its frequency as a separate word, not its frequency as a character. This was done for two reasons. First, this way the manipulation was comparable to that in Finnish. Second, Li et al. (2022) argue that one possible reason for the somewhat mixed results regarding character frequency effects in Chinese is that the studies vary in the frequency measure used, with some studies using the character frequency where its frequency as an independent single-character word is not considered, whereas other studies have limited its frequency to instances where it appears as a single-character word. Li et al. (2022) recommend using the latter alternative, as it is a more sensible way to estimate the relative role of single-character versus whole-word representations in Chinese word recognition. Thus, we followed their recommendation.

Similar to Experiment 1, the target compound words were embedded in single sentences. The sentence frame leading to the target was semantically neutral, so it did not constrain the target word identity. Participants' eye movements were tracked when they read the sentences silently for comprehension.

Methods

Participants. Fifty-six university students (47 female, 9 male) from Shandong Normal University participated in the experiment. All of them were native speakers of Chinese with normal or corrected to normal vision.

Apparatus. Readers' eye movements were recorded with a desk-mounted Eyelink 1000 plus (SR Research, Ontario, Canada) eye-tracker using 1000 Hz sampling rate. Movements of the right eye were tracked. The tracker was calibrated using a 3-point calibration. The sentences were presented in Simple Song 28 font in black on a white background. Each character was about 1.2×1.2 cm^2 in size. The distance between the participant and the screen was 60 cm and each character subtended approximately 1.2° of visual angle.

Materials. A set of 120 two-character Chinese compound words were used as the stimuli. To make them comparable to Finnish compound words used in Experiment 1, we sought compound words whose first components were of similar frequency and meaning across the two languages. Moreover, we sought for two-character compound words whose first-character frequency as a single-character word varied considerably. The matching was done separately with respect to compound words that were short and long in Finnish. However, as all Chinese characters are of equal length, the length manipulation was of no significance in Chinese. There were 60 short and 60 "long" compound words. Similar to Finnish, first-constituent frequency was manipulated in a continuous fashion using six frequency quantiles. The frequency values of the Chinese compound words are presented in Table 7. Word and character frequencies were computed on the basis of the corpus of Cai and Brysbaert (2010).

The log frequency of the first character varied significantly between the quantiles, $F(5,108) = 15.94$, $p < .001$. However, it did not differ between the short and "long" compounds, $F < 1$. Whole-word frequency was generally low and it did not vary between the two word sets or between the quantiles of first-character frequency, $Fs < 1$. Moreover, second-character frequency did not systematically vary between the short and "long" words, $F(1,108) = 1.67$, $p = .20$, or between the quantiles of first-character frequency, $F < 1$. Finally, the number of strokes in a word did not systematically vary as a function of first-character frequency and word set, $Fs < 1.28$. We also tested whether the frequency variables varied systematically between Chinese and Finnish. This was not the case for first-character frequency ($F < 1$ for all effects involving language) or second-character frequency ($F < 1.32$ for all effects involving language). On the other hand, whole-word frequency was statistically marginally higher for Chinese, $F(1,202) = 3.20$, $p = .08$. However, it was low in both languages.

Analogously to Experiment 1, two words from the same first-constituent frequency quantile were paired, one from the short and one from the "long" compound word set, and a sentence frame was written for the pair so that the sentence segment leading to the target word was identical within each pair. The target word always appeared in the middle of the sentence. This time we also attempted to equate the sentence frame after the target word. In the majority of sentence pairs,

Table 7
Logarithmic frequencies (per million values in parentheses) of 1st constituent, 2nd character and whole word for short and "long" Chinese compound words for the six frequency quantiles.

	Quantile 1	Quantile 2	Quantile 3	Quantile 4	Quantile 5	Quantile 6	Total
		1st character frequency					
Short	.36 (3.52)	1.04 (15.6)	1.67 (55.4)	1.98 (102)	2.35 (307)	2.69 (902)	1.68 (231)
"Long"	.09 (1.73)	1.15 (15.7)	1.64 (51.3)	2.05 (118)	2.23 (180)	2.70 (622)	1.65 (165)
		2nd character frequency					
Short	1.70 (67.1)	1.70 (63.4)	1.68 (262)	1.70 (278)	1.50 (65.6)	1.66 (116)	1.66 (142)
"Long"	1.55 (48.0)	1.58 (63.2)	1.49 (92.9)	1.28 (66.5)	1.46 (62.7)	1.49 (187)	1.48 (87)
		Whole-word frequency					
Short	.11 (3.86)	.07 (1.89)	-.04 (2.79)	-.03 (2.33)	.04 (1.88)	.004 (1.66)	.03 (2.40)
"Long"	-.27 (0.78)	-.05 (1.23)	-.09 (1.07)	.13 (3.75)	-.10 (1.63)	.03 (1.58)	-.06 (1.67)
Number of strokes							
Short	17.0	18.8	16.9	13.6	16.0	14.9	16.2
"Long"	18.0	17.4	18.2	18.8	15.0	15.4	17.1

we succeeded to do so, but for some sentence pairs the sentence end was not identical after the target. Close matching of sentence ends after the target was attempted, because the Chinese script is dense so that more information can fit in the parafoveal area than in alphabetic scripts. Thus, the parafoveal information is more likely to affect foveal processing. This had the consequence of not being able to present both sentence pairs to the same participants. Instead, two sentence lists were prepared so that one member of each sentence pair was presented in one list and the other member appeared in the other list. This way each participant read 60 sentences that contained 30 short and "long" compound words. Below, an example sentence pair is presented with an English translation (the target word is underlined for illustration).

Short compound: **唐敏一脚踩进污水里弄脏了鞋子.**

Long compound: **唐敏一脚踩进沙坑里弄脏了鞋子.**

Tang Min stepped into the dirty water and soiled her shoes.

Tang Min stepped into the sand pit and soiled her shoes.

We tested the predictability of the target words by asking 16 native Chinese speakers who did not participate in the experiment to continue each sentence frame by a word that first came to their mind. The sentence frames leading to the target word were presented. We carried out predictability ratings instead of concordance searches done for Finnish, as we could not locate a Chinese corpus, where suitable concordance searches could have been done. The ratings revealed that the target words were highly unpredictable. Only one target word was given as a sentence continuation once, all other words had no mentions.

Procedure. The participants were asked to read the sentences silently for comprehension at their normal reading speed. After 25 % of the sentences, their comprehension was assessed by a true–false statement. They were instructed to press the blue key in the gamepad when the statement was congruent with the meaning of the last read sentence and to press the green key in the gamepad, when it was incongruent. Participants responded correctly to the majority of the statements (the accuracy was 96 %). The target sentences were preceded by 8 practice sentences, 2 of which were followed by a true–false statement. The target sentences were presented in two blocks, which were counterbalanced using a Latin square design such that participants saw each sentence and target word only once. The whole experiment lasted about 15 min.

Results

Analogously to the Finnish experiments, gaze duration was considered the primary measure, as it captures the recognition of the whole compound words. Gaze durations shorter than 80 ms or longer than 1400 ms were removed from the analyses. A total of 0.9 % of the gaze duration data were eliminated. We also conducted separate analyses for the gaze duration on the first and second character to examine whether the first-character frequency effect would be short-lived and apparent only during the reading of single characters. Moreover, the probability of skipping over a single character was also analyzed, as about half of single characters are typically not fixated. Conducting separate analyses for the first and second character allowed us to reveal possible inverted character frequency effects observed previously for second characters as a function of first-character frequency (Cui et al., 2021).

LME models were formed analogously to Experiment 1 with one exception. Word length was a categorical variable, as it referred to two word sets that were matched to the long and short Finnish compound words. In reality, all two-constituent Chinese compound words were of equal length. The descriptive statistics of the dependent variables are given in Table 8.

Gaze duration on the whole word

The results of the LME models are summarized in Table 9. As is apparent from Table 9, in gaze duration the main effects and their interaction all remained clearly non-significant (*ts* < 1.1; see also Fig. 5). To test the credibility of the null effects, we ran Bayes Factors analyses similar to those reported for the pooled analysis of the Finnish results. The analysis revealed strong evidence (BF = 0.08) against a main effect of first-character frequency and moderate evidence against a main effect of word "length" (BF = 0.13) and the interaction (BF = 0.14). The Bayes

Table 8
Means and standard deviations (in parentheses) in Experiment 3 by participants, computed on all the observations.

		First-character frequency quantile					
Measure	Word length	1	2	3	4	5	6
GD/word	Short	283	312	317	287	277	287
		(73)	(117)	(82)	(72)	(90)	(76)
	"Long"	291	298	333	320	278	289
		(72)	(96)	(101)	(108)	(94)	(86)
GD/1st	Short	242	252	269	252	223	242
		(65)	(85)	(94)	(63)	(61)	(67)
	"Long"	234	246	243	246	244	245
		(61)	(76)	(52)	(72)	(76)	(85)
ISP/1st	Short	.44	.45	.48	.49	.57	.47
		(.26)	(.23)	(.28)	(.21)	(.22)	(.22)
	"Long"	.43	.48	.47	.45	.48	.47
		(.28)	(.20)	(.26)	(.26)	(.28)	(.25)
GD/2nd	Short	231	239	243	231	251	243
		(74)	(88)	(60)	(64)	(72)	(71)
	"Long"	248	251	277	247	226	262
		(87)	(79)	(84)	(72)	(71)	(79)
ISP/2nd	Short	.53	.46	.44	.42	.49	.49
		(.23)	(.24)	(.24)	(.23)	(.29)	(.18)
	"Long"	.54	.51	.40	.41	.54	.51
		(.21)	(.23)	(.26)	(.25)	(.26)	(.23)

Note: GD/word = gaze duration on the whole word; GD/1st = gaze duration on the first character; ISP/1st = probability of skipping the first character; GD/2nd = gaze duration on the second character; ISP/2nd = probability of skipping the second character.

Table 9
Linear mixed effect models for gaze duration for the whole word, first character and second character and a generalized mixed linear model for initial skipping probability in Experiment 3 with word length (dichotomous variable) and first-character frequency (continuous variable) entered as the fixed effects. Word length refers to two matched sets of compound words identical in length (see text for details).

Analysis region	Measure	Predictor	b	SE	t/z
Whole word	GD	Word length	0.032	0.031	1.029
		First character frequency (log)	−0.004	0.011	−0.406
		Word length × First character frequency	0.006	0.015	0.408
First character	GD	Word length	−0.012	0.023	−0.505
		First character frequency (log)	−0.008	0.008	−1.059
		Word length × First character frequency	0.012	0.011	1.045
	ISP	Word length	−0.089	0.092	−0.965
		First character frequency (log)	0.037	0.032	1.152
		Word length × First character frequency	−0.021	0.045	−0.473
Second character	GD	Word length	0.041	0.025	1.637
		First character frequency (log)	0.009	0.009	1.030
		Word length × First character frequency	−0.001	0.013	−0.106
	ISP	Word length	0.054	0.099	0.552
		First character frequency (log)	−0.026	0.034	−0.766
		Word length × First character frequency	−0.003	0.049	−0.067

Note: GD = gaze duration; ISP = initial skipping probability.

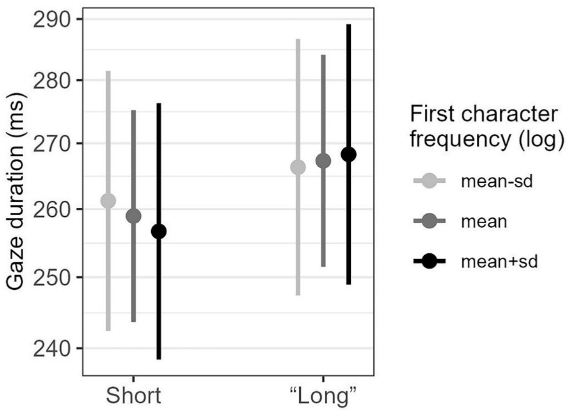

Fig. 5. Gaze duration (ms) on the target compound words in Experiment 3 as a function of word length and first character frequency (log). The "long" word set was the matched set of the long compound words tested in Finnish in Experiment 1 and 2. In reality, the short and "long" compound words were identical in length.

Factors analyses are reported in full in the OSF repository.

Gaze duration and skipping rate for the first and second character

Gaze duration on the first and second character demonstrated no significant effects. Both the main effect of first constituent frequency and its interaction with word length were far from significant (*ts* < 1.6; see Table 9). The evidence for the null effects was moderate to strong (BF = 0.09–0.27).

As single characters are often skipped, it is possible that instead of appearing in gaze duration, it may be observable in the probability of skipping first or second character. Thus, we analyzed the probability of skipping the first and second character using general linear effects models. First character frequency did not affect skipping probability, nor did it interact with word length either for first or second character (*ts* <.1.2; see Table 9). The Bayes Factor analyses showed that the evidence for null effects was strong (BF = 0.07–0.10).

Discussion

The results of Experiment 3 conducted in Chinese are straightforward. There was no evidence that first character frequency affected reading of two-character compound words (see also Cui et al., 2013, 2017; Li et al., 2014; Ma et al., 2015). This became apparent in the gaze duration for the whole word as well as in the gaze duration and skipping probability for the first and second character. The lack of character frequency effect is consistent with the visual acuity principle (Bertram & Hyönä, 2003), the CRM model of Chinese reading (Li & Pollatsek, 2020) and the MCU hypothesis (Zang, 2019).

According to the visual acuity principle, when a compound word does not fit in the fovea, the morphological decomposition route becomes more prominent in compound word processing. The claim can be further tested in Chinese by increasing the text's font size. This was done in Experiment 4, where we enlarged the font size of the so-called long compound words of Experiment 3. These compound words were closely matched to the long Finnish compound words tested in Experiment 1 and 2. If the visual acuity principle holds true, we should observe in Experiment 4 a first-character frequency effect for the long compound words, materializing as a first-character frequency by word length (i.e., font) interaction.

Experiment 4

In Experiment 4, the same materials were used as in Experiment 3 except that the so-called long compound words of Experiment 3 were now in reality also much longer than the short compound words. This was done by enlarging their font size so that their horizontal extent was analogous to that of Finnish long compound words of Experiment 1. Now both characters of long compound words were 2° of visual angle so the long compound words' horizontal extent was 4°, using the viewing distance of 60 cm. Thus, their horizontal extent was twice the size of short compound words, for which each character extended horizontally about 1° of visual angel.

Experiment 4 was designed as a further test of the visual acuity principle (Bertram & Hyönä, 2003). It predicts an effect of first character frequency for long Chinese compound words, but not for the short ones.

Methods

Participants. Fifty university students (40 female, 10 male) from Shandong Normal University participated in the experiment. All of them were native speakers of Chinese with normal or corrected to normal vision. None of them participated in Experiment 3.

Apparatus. The same apparatus was used to record readers' eye movements as in Experiment 3.

Materials. The same materials were used as in Experiment 3, except that the so-called long compound word set of Experiment 3 extended now almost twice the size horizontally to how they were presented in Experiment 3. This was achieved by presenting them in a larger font size (Simple Song 38 font) so that the target compound words extended horizontally 4° of visual angle. The so-called short compound word set was presented in the same Simple Song 28 font as in Experiment 3, which meant that they subtended 2.4° horizontally. The sentences in small and large fonts were displayed in two separate blocks. The order of blocks was counterbalanced across participants.

Procedure. The experimental procedure was identical to that in

Experiment 3. The mean comprehension accuracy was 95 %.

Results

Gaze durations shorter than 80 ms or longer than 1400 ms were removed from the analyses. A total of 1.0 % of the data were eliminated. LME models were formed analogously to Experiment 3 with word length manipulated by font size being a categorical variable and first-character frequency a continuous variable. In the following, we refer to the word length manipulation by font size (small or large), as it was behind the difference in word length. The descriptive statistics of the dependent variables are reported in Table 10.

Gaze duration on the whole word

The results of the LME models are summarized in Table 11. As is apparent from Table 11, in the gaze duration on the whole word the main effects and their interaction all remained clearly non-significant (*ts* < 1; see also Fig. 6). The Bayes Factors analyses showed strong evidence against the main effect of first character frequency (BF = 0.09) and font size (BF = 0.09) and moderate evidence against the interaction (BF = 0.17).

Gaze duration and skipping rate for first and second character

Similar to Experiment 3, we also conducted separate analyses for the gaze duration on the first and second character to examine whether the first-character frequency effect would be short-lived and apparent only during the reading of single characters. Gaze duration on first character revealed a significant main effect of font size and first character frequency (see Table 11). As is apparent from Fig. 7, first characters of two-character compound words presented in large font size were read with shorter gaze durations than first characters appearing in small font size. Moreover, infrequent first characters resulted in longer gaze durations on them than frequent characters. On the other hand, gaze duration on the second character showed no significant effects (*ts* < 1; BF = 0.09–0.27).

Table 10
Means and standard deviations (in parentheses) in Experiment 4 by participants, computed on all the observations.

Measure	Font size	First-character frequency quantile 1	2	3	4	5	6
GD/ word	Small	295 (87)	324 (119)	306 (86)	279 (82)	272 (75)	308 (110)
	Large	288 (80)	285 (75)	294 (87)	307 (79)	280 (86)	295 (93)
GD/1st	Small	239 (58)	250 (80)	239 (55)	231 (56)	204 (48)	225 (54)
	Large	233 (56)	225 (55)	222 (51)	223 (48)	206 (46)	219 (60)
ISP/1st	Small	.52 (.29)	.51 (.25)	.46 (.24)	.46 (.26)	.56 (.31)	.54 (.24)
	Large	.45 (.23)	.42 (.24)	.45 (.23)	.36 (.24)	.5 (.26)	.47 (.24)
GD/2nd	Small	242 (71)	238 (61)	239 (62)	231 (70)	239 (73)	269 (113)
	Large	232 (57)	212 (48)	242 (67)	244 (74)	240 (83)	258 (79)
ISP/2nd	Small	.44 (.25)	.50 (.23)	.41 (.20)	.50 (.25)	.47 (.26)	.43 (.25)
	Large	.48 (.24)	.44 (.24)	.38 (.23)	.43 (.23)	.46 (.25)	.46 (.25)

Note: GD/word = gaze duration on the whole word; GD/1st = gaze duration on the first character; ISP/1st = initial skipping probability for the first character; GD/2nd = gaze duration on the second character; ISP/2nd = initial skipping probability for the second character.

Table 11
Linear mixed effect models for gaze duration for the whole word, first character and second character and a generalized mixed linear model for initial skipping probability in Experiment 4 with font size (dichotomous variable) and first-character frequency (continuous variable) entered as the fixed effects. Statistically significant effects appear in bold.

Analysis region	Measure	Predictor	b	SE	t/z
Whole word	GD	Font size	−0.009	0.034	−0.281
		First character frequency (log)	−0.009	0.012	−0.752
		Font size × First character frequency	0.011	0.017	0.670
First character	GD	**Font size**	**−0.053**	**0.023**	**−2.263**
		First character frequency (log)	**−0.017**	**0.008**	**−2.030**
		Font size × First character frequency	0.007	0.012	0.638
	ISP	**Font size**	**−0.296**	**0.097**	**−3.061**
		First character frequency (log)	−0.003	0.034	−0.088
		Font size × First character frequency	0.005	0.048	0.104
Second character	GD	Font size	−0.012	0.025	−0.467
		First character frequency (log)	0.005	0.009	0.529
		Font size × First character frequency	0.012	0.013	0.931
	ISP	Font size	−0.086	0.100	−0.857
		First character frequency (log)	−0.004	0.035	−0.121
		Font size × First character frequency	−0.012	0.049	−0.250

Note: GD = gaze duration; ISP = initial skipping probability.

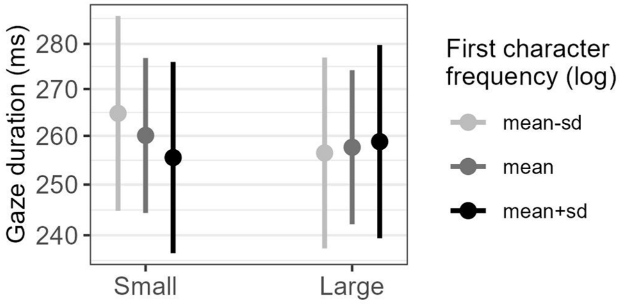

Fig. 6. Gaze duration (ms) on the target compound words in Experiment 4 as a function of font size and first character frequency (log).

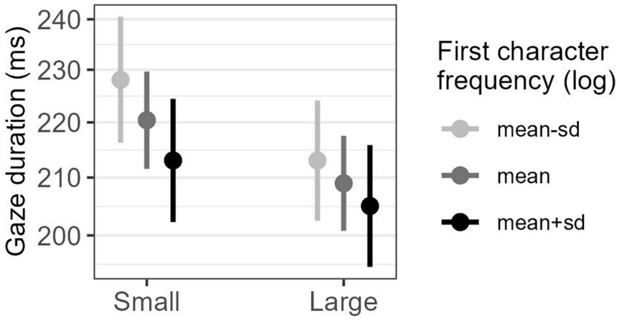

Fig. 7. Gaze duration (ms) on the first character of two-character compound words in Experiment 4 as a function of font size and first-character frequency (log).

We also analyzed the probability of skipping first and second character using general linear effects models. The probability of skipping first character demonstrated a main effect of font size (see Table 11). As is apparent from Fig. 8, first characters of two-character compound words were skipped more often when appearing in small than large font. However, the main effect of first character frequency (BF = 0.06) and the interaction (BF = 0.10) were non-significant in the probability of skipping first character. The probability of skipping the second character revealed no effects (BF = 0.07–0.10).

Pooled analyses of the "long" compound words of Experiment 3 and 4

Even though the results of Experiment 4 provided no evidence for the modulation of the first-character frequency effect by word length, as predicted by the visual acuity principle, we ran further tests of the possible modulation by pooling the data for the long compound word set. If the effect of first character frequency was indeed modulated by word length, we would observe an interaction between first-character frequency and font size. Gaze durations shorter than 80 ms or longer than 1400 ms were removed from the analyses. A total of 1.0 % of the data was eliminated.

The complete set of analyses can be found in OSF. No significant effects involving first character frequency emerged in the analyses (BF = 0.06–0.44). In the analysis of first fixation duration on the first character, a trend ($t = 1.85$) was observed for an interaction between first character frequency and font size. However, the Bayes Factors analysis yields anecdotal evidence for a null result (BF = 0.34).

Discussion

In Experiment 4, we tested the prediction derived from the visual acuity principle (Bertram & Hyönä, 2003), according to which when the entire compound word does not fit in the fovea, the recognition process is initiated by the access of the first compound word character. We tested the prediction in Chinese by enlarging the font size so that a subset of compound words subtended horizontally about 4 degrees of visual angle comparable to the long Finnish compound words tested in Experiment 1. The results did not confirm the prediction. Gaze duration on compound words presented in large font size did not show an effect of first constituent frequency. The same was true for compound words presented in small font, thus confirming the main result of Experiment 3.

The font size manipulation did affect the probability of skipping over the first character. It was more probable to skip over the first character when it was presented in small font. This makes sense in that in the small font the first character was closer to the fovea thus making possible its parafoveal recognition. However, first character frequency did not influence the skipping probability in either font size.

The font size manipulation also produced a bit less intuitive result. Gaze duration on the first character was shorter when presented in large than small font size. It would be tempting to interpret it to suggest that reading is facilitated by presenting the text in large font. However, this appears not to be the case, as there was no effect of font size in the gaze duration on the whole word. Nor did font size affect the gaze duration on the second character. Instead, there seems to be a trade-off between fixation time and probability of skipping. When the first character was presented in small font, it was more often skipped, but when fixated, the fixation time was a bit longer than for the first characters presented in large font that were skipped less. In other words, readers may skip over easily recognizable first characters presented in small font and fixate on the relatively more difficult ones, resulting in longer gaze durations on those. With large font, the easily recognizable first characters were fixated more often, leading to relatively shorter gaze durations.

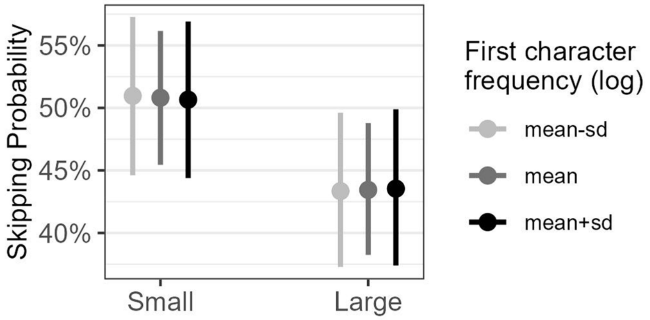

Fig. 8. Probability of skipping the first character of two-character compound words in Experiment 4 as a function of font size and first-character frequency (log).

Finally, even though first character frequency did not affect the gaze duration on whole word, it did so for the gaze duration on the first character. Fixation time was longer when the first character was infrequent than frequent. The effect was present for both font sizes (i.e., no interaction). Finding it for the large font size is consistent with the visual acuity principle. However, it was not expected to emerge also for the small font size. The latter finding is also inconsistent with the results of Experiment 3 where no such effect was visible.

One possible explanation for this effect is that the large font size used for half of the materials encouraged readers to pay relatively more attention to individual characters even when they appeared in small font. In order to test this, we carried out analyses where we added the block order as a factor (large font first versus large font second). If the above suggestion is correct, block order should modify[3] the first-character frequency effect observed in gaze duration on the first character. In other words, the effect should be larger, when the sentences in large font were presented as the first block. However, this did not turn out to be the case; the interaction was not significant ($t = 1.37$). The analysis can be found in the OSF repository.

The present study is not the only study that has observed somewhat inconsistent results regarding character frequency effects in Chinese reading (see the Introduction for references). Many factors may contribute to the inconsistencies (for example, measure of character frequency, grammatical structure and semantic transparency of compound words). However, they cannot explain the inconsistency in the present study, as the target compound words were identical in the two experiments. What is noteworthy, the character frequency effect in Experiment 4 appeared as a fleeting effect only observed in gaze duration of the first character but not in gaze duration of the whole word. The pattern of results where character frequency may have a fleeting effect in compound word processing is readily explained by the CRM model (Li & Pollatsek, 2022). We return to this issue in the General Discussion.

To sum up, the results of Experiment 4 provided no consistent support for the visual acuity principle. Thus, it is fair to say it does not provide a plausible account of compound word recognition in Chinese.

General discussion

Four eye-tracking experiments were conducted to study reading of two-constituent compound words in Finnish and Chinese. The two scripts differ notably from each other with the Finnish script being a transparent alphabetic script while the Chinese script is logographic in nature. On the other hand, they have one feature in common: In both languages, the majority of words in a dictionary are compound words. Thus, in order to understand word recognition during reading of these scripts, it is important to understand how compound words are processed. The present study was designed to shed light on this issue.

Being an orthographically transparent script, each phoneme is denoted in Finnish by a separate grapheme. This feature makes the words in Finnish generally longer than those in many other scripts. The

---

[3] We thank Barbara Juhasz for suggesting this analysis.

contrast is particularly salient with respect to Chinese, where characters are visually dense. According to the visual acuity principle (Bertram & Hyönä, 2003), visual density may have a significant impact on compound word processing. When all letters or characters fit in the fovea where visual acuity is at its best, compound words may be recognized as holistic units. On the other hand, when part of the compound words falls outside the fovea, componential processing is assumed to get a head start.

The visual acuity principle correctly predicted the results of Experiment 1 conducted in Finnish and those of Experiment 3 conducted in Chinese. An effect of first component frequency, a litmus test of componential processing, was obtained for long Finnish compound words but not for short Finnish compound words. A subset of the letters of long compound words were outside foveal reach, whereas all letters of short compound words fitted in the fovea. Analogously, when both characters of Chinese compound words fitted in the fovea, first character frequency exerted no effect.

However, in the subsequent experiments the visual acuity principle fared more poorly. When long Finnish compound words were presented in a condensed font (Experiment 2) so that all the letters fit in the fovea, the first-constituent frequency effect did not cease to exist. This became apparent in the pooled analysis, where the font size did not reliably modulate the effect. Moreover, when a subset of Chinese compound words was presented in a large font (Experiment 4) so that the words were partly outside foveal reach when fixated, the first-character frequency effect was still absent. It did have a fleeting early effect in the fixation time on the first character, which disappeared in the gaze duration on the whole word.

In what follows, we outline what the present results tell about compound word reading in Finnish and Chinese. We end the article by discussing script differences in compound word recognition.

Recognizing compound words in Finnish

The results of the present study suggest that word length modulates the recognition of Finnish compound words, as posited by the visual acuity principle (Bertram & Hyönä, 2003). Short compound words are recognized holistically as suggested by the lack of a first-constituent frequency effect, whereas for long compound words the constituents comprising the compound word play a significant role. However, contrary to the visual acuity principle, foveal constraints are not the only determiner in modulating the nature of compound word processing. The present study demonstrates that the number of letters also plays a significant role. This finding is generally consistent with the results reported by McDonald (2006) and Hautala et al. (2011), who found that word length in terms of number of letters determines fixation times on words, but not word length in terms of horizontal extent. Yet, it is important to note that in the present study the modulation of compound word recognition in Finnish was noticeable for words longer than 10 letters, whereas it was absent for short (6–9 letters) compound words (see Fig. 1A and 4). Thus, the results for short compound words may be explained by foveal constraints (all letters fit in fovea when the word is fixated), as suggested by the visual acuity principle. However, this is not the case for long compound words, as they continued to show a first-constituent frequency effect even when all letters fitted in the fovea (see the condensed font manipulation in Experiment 2).

There appears to be two aspects with respect to the recognition of long compound words. On one hand, when all letters are not within foveal reach, the process is initiated by the access of the first constituent. Here, foveal constraint is likely to be a significant determiner. On the other hand, even when all letters fit in the fovea, there appears to be a limit to the number of letters that can be processed in parallel. Based on the present results, the limit appears to be 10 letters. Interestingly, it compares favorably with the study of New et al. (2006), who in a lexical decision experiment found no or a tiny number-of-letters effect in reaction times for 5–9 -letter words (see their Fig. 2). Yet, a linear increase was observed for 10–13 -letter words. It should be noted that the stimuli of New et al. represented different grammatical categories (nouns, adjectives, verbs, function words); in other words, their effect is not confined to compound words. Thus, the suggested processing limit, if it exists, is general in nature and not specific to morphologically complex words.

In conclusion, we suggest that there may be a limit in the number of letters that can be simultaneously processed and that the limit is not only perceptual (reflecting foveal constraints) but also more cognitive in nature. Thus, the recognition of long compound words (exceeding 10 letters) is inherently decompositional in nature. Future studies may be carried out to further test the claim regarding readers' capacity limitations.

Recognizing compound words in Chinese

The Chinese results lend support to the CRM model (Li & Pollatsek, 2020) of Chinese reading, which posits that compound words are recognized primarily as holistic units. As the Chinese script is non-spaced, segmenting words from other words is an integral part of the recognition process. According to the model, all words within the reader's perceptual span become activated. Two-character combinations receive more activation than single-character words, as they receive activation from both of its characters. Thus, the two-character combination wins the race leading to its recognition and segmentation from the other neighboring characters. Single-character words embedded in the two-character words may receive some activation, but it is short-lived (see the results of Experiment 4 and the pooled analyses) and is inhibited by the activation of the two-character combination.

The Chinese results are also generally compatible with the MCU hypothesis (Zang, 2019; Zang et al., 2021, 2023), according to which not only multiple-character words but also other multi-constituent units (e.g., idioms) will be identified as single units. The compatibility comes from the fact that in the present materials the compound words were composed of characters that were single-character words themselves. Yet, the results suggest that the two-character combinations were readily identified as compound words, not as single-character words. This was despite the fact that in Chinese word boundaries are not demarcated by spaces or other visual signals.

It is likely that the script's visual density significantly contributes to the holistic nature of compound word recognition in Chinese. In typical Chinese script used in experimental studies, two-character compound words are within foveal reach. The present study demonstrates that holistic processing prevails even when part of the word appears in the parafovea, as was the case for compound words presented in large font. Interestingly, the font size manipulation had no overall effect in gaze durations that were highly similar for the two font sizes (around 300 ms). In other words, pushing the second character toward the parafovea did not delay word recognition. It is likely due to the drop in visual acuity being compensated by parafoveal magnification of the second character. This in turn is consistent with the view that readers' perceptual span is governed more by attentional constraints than by visual acuity (Miellet, ÓDonnell, & Sereno, 2009).

It is also noteworthy that the lack of character frequency effects was observed for infrequent compound words (average word frequency was about 1.6 per million). In other words, holistic processing is not confined to frequent compound words, but it also extends to the recognition of relatively infrequent compound words. Thus, the present study demonstrates that holistic processing prevails in Chinese compound word recognition across different font sizes and can also be observed in recognizing infrequent compound words.

Finally, it is important to note that the claim that compound word recognition in Chinese is predominantly holistic in nature does not mean that component characters would have no role to play in compound word recognition. An integral part of word recognition is the segmentation of adjacent characters into words. In other words, as there are no spaces between words, Chinese readers need to decide where one word ends and the next begins. Here the positional probability of the second character of two-character compound words plays a role (Liang, Gao, Li, Wang, Bai, & Liversedge, 2023). When the second character frequently appears as the second character of two-character words, it speeds up segmentation resulting in faster compound word recognition. Moreover, the relationship between the component characters may affect the recognition of Chinese compound words (Cui, Cong, Wang, Zhang, Zheng & Hyönä, 2018). If so, compound word recognition does not solely entail the access to the compound word representation in the mental lexicon, but it may also include relational processing between the components. It is left for the future research to determine whether such relational processing is part and parcel of lexical access or whether the availability of relational information is an end product of lexical access (i.e., it becomes available after lexical access).

Script differences in compound word recognition

In their review of script-universal and script-specific processes in reading, Li et al. (2022) argue that word length may be a more important factor in processing compound words in alphabetic than logographic scripts. This is because the horizontal extent of compound words is generally greater and more variable in alphabetic than logographic scripts. The present results are consistent with this claim. The variability in horizontal extent is intimately linked to visual density; the logographic script is denser than the alphabetic script. This has consequences to the foveal availability of words once they are fixated during reading. The visual density of the Chinese script makes the holistic processing of compound words more likely than with compound words in alphabetic languages like Finnish where they are often long.

Yet, it is noteworthy that the mere visual complexity in terms of number of strokes does not seem a major determiner of word reading times in Chinese. Gaze durations on two-character compound words that differed in number of strokes were found to be similar in length (Fu, Liversedge, Bai, Moosa, & Zang, 2023). This contrasts with the present findings for Finnish, where number of letters significantly affected compound word reading. Strokes and letters are similar in that additional strokes and letters both increase the word’s visual complexity (Liversedge et al., 2023). However, their function is different. Letters represent sublexical components, whereas individual strokes have no such function, as only a combination of strokes represent lexical information. Thus, it is understandable that individual strokes play a smaller role in compound word recognition than individual letters.

As the word and constituent frequencies were carefully matched across the two languages, compound word recognition times may be used to estimate the relative ease of compound word reading between the languages. It turns out that compound words are faster to recognize in Chinese than Finnish. The average gaze duration across the two experiments was 300 ms for Chinese, whereas it was considerably greater for the long Finnish compound words (432 ms) but less so for short Finnish compound words (344 ms). The notable difference in gaze durations between Chinese compound words and long Finnish compound words partly reflects differences in visual density, which in turn has consequences to the words’ foveal availability. However, a notable difference remains even when comparing gaze durations for Chinese compound words presented in large font (295 ms) to those observed when reading long Finnish compound words in condensed font (407 ms). This difference cannot be explained by foveal constraints. Rather, it may reflect qualitative differences in compound word processing: more holistic in Chinese, but more decompositional in Finnish.

Conclusion

Significant differences emerged in compound word recognition during reading of alphabetic (Finnish) and logographic (Chinese) script. In Finnish, compound words longer than 10 letters were recognized using the morphological decomposition route (recognition is initiated by accessing the first compound word constituent), whereas short compound words were recognized via the holistic route. Long Finnish compound words were recognized via the decomposition route regardless of whether or not all the letters fit in the foveal vision. In Chinese, on the other hand, the holistic route prevailed for reading two-character compound words. This was even when they were written in large font, where only a part of the compound word can fit in foveal vision when fixated. The observed pattern of results cannot be readily explained by the visual acuity principle of Bertram and Hyönä (2003).

CRediT authorship contribution statement

Jukka Hyönä: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Lei Cui:** Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Timo T. Heikkilä:** Formal analysis, Software, Validation. **Birgitta Paranko:** Data curation, Formal analysis, Software, Visualization, Investigation. **Yun Gao:** Formal analysis, Investigation, Resources. **Xingzhi Su:** Investigation, Resources.

Declaration of Competing Interest

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

Data availability

Data are available in the Open Science Framework (osf.io/9tc3j).

Acknowledgements

This work was supported by the Academy of Finland (grant number 315963) given to the first author. Lei Cui was supported by a grant from the Social Science Fund of Shandong Province (22CYYJ16) and a grant from the Natural Science Fund of Shandong Province (ZR2020MC222). We thank Annimaaria Kyyhkynen and Anniina Laurila for running Experiment 2. The authors claim no conflicts of interest.

References

Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E-Z Reader’s null predictions. *Journal of Memory and Language, 85*, 76–87. https://doi.org/10.1016/j.jml.2015.07.002

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv. 1506.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48*, 615–634. https://doi.org/10.1016/S0749-596X(02)00539-9

Bertram, R., & Hyönä, J. (2013). The role of hyphens at the constituent boundary in compound word identification: Facilitative for long, detrimental for short compound words. *Experimental Psychology, 60*(3), 157–163. https://doi.org/10.1027/1618-3169/a000183

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One1, 5*(6), e10729.

Chen, H. C., Song, H., Lau, W. Y., Wong, K. F. E., & Tang, S. L. (2003). Developmental characteristics of eye movements in reading Chinese. In C. McBride-Chang, & H. C. Chen (Eds.), *Reading development in Chinese children* (pp. 157–169). Praeger.

Cui, L., Cong, F., Wang, J., Zhang, W., Zheng, W., & Hyönä, J. (2018). Effects of grammatical structure of compound words on word recognition in Chinese. *Frontiers in Psychology, 9*, 258. https://doi.org/10.3389/fpsyg.2018.00258

Cui, L., Häikiö, T., Zhang, W., Zheng, Y., & Hyönä, J. (2017). Reading monomorphemic and compound words in Chinese. *The Mental Lexicon, 12*, 1–20. https://doi.org/10.1075/ml.12.1.01cui

Cui, L., Wang, J., Zhang, Y., Cong, F., Zhang, W., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology, 74*, 610–633. https://doi.org/10.1177/1747021820973661

Cui, L., Yan, G., Bai, X., Hyönä, J., Wang, S., & Liversedge, S. P. (2013). Processing of compound-word characters in reading Chinese: An eye movement- contingent display change study. *Quarterly Journal of Experimental Psychology, 66*, 527–547. https://doi.org/10.1080/17470218.2012.667423

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. https://doi.org/10.1037/xlm0000013

Fu, Y., Liversedge, S.P., Bai, X., Moosa, M., & Zang, C. (2023). Character representations in lexical identification during natural reading. Manuscript submitted for publication.

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research, 51*, 1719–1727. https://doi.org/10.1016/j.visres.2011.05.015

Hyönä, J. (2015). Are polymorphemic words processed differently from other words during reading? In A. Pollatsek, & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 114–128). New York: Oxford University Press.

Hyönä, J., Pollatsek, A., Koski, M., & Olkoniemi, H. (2020). An eye-tracking study of reading long and short novel and lexicalized compound words. *Journal of Eye Movement Research, 13*(4), 3. https://doi.org/10.16910/jemr.13.4.3

Juhasz, B. J. (2008). The processing compound words in English: Effects of word length on eye movements during reading. *Language and Cognitive Processes, 23*, 1057–1088. https://doi.org/10.1080/01690960802144434

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 & Performance, 35*, 876–895. https://doi.org/10.1037/a0013484

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

Li, X., Huang, L., Yao, P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology, 1*, 133–144. https://www.nature.com/articles/s44159-022-00022-6.

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

Liang, F., Gao, Q., Li, X., Wang, Y., Bai, X., & Liversedge, S. P. (2023). The importance of the positional probability of word final (but not word initial) characters for word segmentation and identification in children and adults' natural Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*, 98–115. https://doi.org/10.1037/xlm0001116

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition, 147*, 1–20. https://doi.org/10.1016/j.cognition.2015.10.013

Liversedge, S. P., Olkoniemi, H., Zang, C., Li, X., Bai, X., Yan, G., & Hyönä, J. (2023). Universality in eye movements and reading: A replication with increased power. *Cognition*, in press. https://doi.org/10.1016/j.cognition.2023.105636.

Luotolahti, J., Kanerva, J., Laippala, V., Pyysalo, S., & Ginter, F. (2015). Towards universal web parsebanks. *Proceedings of the International Conference on Dependency Linguistics* (Depling'15).

Ma, G., Li, X., & Rayner, K. (2015). Readers extract character frequency information from non-fixated target word at long pretarget fixations during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 41*(5), 1409–1419. https://doi.org/10.1037/xhp0000072

McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Visual Cognition, 13*, 89–98. https://doi.org/10.1080/13506280500143367

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science, 20*, 721–728. https://doi.org/10.1111/j.1467-9280.2009.02364.x

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for Common designs* (Version 0.9.12.2). http://cran.at.r-project.org/web/packages/BayesFactor/index.html.

New, B., Ferrand, R., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*, 45–52. https://doi.org/10.3758/BF03193811

Oberauer, K. (2022). The importance of random slopes in mixed models for Bayesian hypothesis testing. *Psychological Science, 33*, 648–665. https://doi.org/10.1177/09567976211046884

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception & Performance, 26*, 820–833. https://doi.org/10.1037/0096-1523.26.2.820

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspect of language processing* (pp. 131–154). Erlbaum.

Siegelman, N., Schroeder, S., Acartürk, C., et al. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods, 54*, 2843–2863. https://doi.org/10.3758/s13428-021-01772-6

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning & Verbal Behavior, 15*, 607–620.

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 761–775. https://doi.org/10.1037/0278-7393.23.3.761

Tsang, Y., Huang, J., Lui, M., Xue, M., Chan, Y. W. F., Wang, S., et al. (2018). MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavioral Research Methods, 50*, 1763–1777. https://doi.org/10.3758/s13428-017-0944-0

Tse, C. S., & Yap, M. J. (2018). The role of lexical variables in the visual recognition of two-character Chinese compound words: A megastudy analysis. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 71*(9), 2022–2038. https://doi.org/10.1177/1747021817738965

Wagenmakers, E. J., Love, J., Marsman, M., et al. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*, 58–76. https://doi.org/10.3758/s13423-017-1323-7

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. https://doi.org/10.1348/000712605X70066

Yao, P., Staub, A., & Li, X. (2022). Predictability eliminates neighborhood effects during Chinese sentence reading. *Psychonomic Bulletin & Review, 29*, 243–252. https://doi.org/10.3758/s13423-021-01966-1

Yu, L., Liu, Y., & Reichle, E. D. (2021). A corpus-based versus experimental examination of word- and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of Experimental Psychology: General, 150*(8), 1612–1641. https://doi.org/10.1037/xge0001014

Zang, C. (2019). New perspectives on serialism and parallelism in oculomotor control during reading: The multi-constituent unit hypothesis. *Vision, 3*(50), 1–13. https://doi.org/10.3390/vision3040050

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2021). Foveal and parafoveal processing of Chinese three-character idioms in reading. *Journal of Memory and Language, 119*, Article 104243. https://doi.org/10.1016/j.jml.2021.104243

Zang, C., Fu, Y., Du, H., Bai, X., Yan, G., & Liversedge, S. P. (2023). Processing multi-constituent units: Preview effects during reading of Chinese words, idioms and phrases. *Journal of Experimental Psychology*. https://doi.org/10.1037/xlm0001234. Learning, Memory, & Cognition, in press.