The effect of character contextual diversity on eye movements in Chinese sentence reading

Qingrong Chen¹ · Guoxia Zhao¹ · Xin Huang¹ · Yiming Yang² · Michael K. Tanenhaus^{1,3}

Published online: 30 March 2017
© Psychonomic Society, Inc. 2017

Abstract Chen, Huang, et al. (Psychonomic Bulletin & Review, 2017) found that when reading two-character Chinese words embedded in sentence contexts, contextual diversity (CD), a measure of the proportion of texts in which a word appears, affected fixation times to words. When CD is controlled, however, frequency did not affect reading times. Two experiments used the same experimental designs to examine whether there are frequency effects of the first character of two-character words when CD is controlled. In Experiment 1, yoked triples of characters from a control group, a group matched for character CD that is lower in frequency, and a group matched in frequency with the control group, but higher in character CD, were rotated through the same sentence frame. In Experiment 2 each character from a larger set was embedded in a separate sentence frame, allowing for a larger difference in log frequency compared to Experiment 1 (0.8 and 0.4, respectively). In both experiments, early and later eye movement measures were significantly shorter for characters with higher CD than for characters with lower CD, with no

effects of character frequency. These results place constraints on models of visual word recognition and suggest ways in which Chinese can be used to tease apart the nature of context effects in word recognition and language processing in general.

Keywords Reading · Contextual diversity · Character frequency · Eye movements

Introduction

Word frequency effects have long played a central role in developing and evaluating models of visual word recognition and reading. Although models differ in how they account for frequency effects, most have assumed that lexical representations encode some form of token frequency.

Recently, there has been increased interest in studying word recognition in character-based orthographies because differences from alphabetic orthographies can be used to address a range of theoretical issues in reading (e.g., Rayner, Li, & Pollatsek, 2007). In Chinese, two-character words comprise 72% of the lexicon (Lexicon of Common Words in Contemporary Chinese Research Team, 2008). Studies of frequency effects for two-character words show both word-level frequency effects and, with word-level frequency controlled, smaller frequency effects for the initial character (e.g., Shen & Li, 2012; Yan, Tian, Bai, & Rayner, 2006). This is consistent with models in which, like alphabetic orthographies, the word is the primary unit but with independent effects at the character-level, especially for word-initial characters, which are often semantically related to the entire word.

Emerging evidence suggests that measures that take into account the contexts in which a word are likely to occur provide plausible alternative explanations for effects previously attributed to token frequency. Thus it becomes crucial to

This paper was published with some of the Chinese characters appearing incorrectly in the online version. The original article has been corrected.

- ✉ Qingrong Chen
jscqr80@sina.com
- ✉ Yiming Yang
yangym@jsnu.edu.cn
- ✉ Michael K. Tanenhaus
mtanenha@ur.rochester.edu

¹ School of Psychology, Nanjing Normal University, Nanjing, China

² School of Linguistic Sciences, Jiangsu Normal University, Jiangsu, China

³ Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

establish the empirical foundations for context-dependent measures. Adelman and colleagues (2006) operationalized a measure, *contextual diversity* (CD) – the proportion of texts in a corpus in which a word occurs. CD is typically correlated with frequency: High frequency words typically occur in a wider variety of contexts than lower frequency words. When Adelman, Brown, and Quesada (2006) manipulated frequency, while controlling for CD, and vice versa, naming and lexical decision times were affected by CD but not by token frequency. Subsequent research suggests that the underlying dimension reflected in CD effects may be the semantic distinctiveness of a word used across contexts (Jones, Johns, & Recchia, 2012), which will generally be correlated with CD. We return to this issue in the “General discussion”.

Two recent eye-movement studies found CD effects and the absence of frequency effects (CD-only effects) for words embedded in sentences in English (Plummer, Perea, & Rayner, 2014) and Chinese (Chen, Huang, Xu, Yang, & Tanenhaus, 2017). Chen et al. provide the strongest evidence. First, they demonstrated a CD-only effect in a design which rotated triples of control words, frequency-matched words with higher CD values, and CD-matched words with lower frequency values through the same sentence frames, thus ruling out potential confounds from differences in sentence frames. Second, small differences that were consistent with a residual frequency effect did not increase with an expanded frequency manipulation.

The current study uses the same manipulations as Chen et al. (2017) with two-character Chinese words to examine first character frequency and CD effects, when overall word CD is controlled. The possibilities for first-character effects are: (1) both frequency and CD effects; (2) frequency-only effects; (3) CD-only effects; and (4) neither frequency nor CD effects. Each outcome would place potential constraints on models of word recognition in sentence contexts, and suggest different avenues for research that would likely shed new light on long-standing theoretical debates (e.g., how context influences different aspects of visual word recognition). We consider implications of a few possible outcomes to highlight the theoretical motivation for examining character-based frequency and CD.

Classic word recognition models assume multiple levels of representation. For Chinese, these include radicals and, for two character-Chinese words, the component characters. Whereas word-level CD-only effects could reflect post-lexical integration and/or word-level expectations, character-level frequency effects would be consistent with models in which characters, as pre-lexical units, are primarily influenced by bottom-up processing, and thus sensitive only to token frequency. In contrast, CD-only character effects, especially if they were localized to first-pass measures, would be consistent with prediction-based models (e.g., Smith & Levy, 2013) in which the first character in parafoveal vision has special status. The absence of both character-based frequency and

CD effects would be consistent with models in which all prediction and/or integration effects are word-based.

We controlled word frequency and CD, while independently manipulating frequency and CD of the first character. Experiment 1 rotated yoked triples through the same sentence frame to establish whether there are frequency and/or CD effects, while ruling out any confounding effects of sentence-context. Experiment 2 used a separate sentence frame for each character, allowing for more words with a larger frequency range and longer and more varied sentence frames.

Experiment 1: Contextual diversity effect on characters

We compared three groups of characters: Higher frequency characters with lower CD (Control group); lower frequency characters (LCF group) with a similar CD as the Control group but lower frequency; and higher-CD (HCD) group matched in frequency with the Control group but with higher CD. We adopted this design because there is a dearth of characters with high CD and low frequency (also see Hills, Maouene, Riordan, & Smith, 2010; Perea, Soares, & Comesaña, 2013; Plummer et al., 2014). Comparisons of primary interest are the LCF group with the Control group, which would reveal any effects of frequency with CD controlled, and the HCD group with the Control group, which would reveal any effects of CD with frequency controlled.

Method

Participants

Participants were 30 native speakers of Mandarin with normal or corrected-to-normal vision. Participants received informed consent and were paid 20 RMB.

Materials

Critical characters were chosen from SUBTLEX-CH-CHR (Cai & Brysbaert, 2010), which uses 6,243 films to compute character frequency based on the log10 transformed number of occurrences from 46 million characters and CD based on the log10 transformed number of films in which the character appears. Frequencies from this corpus explain significantly more of the variance in word and character reading than frequencies based on written texts (see Cai & Brysbaert for details).

We selected 27 character triples for each condition (see Table 1). Characters in the LCF group had a similar CD to the control condition ($t(52) = 1.43, p = 0.16$) but a lower character frequency, $t(52) = 18.10, p < 0.001$. Characters in the HCD group had a similar character frequency to the control condition ($t(52) = -0.58, p = 0.56$) but a higher CD, $t(52)$

Table 1 Experimental conditions and exemplar sentences in Experiment 1

Condition	Example	Plausibility
Control	妈妈把 佐 料放到了柜子里。 Mother put the condiment in the cupboard	5.34 (0.51)
LCF	妈妈把 绷 带放到了柜子里。 Mother put the bandage in the cupboard	5.20 (0.60)
HCD	妈妈把 抹 布放到了柜子里。 Mother put the duster in the cupboard	5.42 (0.48)

Note: English glosses are literal translations. The target characters are in bold. English translations of all test sentences for Experiment 1 and Experiment 2 are available by request from the first author (jscqr80@sina.com)

= −12.50, $p < 0.001$. Characters were matched in number of strokes ($ts < 0.64$, $ps > 0.53$), radicals ($ts < 1.22$, $ps > 0.22$), and orthographic neighborhood size ($ts < 0.68$, $ps > 0.50$). All targets were the first character of compound nouns. The characters after the target were matched in frequency ($ts < 1.09$, $ps > 0.28$), CD ($ts < 1.03$, $ps > 0.30$), stroke ($ts < 0.79$, $ps > 0.43$), radical ($ts < 0.41$, $ps > 0.68$), and orthographic neighborhood size ($ts < 1.10$, $ps > 0.27$) across conditions (see Table 2). We

also controlled frequency ($ts < 0.48$, $ps > 0.63$), CD ($ts < 1.10$, $ps > 0.28$), and semantic diversity ($ts < 1.07$, $ps > 0.29$) of the two-character word that began with the target character. We created 27 sentence frames and rotated target characters from each triple through the same frame across three lists such that each list contained nine sentences from each condition, with one version of each sentence.

Control norms

Using a 7-point scale, norming studies with 30 participants who rated plausibility (see Table 1) revealed no significant differences between conditions, $ts < 0.92$, $ps > 0.36$. Character predictability was assessed using 20 students in a cloze completion task with the target word replaced by a blank. The target character appeared in only 2.36% of the completions, with no significant differences across conditions ($ts < 0.22$, $ps > 0.83$). Concreteness values obtained from 30 participants on a 7-point scale (Table 2) revealed no significant differences across conditions for the target characters ($ts < 1$, $ps > 0.34$) and the two-character words they were embedded in ($ts < 1.09$, $ps > 0.28$).

Table 2 Characteristics of the target in each group

Characteristic	Control group	LCF group	HCD group
Target character			
Mean character frequency (log)	3.00 (0.02)	2.59 (0.01)	3.02 (0.02)
Character frequency range (log)	2.89–3.22	0.43–2.68	2.89–3.23
Mean contextual diversity (log)	2.50 (0.02)	2.46 (0.01)	2.84 (0.02)
Contextual diversity range (log)	2.23–2.62	2.36–2.53	2.69–3.06
Mean contextual diversity (standard, %)	5.15 (0.22)	4.69 (0.10)	11.25 (0.49)
Contextual diversity range (standard, %)	2.74–6.70	3.68–5.48	7.93–18.34
Number of strokes	9.33 (0.60)	9.93 (0.72)	9.30 (0.65)
Number of radicals	3.04 (0.24)	3.22 (0.21)	2.67 (0.19)
Orthographic neighborhood size	326.33 (59.35)	303.30 (67.76)	386.26 (65.45)
Concreteness	5.27 (0.29)	5.59 (0.31)	5.68 (0.31)
Post-target character			
Mean character frequency (log)	3.89 (0.18)	3.86 (0.14)	4.16 (0.17)
Mean contextual diversity (log)	3.27 (0.13)	3.34 (0.09)	3.43 (0.08)
Mean contextual diversity (standard, %)	52.13 (6.79)	50.18 (6.12)	60.00 (7.10)
Number of strokes	7.70 (0.68)	6.96 (0.65)	7.15 (0.74)
Number of radicals	2.00 (0.17)	2.07 (0.18)	2.11 (0.22)
Orthographic neighborhood size	296.89 (56.52)	249.82 (67.91)	206.82 (59.80)
Word			
Mean word frequency (log)	1.75 (0.14)	1.65 (0.14)	1.82 (0.15)
Mean contextual diversity (log)	1.50 (0.12)	1.55 (0.14)	1.69 (0.14)
Mean contextual diversity (standard, %)	1.22 (0.26)	1.27 (0.24)	2.16 (0.5)
Semantic diversity	1.56 (0.18)	1.48 (0.15)	1.33 (0.11)
Concreteness	6.37 (0.11)	6.31 (0.18)	6.53 (0.10)

Note. All standard errors are provided in parentheses. According to the *Modern Chinese Dictionary* (2012), semantic diversity was computed based on the number of meanings that a word can have (Liu, Shu, & Li, 2007)

Apparatus

Eye movements were monitored using a SensoMotoric Instruments (Teltow/Berlin, Germany) iView Hi-Speed system, sampling at 1,250 Hz (tracking resolution $< 0.01^\circ$). Viewing was binocular but data were collected only from the right eye. Sentences were presented on a 17-in. CRT monitor. Each character subtended 1.05° of visual angle at a viewing distance of 70 cm.

Procedure

Trials began with presentation of a fixation point left aligned with the first character of each sentence. The order of the 27 experimental sentences and 27 filler sentences was randomized for each participant. Participants were instructed to read silently at a normal rate and answered a yes/no question. The experiment began with eight practice sentences.

Results and discussion

All participants responded correctly on at least 85% of the questions. Prior to analysis, sentences with track losses were excluded (less than 2.49% of trials). Trials with blinks, fixations shorter than 80 ms, longer than 800 ms, and fixations above or below three standard deviations from the mean were deleted (6.73%).

We computed three first-pass measures, considered to primarily reflect early processes, and three measures (see Table 3) which reflect later processes (e.g., Rayner, 1998, 2009). Early measures were first-fixation duration, gaze duration, and skipping rate. Later measures were go-past time, regression rate, and total fixation time.

We implemented planned comparisons using linear mixed-effects models for fixation durations and mixed logit models for skipping rate using the lme4 package (Bates, Maechler, & Bolker, 2012) in R environment (R Development Core Team, 2014). The regression model included fixed effects (e.g., log-transformed character frequency and log-transformed CD) and maximal random effects for participants and items (Barr, Levy, Scheepers, & Tily, 2013; Jaeger, 2008). We report t -

values for linear mixed-effects models, z -values for mixed logit models, and corresponding p -values (see Table 4). For t -values, the lmerTest package was implemented to estimate p values using Satterthwaite approximation for degrees of freedom (Kuznetsova, Christensen, & Brockhoff, 2014).

Effect of character frequency (control group vs. LCF group)

There were no effects of character frequency in first-pass measures, $t_s < 0.86$, $p_s > 0.39$, $\beta_s < 6.20$ (skipping rate, $z = -0.17$, $p = 0.86$, $\beta = -0.08$) and later measures, $t_s < 1.25$, $p_s > 0.22$, $\beta_s < 9.07$ (Reg.in, $z = 1.09$, $p = 0.28$, $\beta = 0.29$). Note that the skipping rates are low compared to some previous studies (e.g., Li, Bicknell, et al., 2014). The reason is that we included a fixation that preceded and was adjacent to a skipped character, which makes it likely that the character was processed on that fixation (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994). We also calculated skipping rates without that fixation. We found similar skipping rates to the most comparable study (Cui, Yan, Bai, Hyönä, Wang, & Liversedge, 2013) using that procedure and repeated all analyses using those data.¹

Effect of contextual diversity (control group vs. HCD group)

For first-pass measures, first-fixation duration ($\beta = -15.64$) and gaze duration ($\beta = -24.69$) were shorter for the HCD characters than for characters with lower CD, respectively ($t_s > -2.28$, $p_s < 0.03$). The effect of CD was not significant in skipping rate ($\beta = 0.54$). For the later measures, HCD target characters had shorter go-past times ($\beta = -28.49$) and total fixation times ($\beta = -33.76$), $t_s > -2.28$, $p_s < 0.007$. There was no difference in regression rate ($\beta = -0.36$).

An additional regression analysis used the eye movement measures as dependent variables and log-transformed character CD and log-transformed character frequency as predictors. There were facilitative effects of character CD on FFD ($\beta = -0.12$, $t = -2.83$, $p = 0.005$), GD ($\beta = -0.16$, $t = -3.77$, $p < 0.001$), TTime ($\beta = -0.15$, $t = -3.45$, $p = 0.001$), and Go-past ($\beta = -0.17$, $t = -3.48$, $p < 0.001$) but no effects of character frequency ($\beta_s < 0.10$, $|t|s/\text{Wald} < 1.50$, $p_s > 0.13$).

In sum, when CD is controlled, character-frequency did not affect reading times, whereas when frequency is controlled, CD affected both first-pass and later measures.

Table 3 Eye movement measures on target characters in each group

Measure	Control group	LCF group	HCD group
FFD (ms)	223 (4.96)	229 (4.90)	207 (4.74)
GD (ms)	284 (6.00)	278 (5.36)	258 (5.71)
Skipping	0.10 (0.02)	0.09 (0.02)	0.12 (0.02)
Go-past (ms)	398 (5.27)	400 (5.23)	369 (4.89)
Reg. in	0.24 (0.02)	0.19 (0.02)	0.20 (0.01)
TTime (ms)	362 (5.05)	371 (5.00)	330 (6.83)

Note. All standard errors are provided in parentheses

¹ When we excluded fixations that were close to the left of a skipped character, skipping rates for Control, LCF, and HCD groups were 0.24, 0.22, and 0.26, respectively. These data are similar to the skipping rate for the first character with high and low frequencies in Cui et al. (2013). Reanalysis of all eye movement measures find the same pattern as analyses that include an adjacent fixation, with no effects of frequency in any of the measures ($|\beta_s| < 7.70$, $|t/zs| < 1.10$, $p_s > 0.33$), and clear effects of CD in FFD ($\beta = -18.68$, $t = -2.94$, $p = 0.007$), GD ($\beta = -28.38$, $t = -3.42$, $p = 0.002$), Go-past ($\beta = -30.69$, $t = -3.74$, $p = 0.001$), and TTime ($\beta = -36.07$, $t = -3.12$, $p = 0.004$).

Table 4 Regression coefficients and test statistics from linear mixed-effects and logistic mixed-effects models for eye movement measures on the target character

		FFD	GD	Skipping	Reg.in	Go-past	TTime
Control group vs. LCF group	Est. β	6.17	−5.13	−0.08	0.29	2.35	9.06
	<i>t</i> -value	0.85	−0.57			0.31	1.24
	<i>z</i> -value			−0.17	1.09		
	<i>p</i> -value	0.399	0.577	0.862	0.28	0.76	0.222
Control group vs. HCD group	Est. β	−15.64	−24.69	0.54	−0.36	−28.49	−33.76
	<i>t</i> -value	−2.28	−2.60			−3.38	−2.93
	<i>z</i> -value			1.26	−0.94		
	<i>p</i> -value	0.028	0.015	0.209	0.35	0.002	0.006

Note. Linear mixed-effects models including two orthogonal contrasts for character frequency and contextual diversity, and covariates (e.g., radical, frequency of two-character word that began with the target character and its CD) were restructured to examine the effects of contextual diversity on the four sets of viewing duration. There were significant effects of character contextual diversity on FFD ($\beta = -21.48$, $t = -2.75$, $p = 0.007$), GD ($\beta = -33.7$, $t = -3.07$, $p = 0.00035$), Go-past ($\beta = -29.96$, $t = -3.17$, $p = 0.004$), and TTime ($\beta = -28.95$, $t = -2.90$, $p = 0.004$). We did not observe effects of character frequency ($|\beta|s < 3.52$, $ts < 0.40$, $ps > 0.65$), word frequency ($|\beta|s < 43$, $ts < 1.60$, $ps > 0.11$), word CD ($|\beta|s < 46$, $ts < 1.60$, $ps > 0.11$), and radical ($|\beta|s < 9$, $ts < 1.70$, $ps > 0.10$) on the eye movement measures

Experiment 2: Contextual diversity effects on characters with large frequency range

In Experiment 1, the direction of the effect for four of the six measures is consistent with small increase in processing difficulty for the LCF characters (see Table 3). If there were an underlying frequency effect, differences between the Control and the LCF conditions, should increase with a larger frequency range. In Experiment 2, we used separate sentence frames for each word, enabling us to increase in the difference in mean log frequencies for the LCF and Control characters from 0.4 to 0.8. This also allowed us to a greater variety of sentence structures than the materials used in Experiment 2.

Method

Participants

Participants were 48 native speakers of Mandarin with normal or corrected-to-normal vision. Participants received informed consent and were paid 20 RMB.

Materials

We selected 48 critical characters, 16 for each of three conditions (see Tables 5 and 6). Characters in the LCF group had a similar CD as the control condition ($t < 1.72$, $p > 0.10$) but with a lower character frequency, $t(30) = 10.59$, $p < 0.001$. Characters in the HCD group had a similar character frequency to the control condition ($t < 0.12$, $p > 0.24$) but with a higher CD value, $t(30) = 9.98$, $p < 0.001$. All targets were the first character of compound nouns, matched in number of strokes ($ts < 0.67$, $ps > 0.50$), radicals ($ts < 1.55$, $ps > 0.13$),

and orthographic neighborhood size ($ts < 0.63$, $ps > 0.53$). Characters prior to and after the target character were matched in frequency ($ts < 1.39$, $ps > 0.17$), CD ($ts < 1.66$, $ps > 0.10$), stroke ($ts < 1.40$, $ps > 0.17$), radical ($ts < 1.45$, $ps > 0.15$), and orthographic neighborhood size ($ts < 1.65$, $ps > 0.38$) across conditions. We also controlled the frequency ($ts < 0.14$, $ps > 0.89$), CD ($ts < 0.92$, $ps > 0.36$), and semantic diversity ($ts < 0.20$, $ps > 0.26$) of the two-character word that began with the target character (Table 5). Each target character was placed in its own sentence frame (see Table 6).

Control norms

Thirty participants rated perceived difficulty and plausibility, using a 7-point scale (see Table 6). Differences between conditions were not significant ($ts < 1.15$, $ps > 0.26$). Character predictability, as assessed by 15 students in a cloze completion task, did not differ across conditions ($ts < 1.09$, $ps > 0.28$). Concreteness ratings by 30 participants using a 7-point scale found no significant differences across conditions for the target characters ($ts < 1$, $ps > 0.54$) and compound words they were embedded in ($ts < 1$, $ps > 0.41$).

Apparatus and procedure

Apparatus and procedure were identical to Experiment 1.

Results and discussion

We used the same exclusion criteria as in Experiment 1 to delete trials with track losses, blinks, and those above or below three standard deviations from the mean (9.6%). Table 7 displays condition means for each eye

Table 5 Characteristics of the target in each group

Characteristic	Control group	LCF group	HCD group
Target character			
Mean character frequency (log)	3.71 (0.06)	2.92 (0.05)	3.80 (0.04)
Character frequency range (log)	3.22–4.12	2.38–3.10	3.41–4.06
Mean contextual diversity (log)	2.94 (0.05)	2.83 (0.05)	3.51 (0.03)
Contextual diversity range (log)	2.41–3.19	2.34–2.97	3.24–3.70
Mean contextual diversity (standard, %)	15.09 (1.37)	11.47 (0.91)	52.95 (3.34)
Contextual diversity range (standard, %)	4.15–24.54	3.48–15.04	27.55–80.15
Number of strokes	9.31 (0.93)	9.94 (0.75)	8.56 (0.63)
Number of radicals	3.00 (0.33)	3.13 (0.22)	2.44 (0.16)
Orthographic neighborhood size	461.69 (104.41)	380.56 (75.86)	315.31 (72.24)
Concreteness	3.89 (0.53)	4.33 (0.47)	4.20 (0.52)
Pre-target character			
Mean character frequency (log)	5.74 (0.12)	5.41 (0.19)	5.54 (0.19)
Mean contextual diversity (log)	3.80 (0.01)	3.78 (0.01)	3.77 (0.01)
Mean contextual diversity (standard, %)	99.79 (0.18)	97.61 (1.31)	95.75 (2.43)
Number of strokes	6.88 (0.51)	6.75 (0.42)	7.69 (0.29)
Number of radicals	2.56 (0.20)	2.50 (0.22)	2.88 (0.13)
Orthographic neighborhood size	89.69 (27.74)	81.06 (31.86)	89.81 (28.50)
Post-target character			
Mean character frequency (log)	3.86 (0.22)	3.95 (0.17)	3.97 (0.15)
Mean contextual diversity (log)	3.25 (0.12)	3.42 (0.09)	3.47 (0.09)
Mean contextual diversity (standard, %)	43.77 (8.78)	54.49 (7.95)	57.84 (7.22)
Number of strokes	7.25 (1.03)	8.75 (0.65)	7.88 (0.69)
Number of radicals	2.06 (0.25)	2.56 (0.24)	2.44 (0.22)
Orthographic neighborhood size	432.25 (90.09)	277.44 (60.63)	319.31 (89.55)
Word			
Mean word frequency (log)	1.89 (0.20)	1.91 (0.20)	1.85 (0.15)
Mean contextual diversity (log)	1.59 (0.20)	1.84 (0.19)	1.75 (0.14)
Mean contextual diversity (standard, %)	2.03 (0.64)	3.69 (1.58)	2.14 (0.95)
Semantic diversity	1.38 (0.16)	1.63 (0.16)	1.56 (0.20)
Concreteness	5.69 (0.30)	5.76 (0.38)	6.01 (0.25)

Note. All standard errors are provided in parentheses

movement measure. Planned comparisons were performed using linear mixed-effects and mixed logit models (see Table 8). We again repeated all analyses with fixations immediately to the left of a skipped target

character excluded (see Table 7). We note that reading times are in general longer for Experiment 2 compared to Experiment 1. This most likely reflects the increased variety, length, and likely complexity of the sentences.

Table 6 Experimental conditions and exemplar sentences in Experiment 2

Condition	Example	Difficulty	Plausibility
Control	拉里说街道两旁回响的凯歌无法换回战友的生命。 Lali said that victorious song that was reverberating on both sides of the street did not let comrade-in-arms live again	1.43 (0.05)	5.50 (0.16)
LCF	我认为过去发生的纠纷可能会明显影响这次合作。 I think the disputes that have occurred in the past may significantly affect this cooperation	1.44 (0.06)	5.43 (0.17)
HCD	妹妹认为这种畅销的坚果非常符合年轻人的口味。 My sister thinks that the best-selling nuts are very suitable for young people's tastes	1.44 (0.05)	5.24 (0.16)

Note: English glosses are literal translations. The target characters are in bold

Table 7 Eye movement measures on target character

Measure	Control group	LCF group	HCD group
FFD (ms)	243 (3.51)	240 (3.52)	227 (2.97)
GD (ms)	342 (6.55)	339 (6.34)	304 (5.48)
Skipping	0.06 (0.01)	0.04 (0.01)	0.07 (0.01)
Go-past (ms)	448 (8.86)	451 (9.16)	398 (7.22)
Reg. in	0.25 (0.02)	0.23 (0.02)	0.22 (0.02)
TTime (ms)	484 (7.51)	490 (6.82)	438 (7.23)

Note. All standard errors are provided in parentheses. When fixations close to the left of the target character were excluded, the skipping rates for the Control, LCF, and HCD groups were 0.21, 0.20, and 0.23, respectively. Analyses with these data found no effects of frequency in any of the measures ($|\beta|s < 11.50$, $t/zs < 1.1$, $ps > 0.29$), and clear effects of CD in FFD ($\beta = -16.47$, $t = -3.10$, $p = 0.004$), GD ($\beta = -37.37$, $t = -3.77$, $p = 0.001$), Go-past ($\beta = -50.95$, $t = -2.88$, $p = 0.008$), and TTime ($\beta = -42.22$, $t = -3.37$, $p = 0.002$)

Effect of character frequency (control group vs. LCF group)

Effects of character frequency were not significant in any of the first-pass ($ps > 0.55$, $\beta s < -3.42$) and later measures ($ps > 0.61$, $\beta s < 6.10$).

Effect of contextual diversity (control group vs. HCD group)

For first-pass measures, first-fixation duration ($\beta = -16.67$) and gaze duration ($\beta = -39.12$) were shorter for the HCD characters than for characters with lower CD, respectively ($ts > -3.13$, $ps < 0.005$), with no main effect of CD on skipping rate, $z = 0.53$, $p = 0.597$, $\beta = 0.26$. For later measures, HCD target characters had shorter go-past time ($\beta = -49.88$) and total fixation time ($\beta = -45.39$), $ts > -2.28$, $ps < 0.007$. The regression rate was numerically, but not significantly, lower for HCD characters ($\beta = -0.12$).

A regression analysis found significant effects of character CD on FFD ($\beta = -0.10$, $t = -3.07$, $p = 0.002$), GD ($\beta = -0.12$, $t = -3.83$, $p < 0.001$), TTime ($\beta = -0.12$, $t = -3.87$, $p < 0.001$), and Go-past ($\beta = -0.12$, $t = -3.58$, $p < 0.001$), but no significant effects of character frequency ($\beta s < 0.30$, $|t|s/Wald < 1.40$, $ps > 0.19$).

General discussion

When word frequency and CD for two-character words were controlled, first-character fixation times were shorter for higher CD characters but were not affected by character frequency. These results complement Chen et al. (2017) who found the same pattern for word-level frequency and CD. These studies establish that CD, a measure that takes into account contexts better accounts for differences that have typically been attributed to token frequency. They go beyond previous findings in demonstrating that the absence of frequency effects in these and previous studies is unlikely to be a consequence of using restricted frequency ranges.

We conclude by outlining some of the theoretical implications. First and foremost, it is important to develop models that can explain both why CD influences word recognition in reading and why effects that have previously been attributed to token frequency are eliminated when CD is controlled. Although standard models of word frequency have assumed that token frequency is a characteristic of form-based lexical representations, it now seems increasingly unlikely that simple form-based (token) frequencies are encoded in lexical representations. We note, however, that future research with measures such as ERP could, in principle, isolate token frequency effects that are not reflected in processing difficulty (cf. Vergara-Martínez, Comesaña, & Perea, 2017)

Table 8 Regression coefficients and test statistics from linear mixed-effects and logistic mixed-effects models for eye movement measures on the target character

		FFD	GD	Skipping	Reg.in	Go-past	TTime
Control group vs. LCF group	Est. β	-3.33	-3.42	-0.08	-0.07	2.03	6.01
	t -value	-0.59	-0.32			0.10	0.50
	z -value			-0.13	-0.50		
	p -value	0.558	0.751	0.896	0.62	0.92	0.619
Control group vs. HCD group	Est. β	-16.67	-39.12	0.26	-0.12	-49.88	-45.39
	t -value	-3.13	-3.59			-2.61	-3.27
	z -value			0.53	-0.78		
	p -value	0.0046	0.0013	0.597	0.43	0.014	0.0027

Note. Similar to Experiment 1, linear mixed-effects models were restructured to consider the potential influence of other item properties and examine the effects of contextual diversity. There were significant effects of character contextual diversity on FFD ($\beta = -17.1$, $t = -2.98$, $p = 0.005$), GD ($\beta = -36.78$, $t = -2.96$, $p = 0.007$), Go-past ($\beta = -56.44$, $t = -2.54$, $p = 0.02$), and TTime ($\beta = -41.72$, $t = -2.67$, $p = 0.01$). We did not observe effects of character frequency ($|\beta|s < 8$, $ts < 0.70$, $ps > 0.50$), word frequency ($|\beta|s < 34$, $ts < 1.30$, $ps > 0.21$), word CD ($|\beta|s < 23$, $ts < 1.20$, $ps > 0.27$), and radical ($|\beta|s < 5$, $ts < 0.8$, $ps > 0.49$)

Setting aside this possibility, why might CD have the effects it does, and what are the implications for models of word recognition and reading? As Jones et al. argue, the relevant dimension approximated by CD is the likelihood that a word will occur in different types of contexts. Indeed, Jones et al. show that a measure that they operationalize as semantic diversity – roughly the likelihood that a word will occur in semantically distinct contexts, as measured by co-occurrence – is correlated with CD but provides a better fit to data which orthogonally varies semantic diversity and CD.

We believe that a promising approach is to embed context-dependent measures in models in which reading times are a function of predictability (e.g., Smith & Levy, 2013) by assuming that predictability effects are at least partially conditionalized on types of contexts. In the absence of knowledge that one is in a *particular* context, the most appropriate prior for a word is the likelihood that it will occur in *any* class of context. In unconstrained contexts, then, CD is a better estimate of the likelihood that a word will be encountered than is token frequency. This approach is related to, but conceptually distinct from, semantic diversity as defined by Jones et al. (2012) in that it predicts that (a) token frequency effects will emerge in constrained contexts and (b) judgments of relatively frequency would be more accurate for words within the same category (e.g., *trumpet* and *trombone* than for words from different categories (e.g., *trumpet* and *daffodil*).

If token frequency effects emerge in constrained contexts when CD is controlled, frequency and CD effects could prove useful in elucidating the grain at which contexts are defined. For example, local within-sentence predictability might be less important than predictability that is tied to a more global context, for example, a text or conversation about a particular sporting event (e.g., baseball or football (soccer in American English) or a type of activity (e.g., cooking).

This, in turn, raises questions about how to operationalize the notion of a context, such that it includes domains which shift the distributions of families of words. One possibility is that a context is theoretically similar to the notion of *topic* that has been explored in models by Steyvers and colleagues (2007). Examining issues related to the likelihood of different word senses appearing in these broad-based contexts, is likely to shed light on the relevant dimension(s) over which readers and listeners generate expectations about forms versus senses. These questions are of particular interest given the emerging evidence for form-based expectation effects in visual word recognition (e.g., Dikker, Rabagliati, Farmer, & Pyllkanen, 2010; Farmer, Christiansen, & Monaghan, 2006; Farmer, Monaghan, Misyak, & Christiansen, 2011; Farmer, Yan, Bicknell, & Tanenhaus, 2015). From this perspective, word-independent, CD-based initial character effects for two-character words occur because form-based predictions about the upcoming character are useful in rapidly estimating surprise. Chinese will be particularly useful for addressing these

issues because of the independent effects of word and character CD in two-character words, and the quasi-systematic relationship between the sense of the first character and the sense of the word.

Acknowledgements This work was supported by grants from the Major State Basic Research Development Program of China (973 Program, 2014CB340502), the Major Program of the Social Science Foundation of China (10&ZD126), the Natural Science Foundation of China (31100814), the China Postdoctoral Science Foundation (2014M560433), and the Postdoctoral Science Foundation of Jiangsu Province (1401061B). MKT was supported by NIH grants HD 27206 and HD 073890. Thanks to Xu Yuhang (Department of Brain and Cognitive Sciences, University of Rochester) for helpful criticism.

References

- Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17, 814–823.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using Eigen and Eigen++. *Journal of Statistical Software*, 65, 1–67.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5, e10729.
- Chen, Q. R., Huang, X., Xu, X. D., Yang, Y. M., & Tanenhaus, M. (2017). The effect of contextual diversity on eye movements in Chinese sentence reading. *Psychonomic Bulletin & Review*.
- Cui, L., Yan, G. L., Bai, X. J., Hyönä, J., Wang, S. P., & Liversedge, S. (2013). Processing of compound-word characters in reading Chinese: An eye-movement-contingent display change study. *The Quarterly Journal of Experimental Psychology*, 66, 527–547.
- R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- Dikker, S., Rabagliati, H., Farmer, T. A., & Pyllkanen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, 21, 629–634.
- Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, 22, 75–87.
- Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 12203–12208.
- Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1318–1325.
- Farmer, T. A., Yan, S., Bicknell, K., & Tanenhaus, M. K. (2015). Form-to-expectation matching effects on first-pass eye movement measures during reading. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 958–976.
- Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in discourse. *Journal of Memory and Language*, 33, 39–68.

- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63, 259–273.
- Institute of Linguistics of Chinese Academy of Social Sciences. (2012). *Modern Chinese Dictionary*. Beijing: The Commercial Press.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66(2), 115–124.
- Kuznetsova, A., Christensen, R. H. B., & Brockhoff, P. B. (2014). LmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version*, 2.
- Lexicon of Common Words in Contemporary Chinese Research Team. (2008). *Lexicon of common words in contemporary Chinese*. Beijing: The Commercial Press.
- Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, 143, 895–913.
- Liu, Y. Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, 39, 192–198.
- Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, 116, 37–44.
- Plummer, P., Perea, M., & Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 275–283.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62, 1457–1506.
- Rayner, K., Li, X., & Pollatsek, A. (2007). Extending the E-Z reader model of eye movement control to Chinese readers. *Cognitive Science*, 31, 1021–1033.
- Shen, W., & Li, X. S. (2012). The uniqueness of word superiority effect in Chinese reading. *Chinese Science Bulletin*, 57, 3414–3420.
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time. *Cognition*, 128, 302–319.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. Hillsdale: Erlbaum.
- Vergara-Martínez, M., Comesaña, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, Affective, & Behavioral Neuroscience*.
- 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(2), 259–268.