



# Event-related brain potentials in lexical processing with Chinese characters show effects of contextual diversity but not word frequency

Jingjing Zhang<sup>1</sup> · Yixiao Zhou<sup>1</sup> · Guoxia Zhao<sup>1</sup> · Xin Wang<sup>2</sup> · Qingrong Chen<sup>1,3</sup> · Michael K. Tanenhaus<sup>4</sup>

Accepted: 18 May 2024  
© The Psychonomic Society, Inc. 2024

## Abstract

The diversity of contexts in which a word occurs, operationalized as CD, is strongly correlated with response times in visual word recognition, with higher CD words being recognized faster. CD and token word frequency (WF) are highly correlated but in behavioral studies when other variables that affect word visual recognition are controlled for, the WF effect is eliminated when contextual diversity (CD) is controlled. In contrast, the only event-related potential (ERP) study to examine CD and WF Vergara-Martínez et al., *Cognitive, Affective, & Behavioral Neuroscience*, 17, 461–474, (2017) found effects of both WF and CD with different distributions in the 225- to 325-ms time window. We conducted an ERP study with Chinese characters to explore the neurocognitive dynamics of WF and CD. We compared three groups of characters: (1) characters high in frequency and low in CD; (2) characters low in frequency and low in CD; and (3) characters high in frequency and high in CD. Behavioral data showed significant effects of CD but not WF. Character CD, but not character frequency, modulated the late positive component (LPC): high-CD characters elicited a larger LPC, widely distributed, with largest amplitude at the posterior sites compared to low-CD characters in the 400-to 600-ms time window, consistent with earlier ERP studies of WF in Chinese, and with the hypothesis that CD affects semantic and context-based processes. No WF effect on any ERP components was observed when CD was controlled. The results are consistent with behavioral results showing CD but not WF effects, and in particular with a “context constructionist” framework.

**Keywords** Contextual diversity · Frequency · ERP · Chinese character

## Introduction

Word frequency, the number of times a word occurs regardless of context, has long played a central role in developing and evaluating models of visual word recognition and reading. However, a pioneering study by Adelman et al. (2006) found that much of the variance previously attributed to WF

is better explained by the diversity of contexts in which a word occurs (for review, see Caldwell-Harris, 2021). Adelman et al. (2006) operationalized a measure, contextual diversity (CD) – the proportion of texts in a corpus in which a word occurs. When controlling for other dimensions that affect lexical processing, CD but not word frequency (WF) affected naming and lexical decision times (Adelman et al., 2006; Adelman & Brown, 2008; Jones et al., 2017).

Adelman et al.’s work was motivated by memory research, where repeated exposure has minimal effects when an item is repeated in the same context (Verkoeijen et al., 2004). If lexical memory follows the same principles, words that occur in more diverse contexts will be better learned and retrieved (see Jones et al., 2012, for a related, learning-based account).

Recently, we proposed a “context constructionist” framework that assumes that (1) lexical representations store fine-grained, contextualized statistical information about word distributions; and (2) these representations are used to actively construct and update a context model that informs

✉ Qingrong Chen  
jscqr80@sina.com

<sup>1</sup> School of Psychology, Nanjing Normal University, Nanjing 210097, China

<sup>2</sup> Human Communication, Development, and Information Sciences, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China

<sup>3</sup> Jiangsu Collaborative Innovation Center for Language Ability, School of Linguistic Sciences and Arts, Jiangsu Normal University, Xuzhou, China

<sup>4</sup> Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

expectations about expected words in that context (Chen et al., in preparation; Yan et al., 2018). Thus, lexical retrieval is optimized to reflect “need probability” (Anderson & Schooler, 1991) –the probability that a word will be encountered in the upcoming text or discourse.<sup>1</sup>

With contextualized word representations, people can form expectations about what words are likely to be encountered in the current task/context. In a specific context, words that are more frequent within that context will be more expected. CD and WF are highly correlated; however, WF is not directly incorporated into lexical representations, and thus is not accessible or easily computed (it would have to be computed by summing the frequency of a word in the range of contexts in which it occurs, weighted by the probability of these contexts). However, the number of distinct contexts in which a word occurs would be more accessible: as CD increases, words are likely to have a larger and more varied set of semantic associations (Adelman et al. 2006; Hoffman et al, 2011), and therefore degree of semantic activation would be a good proxy for need probability.

Because WF and CD are proxies for the same underlying factor, when other variables that affect lexical processing are controlled, it would be surprising to find different effects of both WF and CD. Indeed, in an important study, Plummer et al. (2014) found CD but not WF effects in an eye-tracking study across multiple fixation measures. Because CD and WF cannot be manipulated factorially (HCD words are typically HWF), they introduced a *three-condition design* with words in a contrast/control condition, a *HCD* condition, matched for WF with the contrast condition, but with higher CD, and a *LWF* condition, matched for CD with the contrast condition with lower WF. They found CD (HCD vs. the contrast condition) but not WF (LWF vs. the contrast condition) effects (e.g., HCD words had shorter FFD). The same CD-dominant pattern for three-condition designs has been found in eye-tracking studies for words and characters in sentences in Chinese (Chen, Huang et al., 2017a; Chen, Zhao et al., 2017b), and for lexical decisions with young readers in Portuguese (Perea et al., 2013), and character decision in Chinese (Huang et al., 2021).

Crucially, our account makes novel predictions about CD and WF when contextual constraint increases. First, CD effects should decrease as contextual constraint increases. Second, in strongly constraining contexts with three-condition designs, WF but not CD should affect reading times. We have confirmed these predictions in a three-condition eye-tracking experiment in Chinese and in an analysis of a

corpus of eye-tracking data for natural texts in English (Chen et al., 2025; Yan et al., 2018).<sup>2</sup>

In contrast to behavioral studies with words in isolation, which consistently find effects of CD but not WF, a recent ERP study by Vergara-Martinez et al. (2017) found dissociable effects: Both CD and WF evoked negativities in the 225- to 325-ms time window. However, high CD words elicited larger negativity than low CD words in the anterior region, whereas low-frequency words evoked larger negativity than high-frequency words in the anterior-central region.

The ERP study by Vergara-Martinez et al. is important in clarifying the locus of the CD effect, showing that CD effects have a semantic origin. However, there are two aspects of the results that are noteworthy. First, while the CD but not WF affected response times, the 13-ms CD effects are smaller than observed in previous behavioral studies (e.g., 53 ms in Perea et al., 2013; 65 ms in Plummer et al., 2014).<sup>3</sup> This raises questions about the strength of the CD manipulation. Second, Vergara-Martinez et al. argue that because facilitatory effects are found for both increased CD and increased WF, different effects might be masked in behavioral measures but might be dissociable with a measure like ERP. While this is true in principle, it does not explain why behavioral effects of WF are not found when

<sup>2</sup> An example of a broad (weakly) constraining context and a narrow (strongly) constraining context from Chen, Yan, Mollica, and Tanenhaus (in preparation). In broad contexts CD but not WF affect fixation durations, whereas in narrow contexts, there are WF but not CD effects. The context constructive model predicts this pattern because in a constrained context, need probability is determined by the frequency of the word in that context. Data and materials for this study are available in the Science Data Bank (ScienceDB) data repository: <https://www.scidb.cn/s/BJfmM3>.

Target sentence frame	Broad context	Narrow context
远处的影星引起了大家的注意。 The <i>star</i> in the distance drew everyone's attention.	在本次海选现场的入口处,主持人下车后向粉丝们招手致意。突然,一阵阵尖叫声从人群的边缘传来。 At the entrance to the audition, the host got off the bus and waved to fans. Suddenly, screams came from the edge of the crowd.	据说这部贺岁片的主角都来参加首映礼,在座的粉丝们十分激动。突然,一阵阵尖叫声从人群的边缘传来。 It is said that the main characters of the New Year film came to the premiere, and the fans present were very excited. Suddenly, screams came from the edge of the crowd.

<sup>3</sup> Perea et al.'s (2013) study was conducted with children, in which participants were asked to make a go/no-go lexical decision task on Portuguese words. Plummer et al. (2014) is a study conducted in an adult population, in which participants were asked to complete a "yes/no" lexical decision task on English words.

<sup>1</sup> Yan et al. formalized the context constructivist account as:

$$P(\mathbf{w}) = \sum^C P(\mathbf{w}|\mathbf{c}) * P(\mathbf{c})$$

where  $P(\mathbf{w}|\mathbf{c})$  is the need probability of a word in specific contexts).

CD is controlled. Moreover, it's not clear why larger anterior negativity for higher CD words would map onto faster response times, whereas larger anterior-central negativity for lower WF words would not map onto a response time difference. These observations highlight the importance of replicating the results, especially if the replication showed stronger behavioral effects, which would ensure that the manipulation of CD was robust.

We examined CD and WF effects for characters in Chinese. Characters are the basic orthographic/morphemic unit in Chinese, which minimizes structural complexities associated with morphology, and to some extent orthographic consistency and spelling-to-sound mapping are minimized (Adelman et al. noted that WF is more strongly correlated with word form structural factors than CD). While behavioral and neural studies present different patterns in lexical processing in these two language systems (e.g., Cao et al., 2013; Kim et al., 2016; Zhou & Marslen-Wilson, 2000), behavioral studies using characters in Chinese find the same pattern of CD effects as is found in English and in Portuguese.

Separate neural patterns for WF and CD in a language with a very different orthography would provide compelling support for Vergara-Martinez et al.'s conclusions. Moreover, it would provide strong evidence against any approach, such as ours, in which contextual variability measures and WF are proxies for the same underlying dimension. On the other hand, if we do not find different effects of WF and CD, the results would be consistent with that hypothesis, and importantly, it would pave the way for contextual manipulations that could provide a strong test of the unified hypothesis, a point we return to in the *General discussion*.

We used stimuli drawn from a corpus of Chinese characters used in films (Cai & Brysbaert, 2010) and manipulated character frequency (CF) and CD simultaneously. As in Vergara-Martinez et al., we used a three-condition design.

We predicted that compared with control condition with the same CF but lower CD, character decision times would be faster for the HCD characters, with no effect of CF. As we noted earlier, degree of semantic activation would be a good proxy for need probability. Higher CD characters are likely to be semantically richer than lower CD characters (Adelman et al., 2006; Hoffman et al., 2013; Vergara-Martinez et al., 2017). Therefore, we predicted that HCD characters would induce larger N400 or late positive component (LPC) than characters in the control condition. LPC is a positive component occurring at approximately 500 ms after stimulus onset, with the largest scalp distribution over the posterior region. Although it was initially discussed in relation to syntactic and structural processing, more recent findings demonstrate that LPC is also sensitive to semantic context (for review, see Aurnhammer et al., 2023). Semantic richness effects, which often result in N400 effects for words in

alphabetic languages (e.g., Müller et al., 2010; Rabovsky et al., 2012; Vergara-Martinez et al., 2017), are also realized as effects on the LPC component in Korean and Chinese (Ding et al., 2017; Kwon et al., 2012). In these studies, larger N400 or LPC amplitude is often reported for words with many semantic associates or features than for those with few semantic associates or features. If there are effects of both CD and CF, we should also see ERP differences in the LCF condition compared to the control condition, even if (as expected) there are no behavioral effects between these conditions.

## Methods

### Participants

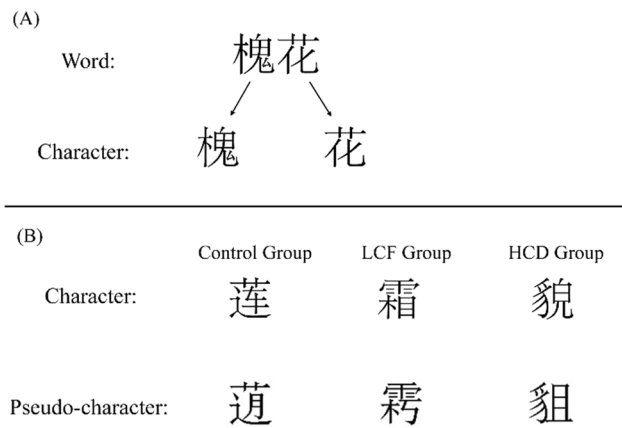
Twenty-nine students participated in the study (15 females, 14 males, age range 21–26 years, mean age 23.72 years). Participants were right-handed, native Mandarin Chinese speakers with normal or corrected-to-normal vision, and no history of neurological or language impairments. Participants were paid for their participation and signed informed consent prior to the experiment.

### Materials

Characters were selected from the SUBTLEX-CH-CHR database (Cai & Brysbaert, 2010). The database provides CF based on the number of occurrences in 33 million words, and CD based on the proportion of films in which a character appears in a 6,243 film-corpus. The CF and CD were both transformed to a log scale. We chose this corpus because frequencies based on this database explain more of the variance in word and character reading than frequencies based on written texts (Cai & Brysbaert, 2010).

We selected 150 single monomorphemic characters from the database, with 50 characters for each condition (Fig. 1B). Characters in the HCD condition have similar CF to the control group ( $t(98) = -1.662, p = 0.10$ ), but they have higher CD ( $t(98) = -16.433, p < 0.001$ ). Characters in the LCF condition have lower CF than the control group ( $t(98) = 25.485, p < 0.001$ ), but they have similar CD ( $t(98) = -1.645, p = 0.103$ ).

Chinese characters are composed of a series of strokes, and those strokes often combined to form sub-character units called “radicals” (Taft et al., 1999; Yan et al., 2012). Different characters may vary in the number of strokes and number of radicals, both of which affect the recognition of characters (Ding et al., 2004; Feldman & Siok, 1997, 1999; Taft et al., 1999; Taft & Zhu, 1997). Therefore, across conditions, characters were matched for number of strokes ( $ts < -0.099, ps > 0.529$ ), radicals ( $ts < 1.003, ps > 0.171$ ), orthographic



**Fig. 1** **A** An example of Chinese word and character. The word (槐花, /huai2 hua1/, the sophora flower) consist of two characters: 槐 (/huai2/, sophora), 花 (/hua1/, flower). Characters are pronounceable and convey meaning. **B** Examples of characters (upper) and pseudo-characters (lower) in different conditions

neighborhood size ( $ts < 1.126$ ,  $ps > 0.263$ ), and semantic polysemy ( $ts < -0.880$ ,  $ps > 0.163$ ). We also controlled for phonological consistency (Hsu et al., 2009; Lee et al., 2005, 2015) and regularity (Cai et al., 2012). Phonological consistency ( $ts < 1.548$ ,  $ps > 0.127$ ) and regularity ( $\chi^2 s < 0.31$ ,  $ps > 0.58$ ) were matched across conditions for phonograms. Regularity and consistency are phonological properties of phonograms (Hsu et al., 2009; Lee et al., 2005; Yum & Law,

2019). Regularity is defined as whether the pronunciation of a phonogram is identical with its phonetic radical, regardless of tone. Consistency is defined as the degree to which a phonetic radical is a reliable cue to the sound of the phonogram containing it. This was calculated by dividing the number of orthographic neighbors with the same pronunciation by the total number of orthographic neighbors.

Twenty-six participants rated concreteness, familiarity, imageability, age of acquisition, valence, arousal, and dominance of each character on 7-point scales. These variables did not differ significantly across conditions ( $ts < 1.472$ ,  $ps > 0.147$ ). The detailed values for each condition are presented in Table 1.

One hundred and fifty pseudo-characters were generated by randomly combining radicals from the original characters: all followed standard orthographic patterns. Using a 7-point scale, 20 students who didn't participate in the EEG experiment rated whether the pseudo-characters looked like real characters. There was no significant difference among conditions ( $ts < 1.36$ ,  $ps > 0.18$ ).

## Procedure

Participants were seated in a sound-attenuating, electrically shielded chamber, approximately 65 cm distant from a computer screen. Following previous studies (e.g., Huang et al., 2021; Zhao et al., 2010), each trial began with a fixation

**Table 1** Characteristics of the target characters in each group

Characteristic	Control group	LCF group	HCD group
Mean character frequency (log)	3.09 (0.02)	2.55 (0.01)	3.14 (0.03)
Character frequency range (log)	2.92–3.50	2.38–2.60	2.91–3.82
Mean contextual diversity (log)	2.49 (0.02)	2.45 (0.01)	3.02 (0.03)
Contextual diversity range (log)	2.19–2.67	2.34–2.52	2.80–3.58
Number of strokes	9.84 (0.49)	9.90 (0.35)	10.24 (0.40)
Number of radicals	2.72 (0.13)	2.56 (0.10)	2.98 (0.14)
Semantic polysemy	2.86 (0.24)	3.16 (0.24)	3.32 (0.23)
Orthographic neighborhood size	407.96 (53.04)	328.96 (45.95)	354.82 (44.70)
Regularity	0.27 (0.05)	0.18 (0.03)	0.24 (0.05)
Consistency	0.33 (0.06)	0.34 (0.05)	0.36 (0.06)
Concreteness	3.35 (0.15)	3.28 (0.16)	3.05 (0.20)
Familiarity	6.75 (0.07)	6.76 (0.07)	6.83 (0.07)
Imageability	3.93 (0.15)	4.04 (0.17)	4.12 (0.17)
Age of acquisition	4.80 (0.16)	4.87 (0.17)	4.45 (0.17)
Valence	3.78 (0.16)	3.66 (0.17)	3.57 (0.16)
Arousal	2.70 (0.20)	2.88 (0.22)	3.09 (0.19)
Dominance	3.44 (0.16)	3.51 (0.16)	3.52 (0.13)

Standard errors are provided in parentheses. Semantic polysemy was calculated by the number of meanings that a character has based on the Modern Chinese Dictionary (2012)

cross in the center of the screen with a random duration ( $M = 1,250$  ms, range = 1,000–1,500 ms). A character was then presented for 200 ms, followed by a blank screen for 2,500 ms. There were six blocks, with each block containing 50 trials. Block order was counterbalanced across participants. Stimuli from the same condition did not appear in more than three consecutive trials and were displayed in a pseudo-randomized order.

Participants performed a character decision task, pressing the “D” or “K” key as accurately and quickly as possible. Assignment of “character” and “pseudo-character” to keys was balanced across participants. The E-Prime software package (Psychology Software Tools, Pittsburgh, PA, USA) was adopted for stimulus presentation and response collection. Response time (RT) was measured from stimulus onset to the participants’ response. The experiment began with a practice session of 20 trials to familiarize participants with the procedure. The entire experiment lasted about 1 h.

### EEG recordings

EEG was continuously recorded by a SynAmp amplifier from 64 Ag/AgCl electrodes, mounted on an elastic cap, located in the Standard International 10–20 System. EEG was referenced online to the left mastoid, and then re-referenced offline to the algebraic average of the left and right mastoids. Vertical electro-oculogram (EOG) was recorded from electrodes located above and below the orbital regions of the left eye. Horizontal EOG was recorded from electrodes located at the outer canthus of each eye. EEG data were digitized at a rate of 1,000 Hz, with a 400-Hz high cut-off filter and a 0.05-Hz low cut-off filter. Electrode impedances were kept below 5 k $\Omega$  throughout the experiment.

### Behavioral data analysis

Planned comparisons used linear mixed-effects models for character decision times and mixed logit models for accuracy using the lme4 package (Bates et al., 2015) in R (R Development Core Team, 2014). The model included fixed effects (conditions) and the maximal random effects structure that would converge as justified by the data with by-participants and by-items random intercepts and slopes (Barr et al., 2013; Jaeger, 2008; Matuschek et al., 2017).<sup>4</sup> The lmerTest package was implemented for significance testing. For linear mixed effects models, we estimated  $p$  values

using the Satterthwaite approximation for degrees of freedom (Kuznetsova et al., 2017).

### EEG data analysis

EEG data were analyzed using MATLAB scripts based on EEGLAB toolbox (Delorme & Makeig, 2004). A digital bandpass filter between 0.1 and 30 Hz was conducted offline. Ocular artifacts were removed via independent component analysis, and other types of EEG artifacts were rejected automatically with criterion of  $\pm 75$   $\mu$ V and manually through visual inspection. Data were segmented from 200 ms before to 800 ms after the onset of the targets, with baseline correction from 200 ms to 0 ms preceding target onset. Incorrectly answered trials were excluded from further analysis. On average, 7.3% of trials were rejected, and  $46.28 \pm 2.89$ ,  $47.07 \pm 2.80$  and  $45.72 \pm 4.71$  trials were included in the control, HCD and LCF conditions, respectively, with no significant difference in number of trials remaining across conditions ( $t_s < 1.56$ ,  $p_s > 0.13$ ).

Based on visual inspection and previous research (e.g., Lartseva et al., 2014), statistical analyses were performed on the mean amplitude between 400 and 600 ms. The midline and lateral electrodes were computed separately. In the midline analysis, there were two factors including character type (LCF/HCD group and Control group) and region (anterior (Fz, FCz), central (Cz, CPz), and posterior (Pz, POz)). In the lateral analysis, there were three factors including character type, Hemisphere (left and right), and Region (anterior, central, and posterior). Lateral electrodes were organized into six regions of interest (ROIs): left anterior (F1, F3, F5, FC1, FC3, FC5), left central (C1, C3, C5, CP1, CP3, CP5), left posterior (P1, P3, P5, PO3, PO5, PO7), right anterior (F2, F4, F6, FC2, FC4, FC6), right central (C2, C4, C6, CP2, CP4, CP6), and right posterior (P2, P4, P6, PO4, PO6, PO8).

We used linear mixed-effects models to analyze the item-based amplitude of the ERP in the time window of 400 to 600 ms. The model included fixed effects (e.g., condition, region, hemisphere) and the maximal random effects structure that would converge, as justified by the data with by-participants and by-items random intercepts and slopes (Barr et al., 2013; Matuschek et al., 2017).<sup>5</sup> Post hoc pairwise comparisons were conducted using the emmeans package with Tukey corrections (Lenth et al., 2018).

<sup>4</sup> RT analysis: `lmer (RT ~ condition + (1 | item) + (1 + condition | subject), control = lmerControl (optCtrl = list (maxfun = 1000)), data)`; ACC analysis: `glmer (ACC ~ condition + (1 | item) + (1 + condition | subject), family = binomial, control = glmerControl (optCtrl = list (maxfun = 1000)), data)`.

<sup>5</sup> Midline analysis: `lmer (avg ~ condition * location + (1 | item) + (1 + condition | subject), control = lmerControl (optCtrl = list (maxfun = 1000)), data)`; Lateral analysis: `lmer (avg ~ condition * hemisphere * region + (1 | item) + (1 + condition | subject), control = lmerControl (optCtrl = list (maxfun=1000)), data)`.

**Table 2** Mean character decision times and average accuracy rates for characters in lexical decision task

Measure	Control group	LCF group	HCD group
Accuracy rates (%)	95.39 (0.64)	94.69 (0.95)	98.00 (0.42)
Response time (ms)	730.61 (5.34)	733.20 (5.40)	686.39 (4.09)

Standard errors are in parentheses

## Results

### Behavioral results

Mean RTs and accuracy rates are presented in Table 2. The average accuracy rates were 95.93% ( $SE = 0.64\%$ ) in the control group, 94.69% ( $SE = 0.95\%$ ) in the LCF condition, and 98.00% ( $SE = 0.42\%$ ) in the HCD condition. Mixed logit models showed that there were no significant effects of character frequency and CD on error rates ( $|β|s < 0.92$ ,  $|z|s < 1.90$ ,  $ps > 0.05$ ).

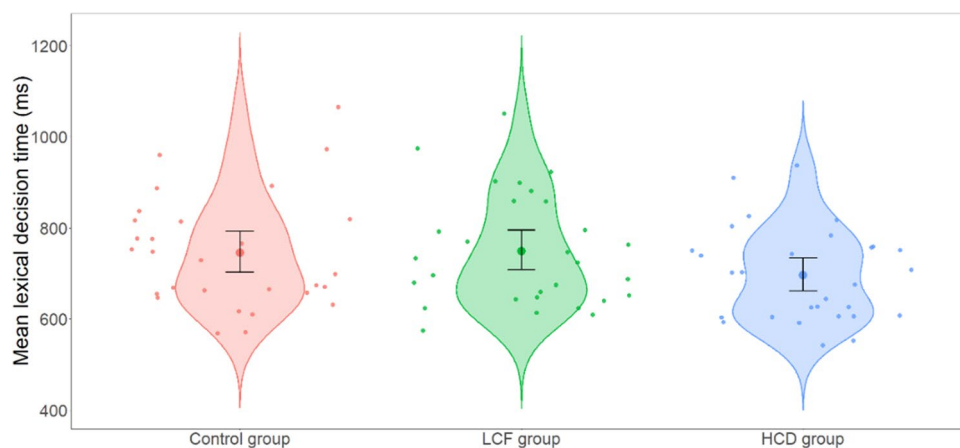
Mean character decision times were 730.61 ms ( $SE = 5.34$  ms) in the control group, 733.20 ms ( $SE = 5.40$  ms) in the LCF condition, and 686.39 ms ( $SE = 4.09$  ms) in the HCD condition (see Fig. 2). As predicted, the CD effect was significant (control group vs. HCD group),  $β = -48.00$ ,  $SE = 12.38$ ,  $t = -3.88$ ,  $p < 0.001$ , whereas the WF effect (control group vs LCF group) was not,  $β = 4.73$ ,  $SE = 13.85$ ,  $t = 0.34$ ,  $p = 0.73$ .

### ERP results

The grand average ERP, time-locked to the onsets of critical characters, is displayed in Fig. 3. Between 400 and

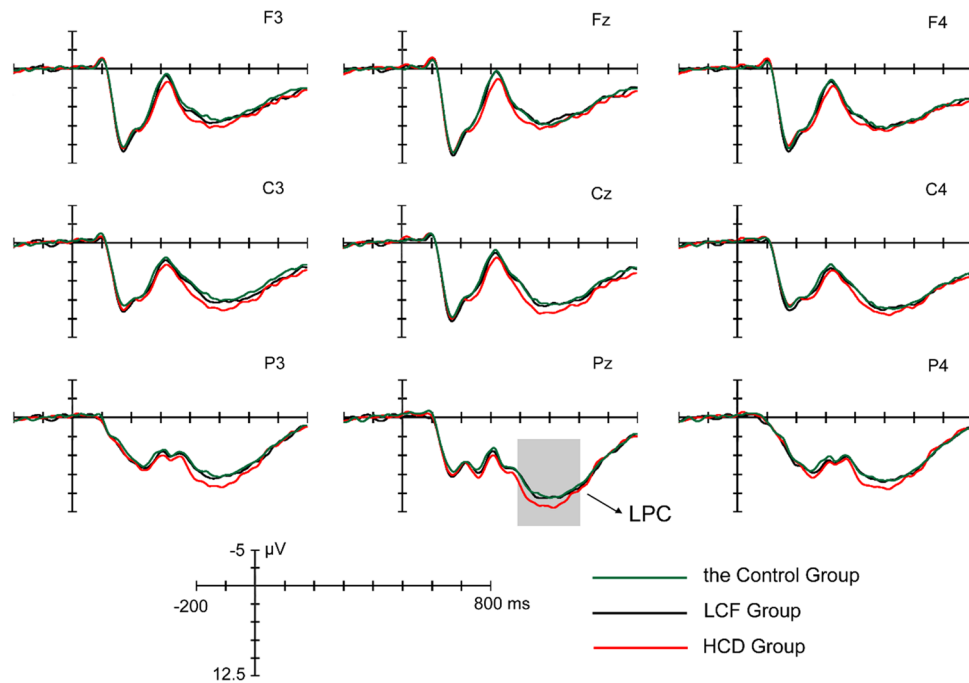
600 ms, there was a main effect of CD in both the midline electrodes ( $F = 7.45$ ,  $p = 0.008$ ) and the lateral electrodes ( $F = 8.44$ ,  $p = 0.005$ ). High-CD characters evoked larger late positive component (LPC) than the control condition (see Fig. 4). The  $CD \times$  region interaction was significant (see Fig. 5),  $F = 3.20$ ,  $p = 0.04$ . Simple effect analyses showed that the effect of CD was largest at the posterior sites ( $β = 0.75$ ,  $SE = 0.22$ ,  $z = 3.48$ ,  $p < 0.001$ ), followed by the central region ( $β = 0.59$ ,  $SE = 0.22$ ,  $z = 2.73$ ,  $p = 0.006$ ), and did not reach significance at the anterior region ( $β = 0.26$ ,  $SE = 0.22$ ,  $z = 1.21$ ,  $p = 0.23$ ). The  $CD \times$  hemisphere interaction was marginally significant,  $F = 2.90$ ,  $p = 0.09$ . We further performed a Bayes factor model comparison using R package “BayesFactor” (Morey & Rouder, 2018). The Bayes factor reflects the ratio of the likelihood probability of two competing models. It has advantages over other model comparison methods such as likelihood ratio tests (Baele et al., 2013). Adding the interaction between CD and hemisphere into the model only improved it by a factor of 0.094, showing no evidence for the potential interaction effect (Jeffreys, 1998).

As shown in Figs. 3 and 4, no main effect of CF was observed in the midline analysis ( $F = 0.67$ ,  $p = 0.42$ ) or in the lateral analysis ( $F = 1.73$ ,  $p = 0.19$ ). The interaction between the CF and hemisphere was marginally significant ( $F = 2.91$ ,  $p = 0.09$ ); however, the Bayes factor shows that adding the interaction between CF and hemisphere into the model only improved it by a factor of 0.08, which is extremely weak evidence for the model with character frequency and hemisphere added. No other interaction with CF was observed,  $F_s < 0.59$ ,  $ps > 0.71$ . Supplemental



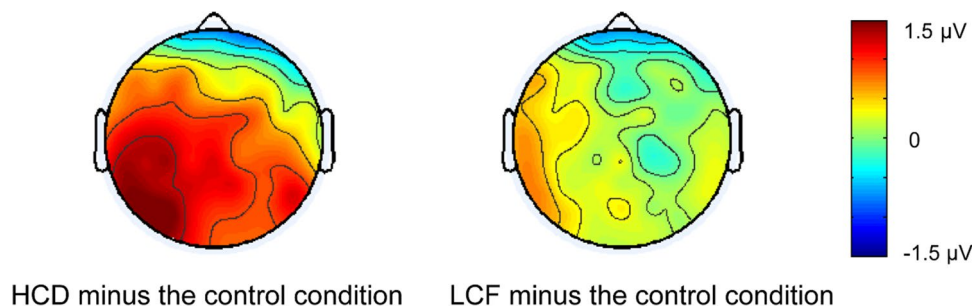
**Fig. 2** Mean response times for characters in lexical decision task. The control group comprises characters with higher character frequency (CF) and lower contextual diversity (CD), the LCF group comprises characters with a similar CD as the control group but lower CF, the HCD group comprises characters with a similar CF as the

control group but higher CD. Bigger dots represent mean amplitude for each condition, and other small dots represent individual mean amplitude for each condition. Error bars represent bootstrapped 95% confidence interval of the mean



**Fig. 3** Grand average event-related potential (ERP) in response to the target characters from nine representative electrodes over the -200- to 800-ms time window. The control group comprises characters with higher character frequency (CF) and lower contextual diversity (CD), the LCF Group comprises characters with a similar CD to the con-

trol group but lower CF, the HCD Group comprises characters with a similar CF to the control group but higher CD. The onset of the critical word is aligned to the zero in the timeline. Analysis windows are shown by the gray-shaded areas



**Fig. 4** Topographical distributions of the contextual diversity (CD) effect and the character frequency (CF) effect in the 400- to 600-ms time window. Mean amplitude differences were calculated across the event-related potential responses to three conditions

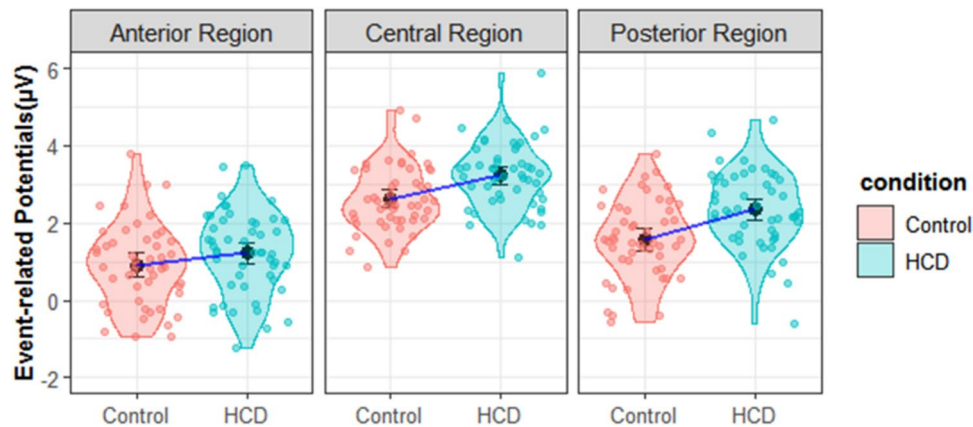
regression analysis also observed significant effect of CD but not CF (see Fig. 6).<sup>6</sup>

<sup>6</sup> We conducted a regression analysis in which the mean amplitude of the ERP in the 400- to 600-ms time window was the dependent variable. Predictors, which were simultaneously entered into the regression, were:  $\log_{10}$  transformed CD and character frequency (both from SUBTLEX-CH-CHR database), number of strokes, number of radicals, orthographic neighborhood size, semantic polysemy, regularity, consistency, concreteness, familiarity, imageability, age of acquisition, valence, arousal, and dominance. The regression analysis found a significant facilitative effect of CD in both the midline electrodes ( $t = 5.19$ ,  $p < 0.001$ ,  $\beta = 1.68$ ) and the lateral electrodes ( $t = 4.65$ ,  $p < 0.001$ ,  $\beta = 1.30$ ), but not of CF ( $|t|s < 1.37$ ,  $ps > 0.17$ ,  $|\beta|s < 0.40$ ).

## Discussion

We manipulated CF and CD for Chinese characters using a character decision task while measuring ERPs. With CF controlled, character decision times were faster for higher CD characters compared to a control condition, whereas there were no effects of CF, with the magnitude of the CD effects consistent with previous behavioral studies.

ERPs were sensitive to CD but not frequency. The LPC, a late positive component that likely reflects degree of semantic activation (Chen et al., 2016; Juottonen et al., 1996; Zou



**Fig. 5** Mean amplitude of the event-related potentials in the 400- to 600-ms window elicited by control group (high frequency and low CD characters) and HCD group (high frequency and CD characters) in each region. Black dots represent mean amplitude for each condi-

tion and error bars represent bootstrapped 95% confidence interval of the mean. Small dots represent individual mean amplitude for each condition

et al., 2019), and which is sensitive to linguistic context (Aurnhammer et al., 2023), was larger for higher CD characters compared to lower CD, matched-frequency controls. Importantly, the CD effect obtained in the present study cannot be explained in terms of other semantic variables (e.g., concreteness, imageability) or emotional variables (e.g., valence, arousal), as the experimental characters were matched in these factors (see Table 1). Compared to low CD characters, contextual information is richer and more available for high CD characters, resulting in a larger LPC amplitude. Notably, in previous ERP studies using Chinese words or characters, which manipulated word and character frequency but not CD, frequency effects were also reflected in LPC (e.g., Guo et al., 2004; Ye et al., 2019; Yum & Law, 2019; Zhang et al., 2006). Moreover, the direction and the central-posterior distribution of the CD effects resembles the results obtained in other ERP studies that manipulated factors related to context (e.g., Kwon et al., 2012).

There are similarities and differences between our findings and those of Vergara-Martinez et al. (2017). The most important similarity is that the LPC locus of the CD effects support Vergara-Martinez et al.'s conclusion that the ERP effects of CD are “the result of larger semantic networks that become temporally active for words that appear in many contexts” (Vergara-Martinez et al., 2017, p. 467).

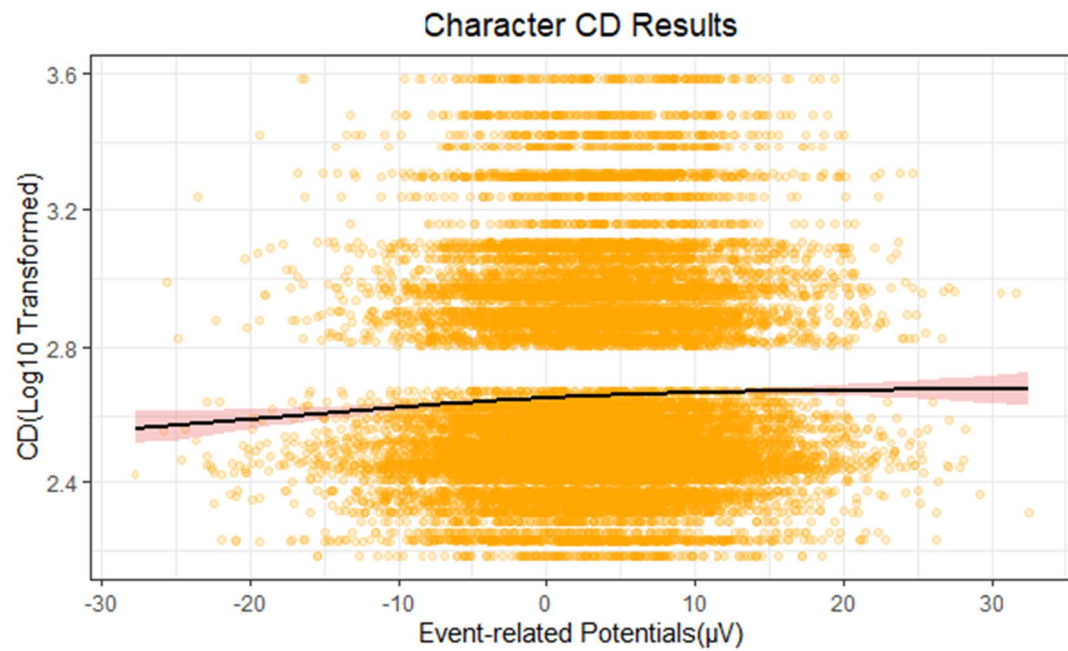
There are two notable differences. First, Vergara-Martinez et al. found CD effects on N400, whereas in our study CD affected LPC, a later component. This difference is not surprising. While frequency effects on N400 have been observed in Chinese, frequency consistently affects LPC, which follows N400 and is sensitive to semantic and contextual variables. In behavioral studies where both CD and WF were manipulated, character and lexical decision times (the current study and Huang et al., 2021) and reading times (Chen,

Huang et al., 2017a; Chen, Zhao et al., 2017b) to Chinese words and characters were slower than those to English (Plummer et al., 2014), Spanish (Vergara-Martinez et al., 2017), and Portuguese (Perea et al., 2013). The different time-course of the CD effects likely reflects slower access of semantic/lexical information in Chinese compared to alphabetic languages, with the time course of the LPC consistent with character-decision times (for review, see Li et al., 2022).

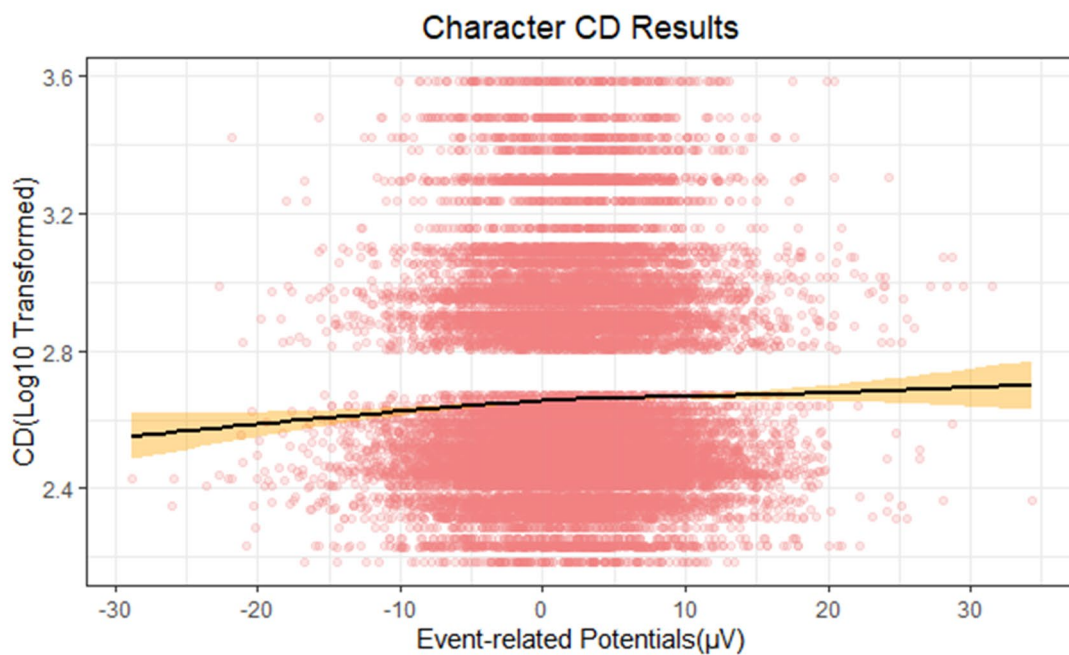
The second, and most important, difference is that Vergara-Martinez et al. found ERP effects of both CD and WF with CD and WF effects differing in their direction and distribution, whereas we found effects of character CD but not frequency, which is consistent with results using behavioral measures. Further research will be needed to determine whether this difference can be attributed to properties of alphabetic compared to character-based orthographies, or to some other aspect of the materials, for example, structural characteristic of word forms that are correlated with WF but not CD (Adelman et al. 2006; Vergara-Martinez et al., 2017). One promising approach would be to use three-condition designs in which context manipulations result in either CD or WF effects, depending upon the strength of the contextual constraint (see note 2 for an example).

The results are consistent with our context constructivist account in which both CD and WF effects reflect need probability (e.g., predictability) of a word. On this account, lexical representations store only context-contingent frequencies. Thus, token frequency is not easily accessible/computable. However, the range of contexts in which a word will occur (which is correlated with semantic richness) is accessible and thus a good proxy for need probability for words in isolation or weakly constraining contexts. Because WF and CD are both proxies for need probability we do not predict dissociable effects of these





(a) Midline analysis



(b) Lateral analysis

**Fig. 6** The amplitude of event-related potentials in the time window of 400- to 600-ms as a function of the contextual diversity values ( $\text{Log}_{10}$  transformed). Red dots (Fig. 6a) and yellow dots (Fig. 6b) rep-

resent individual amplitude for each item in midline and lateral analyses respectively. Colorful shaded regions represent 95% confidence intervals on the slopes

two variables in three-condition designs in which other variables that affect lexical access, many of which are correlated with WF, are factored out. In ERP studies, depending on the time course of semantic effects, CD should be

reflected in components sensitive to richness of context, such as N400 or LPC.

The results are also consistent with two proposals that do not incorporate need probability. The first is

the “context availability model” (Holcomb et al., 1999; Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983), which is often used to explain concreteness effect. This model argues that comprehension is heavily reliant on contextual information provided by either the preceding context or the comprehender’s mental knowledge. In the absence of context, lexical decisions are shorter for high-CD characters because of the increased availability of related contextual information, which also results in a larger LPC amplitude. However, the context constructivist model differs from the context availability model in making specific claims about how context is incorporated into lexical representations and in predicting word frequency effects in constrained contexts.

Our approach differs from Adelman et al. (2006) and Jones et al. (2017) in that it incorporates context into lexical representations and assumes that need probability underlies both CD and WF effects. Our approach makes novel predictions about how WF and CD effects will be modulated by contextual constraint, which can be manipulated in three-condition designs. We suggest that neural-imaging studies adopting this approach would be a fruitful avenue for understanding the neural basis of CD and WF effects, including whether they are dissociable.

**Funding** This work was supported by the Major Project of the National Social Science Foundation of China [grant number 21&ZD288] to Qingrong Chen.

**Data availability** The data and materials are available at <https://www.scidb.cn/anonymous/ajZqcVFy>.

**Code availability** The analysis code is available at <https://www.scidb.cn/anonymous/ajZqcVFy>.

## Declarations

**Conflicts of interest** The authors have no competing interests to declare that are relevant to the content of this article.

**Ethics approval** This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the School of Psychology, Nanjing Normal University.

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

**Consent for publication** The authors affirm that participants signed informed consent regarding publishing their data.

## References

Adelman, J. S., & Brown, G. D. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review*, *115*(1), 214–229.

- 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*(9), 814–823.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408.
- Aurnhammer, C., Delogu, F., Brouwer, H., & Crocker, M. W. (2023). The P600 as a continuous index of integration effort. *Psychophysiology*, *60*(9), e14302.
- Baele, G., Lemey, P., & Vansteelandt, S. (2013). Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinformatics*, *14*(1), 85–103.
- 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*(3), 255–278.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, *5*(6), e10729.
- Cai, H. D., Qi, X. L., Chen, Q. R., & Zhong, Y. (2012). Effects of phonetic radical position on the regularity effect for naming pictophonetic characters. *Acta Psychologica Sinica*, *44*(7), 868–881.
- Caldwell Harris, C. L. (2021). Frequency effects in reading are powerful—But is contextual diversity the more important variable? *Language and Linguistics Compass*, *15*(12), e12444.
- Cao, F., Vu, M., Lung Chan, D. H., Lawrence, J. M., Harris, L. N., Guan, Q., Xu, Y., & Perfetti, C. A. (2013). Writing affects the brain network of reading in Chinese: A functional magnetic resonance imaging study. *Human Brain Mapping*, *34*(7), 1670–1684.
- Chen, W., Chao, P., Chang, Y., Hsu, C., & Lee, C. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Language*, *157*, 51–62.
- Chen, Q., Huang, X., Bai, L., Xu, X., Yang, Y., & Tanenhaus, M. K. (2017a). The effect of contextual diversity on eye movements in Chinese sentence reading. *Psychonomic Bulletin and Review*, *24*(2), 510–518.
- Chen, Q., Zhao, G., Huang, X., Yang, Y., & Tanenhaus, M. K. (2017b). The effect of character contextual diversity on eye movements in Chinese sentence reading. *Psychonomic Bulletin and Review*, *24*(6), 1971–1979.
- Chen, Q., Yan, S. R., Mollica, F., & Tanenhaus, M. K. (in preparation, 2025). A context constructivist account of contextual diversity and word frequency. *The Psychology of Learning and Motivation*, *83*.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Ding, G., Peng, D., & Taft, M. (2004). The nature of the mental representation of radicals in Chinese: A priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 530–539.
- Ding, J., Liu, W., & Yang, Y. (2017). The influence of concreteness of concepts on the integration of novel words into the semantic network. *Frontiers in Psychology*, *8*, 2111.
- Feldman, L. B., & Siok, W. W. (1997). The role of component function in visual recognition of Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 776–781.
- Feldman, L. B., & Siok, W. W. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language*, *40*(4), 559–576.

- Guo, C. Y., Zhu, Y., Ding, J. H., & Fan, S. L. (2004). An event-related potential study on the relationship between encoding and stimulus distinctiveness. *Acta Psychologica Sinica*, *36*(4), 455–463.
- Hoffman, P., Rogers, T. T., & Ralph, M. A. L. (2011). Semantic diversity accounts for the “missing” word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience*, *23*(9), 2432–2446.
- Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of contextual variation in word meaning based on latent semantic analysis. *Behavior Research Methods*, *45*(3), 718–730.
- Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721–742.
- Hsu, C. H., Tsai, J. L., Lee, C. Y., & Tzeng, O. J. L. (2009). Orthographic combinability and phonological consistency effects in reading Chinese phonograms: An event-related potential study. *Brain and Language*, *108*(1), 56–66.
- Huang, X., Lin, D., Yang, Y. M., Xu, Y. H., Chen, Q. R., & Tanenhaus, M. (2021). Effects of Character and Word Contextual Diversity in Chinese Beginning Readers. *Scientific Studies of Reading*, *25*(3), 251–271.
- Institute of Linguistics of Chinese Academy of Social Sciences. (2012). *Modern Chinese Dictionary*. 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*(4), 434–446.
- Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *66*(2), 115–124.
- Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. *Psychology of Learning and Motivation*, *67*, 239–283.
- Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, *4*(2), 99–107.
- Kim, S. Y., Qi, T., Feng, X., Ding, G., Liu, L., & Cao, F. (2016). How does language distance between L1 and L2 affect the L2 brain network? An fMRI study of Korean–Chinese–English trilinguals. *Neuroimage*, *129*, 25–39.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.
- Kwon, Y., Nam, K., & Lee, Y. (2012). ERP index of the morphological family size effect during word recognition. *Neuropsychologia*, *50*(14), 3385–3391.
- Lartseva, A., Dijkstra, T., Kan, C. C., & Buitelaar, J. K. (2014). Processing of emotion words by patients with autism spectrum disorders: Evidence from reaction times and EEG. *Journal of Autism and Developmental Disorders*, *44*(11), 2882–2894.
- Lee, C. Y., Tsai, J. L., Su, C. I., Tzeng, J. L., & Hung, L. (2005). Consistency, regularity, and frequency effects in naming Chinese characters. *Language and Linguistics*, *6*(1), 75–107.
- Lee, C. Y., Hsu, C. H., Chang, Y. N., Chen, W. F., & Chao, P. C. (2015). The feedback consistency effect in Chinese character recognition: Evidence from a psycholinguistic norm. *Language and Linguistics*, *16*(4), 535–554.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3.
- Li, X., Huang, L., Yao, P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology*, *1*(3), 133–144.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.
- Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of bayes factors for common designs. R package v0.9.12-4.2. URL: <http://CRAN.R-project.org/package=BayesFactor>
- Müller, O., Duñabeitia, J. A., & Carreiras, M. (2010). Orthographic and associative neighborhood density effects: What is shared, what is different? *Psychophysiology*, *47*(3), 455–466.
- 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*(1), 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*(1), 275–283.
- R Development Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>
- Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). The time course of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 11.
- Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 82–102.
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499–520.
- Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 761–775.
- Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language*, *40*(4), 498–519.
- Vergara-Martínez, M., Comesaña, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(3), 461–474.
- Verkoeijen, P. P., Rikers, R. M., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 796–800.
- Yan, G., Bai, X., Zang, C., Bian, Q., Cui, L., Qi, W., Rayner, K., & Liversedge, S. P. (2012). Using stroke removal to investigate Chinese character identification during reading: Evidence from eye movements. *Reading and Writing*, *25*(5), 951–979.
- Yan, S. R., Mollica, F., & Tanenhaus, M. K. (2018, July). A context constructivist account of contextual diversity (pp. 1205–1210). *Proceedings of the 40th Annual Meeting of the Cognitive Science Society, USA*.
- Ye, J., Nie, A., & Liu, S. (2019). How do word frequency and memory task influence directed forgetting: An ERP study. *International Journal of Psychophysiology*, *146*, 157–172.
- Yum, Y. N., & Law, S. P. (2019). Interactions of age of acquisition and lexical frequency effects with phonological regularity: An ERP study. *Psychophysiology*, *56*(10), e13433.
- Zhang, Q., Guo, C., Ding, J., & Wang, Z. (2006). Concreteness effects in the processing of Chinese words. *Brain and Language*, *96*(1), 59–68.
- Zhao, X., Chen, A., & West, R. (2010). The Influence of Working Memory Load on the Simon Effect. *Psychonomic Bulletin & Review*, *17*(5), 687–692.
- Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal*

*of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1245–1265.

Zou, Y., Tsang, Y., & Wu, Y. (2019). Semantic radical activation in Chinese phonogram recognition: evidence from event-related potential recording. *Neuroscience*, 417, 24–34.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.