Applied Psycholinguistics (2020), **41**, 727–751
doi:10.1017/S0142716420000181

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Reading minds in motion: Mouse tracking reveals transposed-character effects in Chinese compound word recognition

Yu-Cheng Lin[1,*] and Pei-Ying Lin[2]

[1]University of Texas Rio Grande Valley and [2]University of Saskatchewan
*Corresponding author. Email: yucheng.lin@utrgv.edu

(Received 06 February 2018; revised 31 March 2020; accepted 14 April 2020; first published online 28 August 2020)

Abstract

This study investigated the development of character transposition effects during Chinese compound word recognition via computer mouse movements instead of the conventional key presses. Empirical evidence to reveal the impacts of vocabulary knowledge, grade level, and whole word frequency on Chinese transposed-character effect is lacking. In the present study, we measured the transposed-character effect in two groups of Taiwanese children (second and fourth graders) in a mouse-tracking lexical-decision task including nonwords derived from real words by transposing two characters (e.g., “習學” from “學習” [learning]) and control nonwords in which two characters are replaced (e.g., “以修”). Our results indicate that participants showed longer mouse movement times and larger spatial attraction in recognizing transposed-character nonwords than in replaced-character nonwords, suggesting that the dominant role of whole-word representation in processing Chinese compound words. Our results also further demonstrate that how the degree of character transposition was affected by vocabulary knowledge, grade level, and word frequency.

Keywords: mouse-tracking; reading development; transposed-character effect

How do readers process written words? Reading is a complicated skill involving the orchestration of a number of lexical representations in terms of orthography, phonology, and semantics. In recent decades, there has been a dramatic increase the investigation of the orthographic processing in the mental lexicon. It is widely reported that there are two different orthographic processes of visual word recognition: letter identity and letter position. Successful visual word recognition not only requires actual identification of the letter within a word but also encoding of the letter position. The flexible letter position coding is well documented in current literature. Previous masked priming studies on alphabetic languages have consistently shown that transposed-letter nonwords (e.g., “jugde”–JUDGE) facilitate target word recognition relative to orthographic controls in which two different letters are replaced (e.g., “jupre”–JUDGE; Perea & Lupker, 2003, 2004; Perea et al., 2005).

The transposed-letter (TL) facilitation effect suggests that skilled readers are tolerant to disruptions of letter positions, also showing that their whole-word orthographic representations are automatically activated even when internal letters have been transposed in a word.

The location uncertainty in letter position coding has also been extended to the nonalphabetic languages such as Chinese. The flexibility of Chinese character position coding of compound word recognition has been repeatedly observed in adults (Bai et al., 2011; Cheng, 1981; Mattingly & Xu, 1994; Peng et al., 1999; Taft et al., 1999). However, remarkably little research has been conducted with children to determine whether the character transposition effect could emerge in the early Chinese reading development. This information gap prevents us from fully understanding the developmental trajectory of character position coding and mechanisms for the promotion of effective visual word recognition skills that are unique to Chinese readers. Liu et al. (2014) found that Chinese elementary school students (second graders and third graders) and high school students responded to transposed-character nonwords (e.g., "麗美") more slowly and less accurately than they did to replaced-character nonwords (e.g., "麗新") and real words ("圖書") in a lexical-decision task. Similar results were reported by Liu et al. (2010) of fourth graders using a similar lexical-decision task. Furthermore, Liu et al. (2014) demonstrated that the magnitudes of transposed-character effects are similar among second and third grade students, suggesting that whole-word orthographic representation emerges at an early stage of reading acquisition. Although previous studies reveal some uncertainty surrounding character position coding of Chinese reading development, they do not address to what extent stimuli and subjects modulate the magnitude of the transposed-character effect of Chinese compound words. To bridge the knowledge gap, this study examines the impacts of word frequency and vocabulary size on the character position coding of Chinese compound words by focusing on the developmental trajectories of developing readers, which are both insufficiently understood and understudied.

The effect of whole-word frequency

Substantial evidence on visual word recognition in alphabetic languages has identified several lexical variables, including the orthographic, phonological, and semantic properties such as whole-word frequency, phonological feature, and semantic transparency, that significantly affect the processing efficiency of visual word recognition (see Giraudo & Dal Maso, 2016, for a review). Of these lexical variables, word frequency (i.e., the number of times a reader has encountered a word) is considered one of the strongest predictors of processing efficiency for visual word recognition among readers. Understanding the influence of word frequency on lexical access can help us infer the organization, structure, and processing of the mental lexicon (Joseph et al., 2013).

One of the main sources of evidence for the influence of word frequency on the transposed-letter effect comes from unmasked lexical-decision and text reading studies (see Andrews, 1996; O'Connor & Forster, 1981; Perea et al., 2005; White et al., 2008) that have examined the important role of whole-word frequency in the letter position coding of visual word recognition. For example, O'Connor

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

and Forster (1981) found that English-speaking adults exhibited more errors in responding to TL nonwords derived from high-frequency words than in responding to those derived from low-frequency words in a lexical-decision task. Likewise, Andrews (1996) found a larger delay in the response time (around 34 ms) to TL nonwords than to orthographic controls (replaced-letter nonwords) when TL nonwords were created from high-frequency words. However, the TL effect was not observed in the lexical-decision task when TL nonwords were created from low-frequency words. The aforementioned results indicate that high-frequency TL nonwords are more effective at eliciting orthographic representations of their base words than low-frequency TL nonwords are.

A recent event-related potential (ERP) study has also lent further support to the idea that the size of the transposition effect varies as a function of whole-word frequency. For example, in both semantic categorization and lexical-decision tasks, Vergara-Martínez et al. (2013) demonstrated that words that are easy to identify (e.g., high-frequency words) can induce a larger transposition effect than difficult-to-identify words (e.g., low-frequency words) can. Vergara-Martínez et al. (2013) also found that TL nonwords created from high-frequency words exhibited N400 amplitudes similar to those of their base words. In contrast, replaced-letter nonwords created from low-frequency words triggered larger N400 amplitudes than their base words did. The results suggested higher activation of the whole-word representation of the TL nonwords derived from high-frequency words than that of TL nonwords derived from low-frequency words. The ERP results clearly showed that the whole-word access route takes priority in the accessing of high-frequency words, whereas the decompositional route plays a central role in the identification of low-frequency words. Although previous studies have revealed the influence of whole-word frequency on the magnitude of the transposition effect in adults who read morphologically simple words, the question of whether whole-word frequency modulates the transposition effect of morphologically complex words such as compound words remains unanswered in the current literature. In particular, no research has examined the influence of whole-word frequency on the transposed-character effect of Chinese compound words in developing readers during the elementary school years.

The effects of vocabulary knowledge and age

In addition to word frequency, previous studies (e.g., De Zeeuw et al., 2015; Hasenäcker & Schroeder, 2019; Hasenäcker et al., 2017; Häikiö et al., 2011) have demonstrated that morphological processing is strongly tied to vocabulary knowledge (indexed by vocabulary size) in reading development. Developmental studies using the masked priming paradigm have demonstrated that transposed-letter priming decreases with age, indicating that letter position coding becomes stricter as development progresses (Acha & Perea, 2008; Castles et al., 2007). This finding has been explained by a theoretical account called the lexical tuning hypothesis (Castles et al., 1999, 2003, 2007). The hypothesis assumes that the children's age might affect the magnitude of the transposition effect during visual world word recognition. Specifically, the hypothesis predicts that a modification in the lexical

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

tuning of the word recognition system responds to a child's vocabulary size. For old children, high-vocabulary sizes precisely tuned to match the criterion of the visual word recognition leads to a strict letter position coding (a smaller magnitude of transposition effect). For younger children, the matching criterion of word recognition might be less finely tuned by their limited vocabulary sizes in the mental lexicon. Younger children therefore are highly tolerant of the confusability of letter position information and exhibit a larger magnitude of transposed-letter effect in word recognition. Prior alphabetic research has suggested that vocabulary size and age might affect the flexible letter position coding in visual word recognition. Nonetheless, conclusions on nonalphabetic scripts such as Chinese have not been reached yet. The reason for this is that cross-linguistic generalizations might be problematic because the morphological structures are very different between English and Chinese. Morphological systems vary considerably across English and Chinese. For example, Chinese has highly productive lexical compounding, so it contains a relatively large number of compound words but few derived or inflected words. In contrast, lexical compounding in English is less prevalent than inflection or derivation. To address this differentiation, the present study tackled the question of the effects of vocabulary size and age (indexed by grade level) on the magnitude of transposed-character effect in Chinese compound word recognition.

The mouse-tracking lexical-decision paradigm

In the past few years, previous studies (Anderson et al., 2011; Cargill et al., 2007; Hermens, 2018; Krueger et al., 2018; Schroeder & Verrel, 2014) have shown that computer mouse tracking and other motor-tracking techniques (e.g., finger tracking or three-dimensional reach tracking) are feasible for studying language processing in both younger and older children (for a more comprehensive overview of using the mouse-tracking technique with child participants, see Erb, 2018). In particular, the mouse-tracking lexical-decision paradigm has proven to be a powerful and popular research tool for assessing the cognitive processing that underlies visual word recognition (e.g., Barca & Pezzulo, 2012, 2015). In a mouse-tracking lexical-decision task, Lin and Lin (2016) demonstrated that the letter transposition effect could be revealed by not only key press responses (e.g., total response times and error rates) but also mouse movement curvatures, such as the area under the curve (AUC) and maximum deviation, during visual word recognition. Schroeder and Verrel (2014) have shown that hand movement trajectories complement the traditional behavior measures and provide wholly new insights into lexical access in younger children (elementary school Grades 2 and 3), older children (secondary school Grades 5 and 6), and adults. In the present study, the kinematics of hand movements (e.g., computer cursor movement trajectories) were considered as a reliable indicator of the relative lexical activation level of "word" and "nonword" response options underlying the transposed-character effect in developing readers from different grades.

Previous studies (e.g., Bai et al., 2011; Liu et al., 2010, 2014; Peng et al., 1999) on the transposed-character effect in Chinese compound word recognition mainly used outcome-based lexical-decision measures such as response time and errors. This response-focused method has an obvious limitation in that it only reflects the end point of lexical processing, not the real-time processing underlying visual word recognition. Studies using the novel computer mouse-tracking method (e.g., Barca & Pezzulo, 2015; Dale & Duran, 2011; Lin et al., 2015; Lin & Lin, 2016; Tomlinson et al., 2013) instead of traditional button-press measures have shown that the continuous nature of mouse movements presents a more detailed view of the real-time mental processing of lexical processing. Furthermore, compared with other time-sensitive measures such as ERPs, mouse tracking can quantitatively distinguish different decision-making processes from multiple response alternatives. Although ERPs provide an index of the specific timing of neural activity associated with the presentation of a single event or response option, it is difficult to track the parallel and continuous nature of lexical activation from multiple response alternatives. In contrast, mouse-tracking provides a more direct measure of multiple response activations and how the participant accumulates evidence to drive the lexical decision over a span of hundreds of milliseconds.

The unique contribution of the present mouse-tracking study is that it targeted two distinct inhibition processes underlying the transposed-character effect across reading development: a response threshold adjustment process versus a conflict resolution process. According to a reach-tracking study (Erb et al., 2016), these two inhibition processes can be identified by initial time (the time until reach movement onset) and reach movement curvature (the degree to which a movement deviates from a direct path to the selected target). By employing these two key measures, the present study revealed how the response threshold adjustment and conflict resolution processes contribute to the transposed-character effect during compound word recognition. Therefore, one can expect that the presence or absence of transposition effects in the mouse movement parameters (initiation time and AUC) will reveal how a Chinese reader uses different inhibition processes to resolve competing interpretations of a transposed-character stimulus in visual word recognition.

The current study

The present study examined two research questions with regard to the character transposition effect with Chinese compound words. First, we investigated whether participants' hand movement trajectories can be used to index flexible character position coding during visual word recognition. If this is the case, we would expect that developing readers' behavioral responses would be affected by recognizing transposed-character nonwords in a mouse-tracking lexical-decision task. We expected longer movement times for transposed-character nonwords than for replaced-character nonwords. Furthermore, regarding mouse movement trajectories (for example, AUC), we anticipated that the transposed-character nonwords would trigger a greater overall spatial attraction to the "YES" response option indicating the base word rather than the replaced-character nonwords on the mouse movement trajectories. One expect that the mean of the AUC value of a trajectory can serve as a reliable index of spatial attraction toward the unselected alternative (Freeman & Ambady, 2010). A larger AUC indicates a greater deviation in movement trajectory toward the unselected alternative.

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

Second, we examined whether the item (whole-word frequency) and subject (vocabulary size and grade) affect the magnitude of the transposition effect. In the present study, we manipulated the whole-word frequency of a base word (high frequency vs. low frequency) to investigate its influence on transposition effects during visual word recognition. As suggested by the earlier studies in alphabetic languages, high-frequency transposed-letter nonwords should be more effective in activating the whole-word orthographic representation of their base words, as they are perceptually more similar to their base words than to low-frequency nonwords. In a similar vein, we expected that a higher degree of word similarities could be elicited by high-frequency transposed-character nonwords and that they might lead to a larger-magnitude transposition effect in Chinese compound word recognition. The whole-word frequency could thus play a central role in modulating the activation level of whole-word orthographic representation of Chinese compound words. In addition to the word-frequency, we examined the impact of vocabulary size and grade level on the character position processing of Chinese compound word recognition over the course of reading development. This prediction would expand upon previous evidence on compound word processing research in children from a developmental perspective. Based on a recent study by Hasenäcker and Schroeder (2018), we hypothesized that vocabulary size has an independent role in influencing the overall processing and accuracy of compound word reading, unrelated to whole-word frequency. In regard to the effect of grade levels, we predicted that the magnitude of transposed-character effects should be similar across different grades (the second grade vs. the fourth grade, particularly) in elementary school children. This prediction is consistent with a finding from Liu et al. (2014) suggesting that whole-word representation of Chinese compound words can emerge at an early stage of reading development.

Method

Participants

A sample of 82 Taiwanese children separated into two groups by grade completed the mouse-tracking lexical-decision task. The sample consisted of 32 second graders (mean age 9.8 years, $SD = 0.4$, range 9.0–10.0; 12 males and 20 females) and 50 fourth graders (mean age 11.4 years, $SD = 0.5$, range 11.0–12.0; 26 males and 24 females). All participants were native speakers of Mandarin Chinese and Taiwan Southern Min. All participating children had received exposure to Mandarin Chinese and Taiwan Southern Min from birth and literacy instruction at home and/or in after-school programs. They also studied English as a foreign language. The participants were recruited from two public elementary schools in Taiwan. Their homeroom teachers reported no reading difficulties or dyslexia in these participants. Children in the second grade should have developed good knowledge of lexical compounding and are expected to make significant progress beyond the fourth grade, so we included children in the second and fourth grades in our study to provide credible evidence of age-related differences in the effect of transposed characters.

Vocabulary knowledge measures

The Mandarin Chinese version of the Peabody Picture Vocabulary Test–Revised (PPVT; Lu & Liu, 1998) was used to test the children's vocabulary sizes. The PPVT was used to test the children's receptive vocabulary by asking them to point to one of four pictures that represented a spoken word. The maximum score was 60. The measure of word reading reflected larger vocabulary sizes in the fourth graders ($M = 55.1$, $SD = 3.3$) than in the second graders ($M = 50.5$, $SD = 4.9$), with a significant difference between these two groups, $t = -22.0$, $p < .001$. The split-half reliability of the PPVT–R ranges from .90 to .97, with alternate-forms reliability ranging from .60 to .91 and test–retest reliability ranging from .84 to .90. In addition, the children's Chinese character reading ability was assessed with the Chinese Character Recognition Test for School–Age Children (Huang, 2001). Both the internal consistency reliability and split-half reliability of this measure is .99, with test–retest reliability ranging from .81 to .95. In the test, the children were asked to write down the characters one by one. There was a significant difference in character reading between the fourth graders ($M = 124.7$, $SD = 24.2$) and the second graders ($M = 86.8$, $SD = 22.8$), $t = -33.9$, $p < .001$. The maximum score was 200.

Materials

The full lists of the experimental stimuli for second and fourth graders are described in more detail in Appendix A. The stimuli were adapted from the stimulus sets from Experiment 1 in Lau's study (2012), which were chosen from the Hong Kong Corpus of Primary School Chinese (Leung & Lee, 2002). For the second grade stimuli list, a total number of 54 experimental items were selected, as follows: 18 Chinese two-character compound base words, 18 transposed-character nonwords, and 18 replaced-character nonwords. For the fourth grade stimuli list, 78 experimental items (i.e., 26 base words, 26 transposed-character nonwords, and 26 replaced-character nonwords) were selected. Each compound base word was matched with two non-words: (a) a transposed-character (TC) nonword, created by transposing two characters of the base words (e.g., "學習" [learning]–"習學"), and (b) a replaced-character (RC) nonword, created by replacing both characters of the base words with two other unrelated characters (e.g., "學習"–"以修"), which served as an orthographic control for the TC condition.

Example stimuli are provided in Table 1. Four running lists were created such that the experimental items and the positions (top-right and top-left corners) of the response box were counterbalanced across conditions. Half of the base words were presented in the TC condition, and the other half in the RC condition. Each participant received only one of the lists to ensure that no participant saw the same target more than once. That is, each second grade participant saw a total of 36 trials (18 base words, 9 TC nonwords, and 9 RC nonwords), and each fourth grade participant saw a total of 52 trials (26 base word, 13 TC nonwords, and 13 RC nonwords) in the mouse-tracking lexical-decision task.

To reduce the variations in word usage between Hong Kong and Taiwan, we excluded some words that are not used in Taiwan and made sure that the critical words were suitable for Taiwanese elementary students by using the database of the List of Commonly Used Characters and Words for Taiwanese Elementary Students

Table 1. Example stimuli

Grade	Condition	Similarity type	Word frequency	Base word target	Nonword target
Grade 2	HF – TC	Transposed	HF	學習	習學
	HF – RC	Replaced	HF	學習	以修
	LF – TC	Transposed	LF	親切	切親
	LF – RC	Replaced	LF	親切	開情
Grade 4	HF – TC	Transposed	HF	希望	望希
	HF – RC	Replaced	HF	希望	藝正
	LF – TC	Transposed	LF	旅遊	遊旅
	LF – RC	Replaced	LF	旅遊	權唱

Note: HF, high frequency. LF, low frequency. TC, transposed character. RC, replaced character.

(Ministry of Education, 2002). For example, some high-frequency (e.g., "課室") and low-frequency (e.g., "齊備") words could not be used because those items exist in Hong Kong Mandarin vocabulary repositories but not in Taiwan Mandarin.

Half of the base words were high-frequency whole-word compounds, and half were low-frequency whole-word compounds. The high-frequency and low-frequency base words were matched for morpheme/character frequency, constituent morphological family size, number of strokes, and imageability of the first and second constituent characters. Details of the selected stimuli are listed in Table 2. A small number of stimuli remained after the lexical variables were matched. In addition, to avoid the possible confounding variable of semantic transparency, we selected only semantically transparent words for this study. This word selection criterion limited the number of stimuli that could be used in the present study.

According to Lau's argument (2012), "a lack of developmental perspective in choosing stimuli might lead to severe errors because a high-frequency word at the grade four levels could have relatively low frequency at the grade two levels" (see p. 73). In our pilot study, we did find that it was difficult for either group to read the same set of stimuli when they were not age-inappropriate materials for some students. Accordingly, the properties of the stimuli in our materials were carefully controlled from a developmental perspective according to the children's grade levels. For example, for the second grade stimuli, the average whole-word frequency count of the high-frequency words was 4.2% ($SD = 0.4\%$), and for the low-frequency words, it was 1% ($SD = 0\%$). Similarly, for the fourth grade stimuli, the average whole-word frequency count of the high-frequency words was 4.2% ($SD = 0.4\%$); for the low-frequency words, it was 1% ($SD = 0\%$). In the present study, each word item was labeled as either a high- or low-frequency word, following from a Hong Kong Word Corpus of Primary School Chinese consisting of 8,142 words coded for word frequency (Lau, 2012). High- and low-frequency words were operationalized as having a count of 4 or more occurrences and 1 or fewer occurrences per 8,142 words, respectively. Furthermore, the Cronbach's α internal

Table 2. Means and standard deviations of whole-word frequencies, constituent morpheme frequencies, and constituent family sizes of the HF words and LF words selected for different grade levels

	Grade 2 stimuli ($n = 18$)		Grade 4 stimuli ($n = 26$)	
	HF words	LF words	HF words	LF words
Word frequency	4.0 (0.0)	1.0 (0.0)	8.8 (1.2)	1.0 (0.0)
Morpheme frequency				
First morpheme	17.9 (10.7)	14.8 (7.7)	20.5 (14.3)	20.4 (15.2)
Second morpheme	16.2 (18.0)	10.4 (9.4)	29.8 (21.2)	29.5 (21.2)
Family size				
First morpheme	6.9 (5.9)	4.6 (1.3)	6.5 (5.9)	8.7 (3.5)
Second morpheme	4.7 (3.0)	3.1 (2.1)	7.6 (5.8)	9.5 (5.3)
Imageability				
First morpheme	6.1 (0.3)	5.5 (0.8)	5.4 (1.0)	5.3 (1.1)
Second morpheme	5.1 (1.3)	5.9 (0.7)	5.5 (0.5)	5.8 (0.8)
Number of strokes				
First morpheme	11.0 (4.2)	10.0 (4.0)	11.8 (4.2)	9.3 (3.1)
Second morpheme	8.8 (3.3)	11.3 (4.7)	9.7 (3.8)	10.7 (2.9)

Note: HF, high frequency. LF, low frequency.

consistency coefficients were .85 for the second grade stimuli, .86 for the fourth grade stimuli, and .86 for the entire task.

Procedure

Participants sat in a quiet testing room in front of a computer screen. They first performed 10 practice trials to familiarize themselves with the procedure; the experimental trials followed. Stimulus presentation was controlled by MouseTracker software (sampling rate of mouse pointer, 70 Hz; Freeman & Ambady, 2010). As shown in Figure 1, participants initiated each trial by clicking a "開始/START" button centrally located at the bottom of the screen. This button was then replaced by two-character compound words or nonwords in Chinese. The character string remained on the screen for 3000 ms. Participants were asked to quickly and accurately determine whether or not the word was a real word by moving to and clicking a "是/YES" (indicating it is a word) or "否/NO" (indicating it is a nonword) box in the top right and left corners of the screen, respectively. The stimulus order was randomized.

If a participant's total response time for a lexical decision on a particular trial exceeded 3000 ms, "TIME OUT" appeared on the screen. We excluded trials with total response times slower than 3000 ms and trials where the participant did not respond correctly. For the remaining correct trials, we eliminated values above and below the ±3 *SD* limits around the grand mean of the participants. Participants were given feedback (i.e., whether their responses were correct or incorrect) at the end

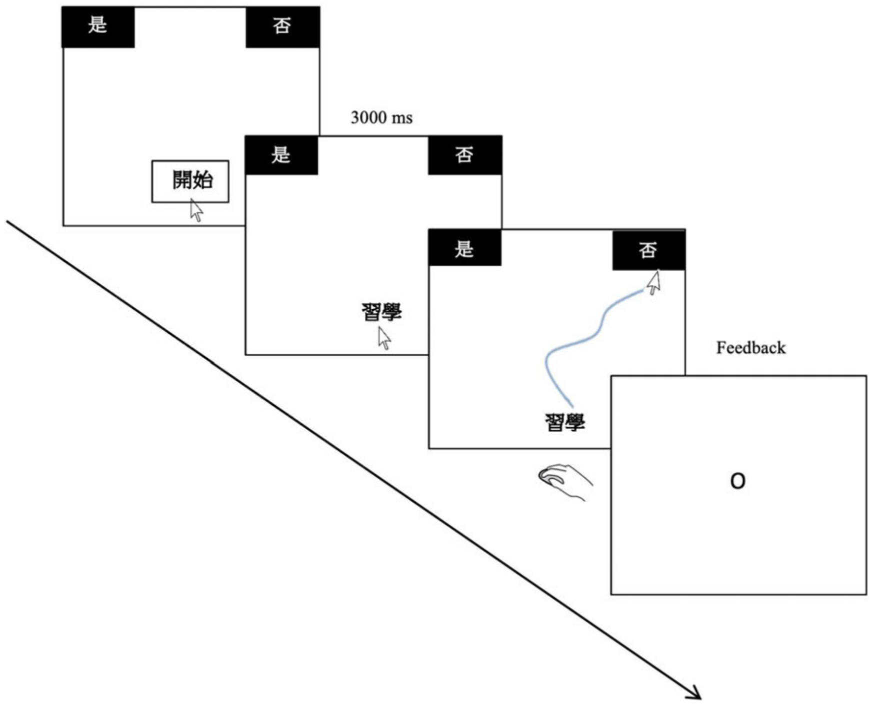

Figure 1. Schematic of a single trial for the mouse-tracking lexical-decision task.

of each trial to help them self-monitor their performance. The temporal characteristics and *x*-, - coordinates (in pixels) of mouse trajectories were recorded by the MouseTracker software.

Results

All analyses were performed in R (R Core Team, 2015) and the R package "lme4" (Bates et al., 2015). We used linear mixed models (LMMs) analysis[1] to evaluate the impacts of transposed-character similarity, grade, and word frequency on total response time (TRT), initiation time (IT), movement time (MT), and AUC. TRT was the time elapsed between clicking of the START button and the response completion, and it was similar to the response time of discrete responses (e.g., key presses or verbal responses) in a traditional lexical-decision task. IT referred to the time in macroseconds at which the first mouse movement was initiated (the time elapsed between stimulus presentation and mouse movement initiation). If the TC effect was observed in IT, we would expect the TC effect to precede mouse movement execution. In contrast, MT (MT = TRT – IT) was defined as the interval between the onset of the initial mouse movement and response completion. If the TC effect affected MT, then one would assume that the TC effect would continue to influence actions even after a mouse movement had been initiated.

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

Separate LMMs were created for each dependent measure. Log transformations were conducted on TRTs, ITs, and MTs to correct for distribution skews. We also performed a generalized linear mixed-effects (GLMMs) analysis on the accuracy. As fixed effects, we entered similarity type, grade, word frequency, and their interactions into the models. As random effects, we had intercepts for subjects and items. Each of the categorical variables was effect (sum) coded.

Additional analyses paralleled the original analysis and were conducted to examine whether the magnitude of transposed-character effects was related to the participants' vocabulary size and single-character reading. Previous studies (e.g., Liu et al., 2010; Wang & McBride, 2016) found that character and word reading might constitute somewhat different processes in Chinese children. To investigate their effects on the similarity type, we separately entered them following two continuous measures, such as vocabulary size and character reading, as fixed factors and their interactions with similarity type in the LMMs and GLMMs models. These two continuous measures were centered around their respective means.

The 82 participants completed a total of 1,876 trials. Of these, 499 trials (27%) resulted in incorrect responses, and they were excluded in data analysis. For the time analysis, outliers were removed by excluding TRT, INIT, MT, and AUC values that exceeded three standard deviations (TRT: 9 trials, 0.5%; IT: 24 trials, 1.3%; MT: 10 trials, 0.6%; AUC: 3 trials, 0.2%) from the mean for each participant across the correct trials. We also excluded an additional 41 trials (2.3%) for which response times exceeded the deadline (3000 ms). All subsequent analyses of RTs (1,327 trials), ITs (1,312 trials), MTs (1,326 trials), and AUCs (1,333 trials) were conducted on the remaining trials from the participants. The descriptive statistics of the TRTs, ITs, MTs, AUCs, and accuracy rates in the mouse-tracking lexical-decision task are provided in Tables 3 and 4.

Total response times

To determine whether the main effects of similarity type (transposed character vs. replaced character), grade (second grade vs. fourth grade), compound word frequency (low- vs. high- whole-word frequency), and their interactions were significant, the Kenward–Roger approximation[2] (Kenward & Roger, 1997) was implemented using the mixed function from the *R* package (Singmann et al., 2015) to estimate the *p* values in our models. The LMM model of the total response time revealed significant main effects of similarity type ($B = 0.04$, $t = 3.53$, $p < .001$), grade ($B = -0.24$, $t = -6.45$, $p < .0001$), and frequency ($B = -0.04$, $t = -3.13$, $p < .01$). We found that our participants' response times with the transposed-character nonwords were 84 ms slower than response times with the RC nonwords (1578 vs. 1493 ms), indicating TC effects in both second and fourth graders. In addition, the second graders spent significantly more time recognizing the nonwords (across TC and RC nonwords) than the fourth graders did, as indicated by the negative estimate (1743 vs. 1375 ms). Furthermore, the lexical decision times were longer for low-frequency words (1562 ms) than for high-frequency words (1509 ms). However, the three-way interaction among similarity type, grade, and word frequency was not significant ($B = -0.03$, $t = -0.63$, $p = .53$). In addition, neither the interaction of Similarity Type × Grade ($B =0 .02$, $t = 0.82$, $p = .41$) nor

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

Table 3. Mean and standard errors for total response times, initiation times, and movement times by word frequency, similarity type, and grade

	Total response times							
	TC				RC			
	HF		LF		HF		LF	
Grade	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Second	1751.4	36.97	1800.17	34.57	1690.11	32.99	1731.5	36.93
Fourth	1393.42	28.72	1444.67	29.47	1329.18	26.17	1350.36	28.13
	Initiation times							
	TC				RC			
	HF		LF		HF		LF	
Grade	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Second	208.76	13.61	205.97	13.92	203.31	12.8	207.12	13.9
Fourth	178.35	9.97	140.48	11.47	188.94	10.23	178.72	10.76
	Movement times							
	TC				RC			
	HF		LF		HF		LF	
Grade	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Second	1537.57	38.31	1573.51	36.16	1466.88	32.15	1496.74	36.83
Fourth	1216.61	30.92	1253.73	31.54	1133.79	27.17	1161.3	29.34

Note: HF, high frequency. LF, low frequency. TC, transposed character. RC, replaced character. Table 3 shows untransformed but trimmed reaction times, initiation time, and movement times for correct responses.

Table 4. Mean and standard errors for area under the curve values and accuracy rates by word frequency, similarity type, and grade

	Area under the curve values[a]							
	TC				RC			
	HF		LF		HF		LF	
Grade	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Second	2.31	0.14	2.24	0.13	1.73	0.14	1.90	0.15
Fourth	2.62	0.12	2.43	0.13	1.99	0.12	2.14	0.13
	Accuracy rates (%)							
	TC				RC			
	HF		LF		HF		LF	
Grade	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Second	75	3	81	3	82	3	73	3
Fourth	80	2	60	3	86	3	76	3

Note: HF, high frequency. LF, low frequency. TC, transposed character. RC, replaced character. [a]Untransformed but trimmed area under the curve values for correct responses.

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

that of Similarity Type × Frequency ($B = -0.02$, $t = -0.76$, $p = .44$) was significant, suggesting that grade level and frequency did not affect the magnitude of the TC effect.

Similar to the initial data analysis, we also found a main effect of similarity type when both vocabulary size and character reading ($B = 0.0347$, $t = 2.57$, $p < .05$) were added to the model. There were also main effects of vocabulary size and character reading ($B = -0.0207$, $t = -3.24$, $p < .01$, and $B = -0.0025$, $t = -3.13$, $p < .01$), indicating that both a larger vocabulary size and a greater character reading ability resulted in faster response times. However, there were no interactions between similarity type and vocabulary size ($B = -0.0039$, $t = -0.934$, $p = .3504$) or similarity type and character reading ($B = 0.0007$, $t = 1.367$, $p = .1720$).

Initiation times

As before, using the Kenward–Roger approximation, we did not find any significant main effects of similarity type, grade, frequency, or their interactions. Our results revealed neither main effects of similarity type ($B = -0.014$, $t = -0.275$, $p = .783$) nor interactions between similarity type and vocabulary size ($B = 0.018$, $t = 1.08$, $p = .283$) or similarity type and character reading ($B = -0.0005$, $t = -0.238$, $p = .812$). There were no significant main effects of vocabulary size ($B = -0.021$, $t = -1.050$, $p = .297$) or character reading ($B = 0.002$, $t = 0.956$, $p = .342$).

Movement times

Movement time results indicated significant effects of similarity type ($B = 0.051$, $t = 3.46$, $p < .001$), grade ($B = -.258$, $t = -5.79$, $p < .0001$), and frequency ($B = -0.038$, $t = -2.56$, $p < .011$). The participants' movement times for the TC nonwords were 93 ms slower than those for the RC nonwords (1378 vs. 1285 ms), indicating TC effects in both second and fourth graders. Moreover, the second graders spent significantly more time recognizing the nonwords across TC and RC nonwords than did the fourth graders (1519 vs. 1187 ms). Furthermore, lexical decision times were longer for low-frequency words (1353 ms) than for high-frequency words (1309 ms). None of the other effects were significant.

After entering both vocabulary size and character reading into the final data analysis, we also found significant main effects of similarity type ($B = 0.047$, $t = 2.741$, $p < .01$), vocabulary size ($B = -0.020$, $t = -2.723$, $p = .0079$), and character reading ($B = -0.004$, $t = -3.822$, $p = .0003$). Neither the interaction between similarity type and vocabulary size ($B = -0.005$, $t = -1.004$, $p = .3155$) nor that between similarity type and character reading ($B = 0.0006$, $t = 0.827$, $p = .4082$) was significant.

Accuracy rates

In the GLMM with accuracy rates as the dependent outcome, our results indicated a three-way interaction among similarity type, grade, and word frequency ($B = 1.17$, $t = 2.35$, $p < .05$). The results also indicated significant effects of word frequency ($B = 0.47$, $t = 3.79$, $p < .001$) and similarity type ($B = -0.31$, $t = -2.46$, $p < .05$). There was no significant interaction between similarity type and word frequency ($B = 0.32$, $t = -1.27$, $p = .2053$).

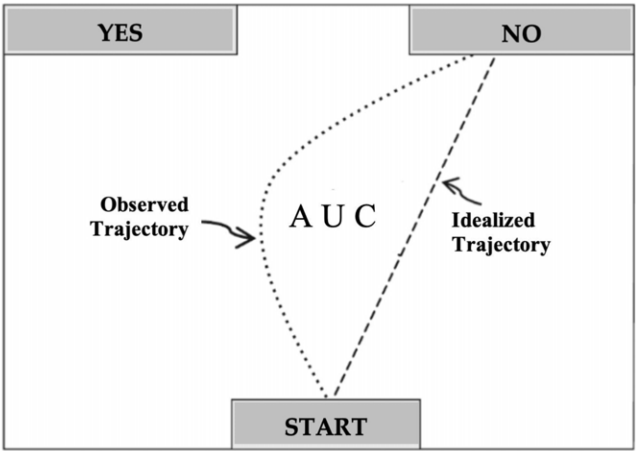

Figure 2. Illustration of the area under the curve (AUC) in our mouse-tracking lexical-decision task.

However, we did find a significant interaction between similarity type and grade ($B = -0.72$, $t = -2.88$, $p < .01$). An examination of simple main effects revealed that the effect of similarity type was significant for fourth graders ($B = -0.67$, $t = -4.29$, $p < .0001$), but not for second graders ($B = 0.06$, $t = 0.31$, $p = .753$). The results showed that the fourth graders had lower accuracy rates on the TC nonwords than on the RC nonwords (70% vs. 81%), whereas there was no difference in accuracy rates between TC nonwords and RC nonwords for the second graders (78% vs. 77%). This suggests that, unlike the second graders, the fourth graders exhibited a reliable transposed-character effect. Neither the interaction between similarity type and vocabulary size ($B = 0.01$, $t = 0.172$, $p = .8636$) nor the interaction between similarity type and character reading ($B = -0.004$, $t = -0.93$, $p = .353$) was significant.

AUC

To evaluate the overall spatial attraction of the mouse trajectory, AUC values were also introduced into the LMM analysis. As shown in Figure 2, the AUC is the geometric area enclosed by the observed mouse-trajectory and the idealized straight-line trajectory connecting the starting position and the end position (Freeman & Ambady, 2010). The main effect of similarity type was significant ($B = 0.521$, $t = 6.22$, $p < .0001$); the results demonstrated that the TC nonwords exhibited a greater overall spatial attraction to the "YES" response option than did the RC nonwords across grades (see Figure 3). As we expected, this result indicated that the TC nonwords elicited a larger spatial attraction to the base word response option. Nonetheless, the main effects of grade ($B = 0.227$, $t = 1.14$, $p = .2576$) and word frequency were not significant ($B = -0.153$, $t = -1.823$, $p = .0686$), nor were the other two-way or three-way interactions significant. Most important, our results indicated a significant interaction between similarity type and word frequency ($B = 0.37$, $t = 2.22$, $p < .05$). The interaction between similarity type and frequency was explored employing the *R* package (De Rosario-Martínez, 2015). An examination

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

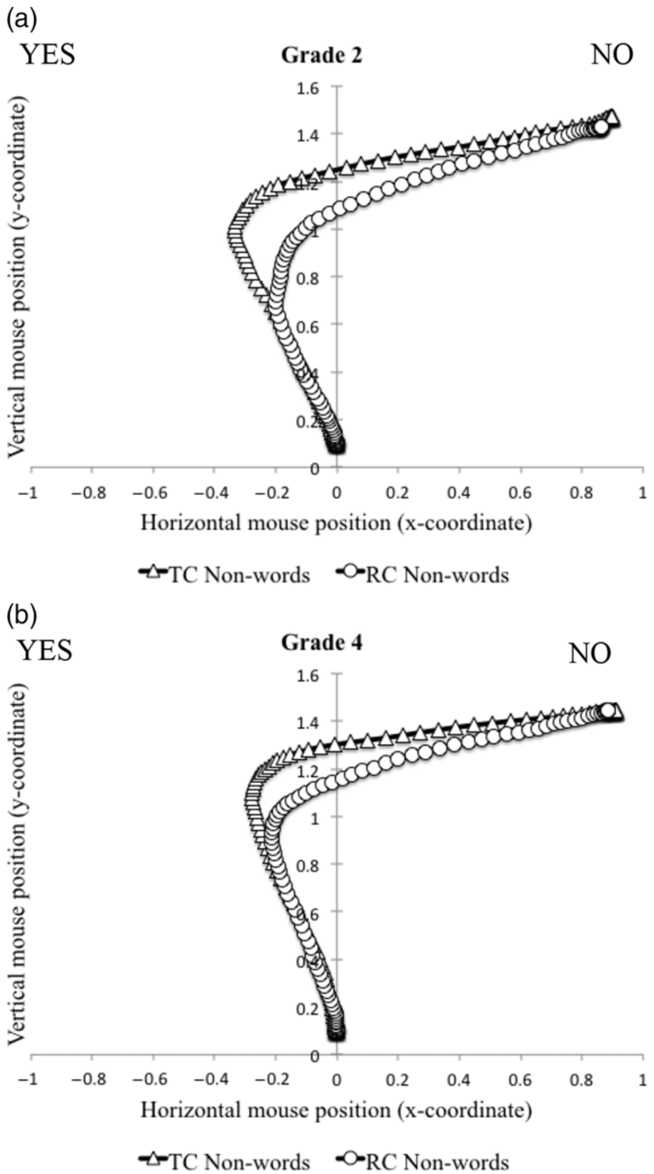

Figure 3. Schematic illustrations of mean movement trajectories for transposed-character nonwords (TC nonwords, triangles) and replaced-character nonwords (RC nonwords, circles) in the (a) second graders and (b) fourth graders. In both panels, it is clear that the trajectories for the TC nonwords were strongly attracted to the "YES" word response for a significant portion of the movement response as compared to RC nonword trajectories.

of simple main effects revealed that the effect of similarity type was significant for the nonwords derived from both high-frequency words ($\chi^2 = 39.40$, $df = 1$, $p < .0001$) and low-frequency words ($\chi^2 = 7.58$, $df = 1$, $p < .01$). Of interest, the AUC data further showed that the magnitude of the effect of similarity type was

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

greater for high-frequency words than for low-frequency words ($\chi^2 = 5.25$, $df = 1$, $p < .05$). As for the effects of vocabulary size and character reading, we only found significant main effects of similarity type ($B = 0.5243$, $t = 5.329$, $p < .000$), but no interactions between similarity type and either vocabulary size ($B = 0.008$, $t = 0.273$, $p = .785$) or character reading ($B = -0.005$, $t = -1.224$, $p = .221$). None of the other effects were significant.

Discussion

In the present study, we used a mouse-tracking lexical-decision task to examine the emergence of whole-word processing of compound words via the TC effect in early Chinese reading development. Mouse tracking was used to record the overall spatial attraction of computer mouse movement (AUC) toward the alternative response option, suggesting that lexical processing cascades into motor execution. We found that the TC effect can be observed in the kinematics of mouse movements in second and fourth graders. As predicted, TC nonwords triggered a greater overall spatial attraction to the real-word response option (the base word) relative to the RC nonwords, as exhibited by the mouse movement trajectories over time during lexical decision. That is, TC effects were observed in the AUC values, in both second and fourth graders. Based on this finding, movement trajectories of the computer mouse cursor can serve as a reliable index of the TC effect in Chinese reading development and mouse tracking presents a new tool to study the links among perception, cognition, and action across reading development, which is a major strength of the current study. Our finding also suggests that the mouse trajectory and other conventional dependent variables (e.g., accuracy and reaction time) are complementary and they can be tapped into similar TC effects. Similar to the mouse trajectory results, the total response times and movement times of all the children were higher in the TC condition than in the RC condition. Our results on total response times and movement times are similar to those of previous studies in Chinese adults and children (e.g., Bai et al., 2011; Liu et al., 2010, 2014; Peng et al., 1999; Taft et al., 1999) performing lexical-decision tasks, indicating that Chinese readers tend to spend more time and make more errors in the TC condition than in the RC condition. For example, Liu et al. (2014) demonstrated that both second- and third-grade elementary students showed very similar sizes of the TC effect on both response times and errors in a lexical-decision task.

Combining the results from mouse movement trajectories, total response times, movement times, and accuracy rates, we can conclude that there was a robust TC effect, showing that developing Chinese readers exhibited stronger spatial attraction to the base words and more motor response conflict for TC nonwords than for RC nonwords. The results of the present study lead to an assumption that the whole-word access route of Chinese compound word reading could emerge in the early reading acquisition of children.

Unlike a previous study of Liu et al. (2014) with a conventional lexical-decision task, the unique contribution of the present mouse-tracking study not only informs us whether a lexical-decision process is fast or slow but also provides a more direct real-time and data-rich window into the dynamics of inhibition processes

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

underlying compound word recognition. The present study has identified two distinct types of inhibition processes during the lexical-decision task: the response threshold adjustment and conflict resolution processes underlying the TC effect of Chinese compound word recognition. These two motor movement measures (i.e., initiation time and AUC) were documented in previous literature (e.g., Erb et al., 2016) to index the response threshold adjustment process and the conflict resolution process, respectively. In the current study, we did not find a significant TC effect on initiation time (i.e., the time elapsed between the beginning of a trial and the initiation of mouse movement), which suggests that there was no difference between TC and RC nonwords when participants initiated their movement responses. Based on this result, it is likely that TC and RC nonwords might trigger comparable degrees of motoric stopping before a movement is started, which might be due to their similar response thresholds. The result shows us the TC effect might not arise at an early representational stage, which implies that the activation of whole-word orthographic representation did not occur before motor response selection stage of lexical access. Still, the stimuli used and other confounding factors may contribute to the nonsignificant effect; they have yet to be fully explored using real-time metrics of conflict, and we look forward to future research in testing the assumption.

In contrast, we detected the transposition effect in the mouse movement trajectories and movement times (the time elapsed between movement onset and response completion) throughout ongoing motor responses. It is more likely that the observed transposition effect occurs at a lexical-decision stage in which two orthographic codes compete (i.e., the competition between real-word and nonword orthographic nodes) and the competition feeds forward into the later motor response stage. The character transposition effect arises as response competition during the preparation and execution of the manual response stage. A similar explanation was elaborately explored in other mouse-tracking studies using a lexical-decision task (see Barca & Pezzulo, 2012; Barca et al., 2017), which reported that the interference arises at the manual response/late stage, in which two orthographic codes (i.e., the real-word versus nonword) are activated in functionally independent routes, and then these codes compete at a motor decision stage during lexical access. Such a conflict resolution process at the motor response stage also has been explained by the late interaction model (see Santens & Verguts, 2011, for a detailed discussion about the shared-decision account) in other cognition domains (e.g., numerical cognition). Another important finding of the present study is that the magnitude of the transposition effect did not vary as a function of grade level and vocabulary knowledge. Consistent with one finding of a past study by Liu et al. (2014), our results indicate that the Taiwanese children exhibited the TC effect regardless of their different grades. The second graders (mean age = 9.8 years) showed a magnitude of the TC effect similar to that of the fourth graders (mean age 11.4 years). Liu et al. (2014) found a nonsignificant interaction effect between grade level and similarity type, which also demonstrated that the magnitude of the TC effect was similar in young readers of Chinese in the second (mean age = 7.3 years) and third (mean age = 8.3 years) grades. These results imply flexibility in character position encoding during visual word recognition throughout development and that developing readers of Chinese are less sensitive to constituent morphemes during compound word recognition in early stages of reading acquisition.

Moreover, we found that individual differences in vocabulary knowledge did not affect whole-word processing in children's reading acquisition and development in elementary school. Although previous studies reported that the developmental trajectories of morphological processing are moderated by individual differences in vocabulary knowledge (Hasenäcker et al., 2017), our empirical results are not in line with the previous findings. Our finding suggests a weaker effect of vocabulary knowledge on orthographic or morphological processing mechanisms in developing readers of Chinese, though vocabulary knowledge of the quality lexical representations plays an important role in the crucial building blocks of reading development. There are two plausible explanations for the finding. First, we reason that written vocabulary knowledge can instead be expected to feed more directly into the semantic process, and this might not reinforce the character transposition in orthographic or morphological processes in early reading development. Second, it is also possible that individual differences in vocabulary knowledge are associated with differences in the manner and/or extent of the decompositional route instead of the whole-word route, as Andrews and Lo (2013) suggested may be true in skilled readers. As a result, a more comprehensive understanding of direct comparison of the whole-word frequency and constituent frequency of compound words in reading development across the elementary school years will be needed. It is important to concurrently manipulate constituent frequency and whole-word frequency in children within the scope of developmental studies on Chinese compound word recognition. Doing so would provide further insight into the character position processing mechanisms underlying the emerging ability to recruit the whole and decompositional access routes in visual word recognition.

Furthermore, we found that whole-word frequency significantly modulated the TC effect in accuracy and hand movement trajectory (i.e., AUC) rather than response durations (i.e., total response time, initiation time, and movement time). Our results suggest that word frequency variation mainly affected the participants' accuracy and hand trajectories when they categorized lexical stimuli but did not impact their processing speed. The previous literature on the character position processing is predominantly based on accuracy and latency measurements of key-press responses. The current study expands on existing evidence by revealing the influence of lexical frequency on character position processing during Chinese compound word recognition through the measurement of computer mouse-movement trajectories. Our data clearly indicate that a larger magnitude TC effect exists for nonwords derived from high-frequency words than for nonwords derived from low-frequency words, suggesting that high-frequency TC nonwords are more effective in activating the whole-word orthographic representation of their base words and that they are perceptually closer to their base words than low-frequency nonwords are. Similar findings were demonstrated in previous research in English-speaking adults with lexical-decision, naming, and semantic categorization tasks (see Andrews, 1996; O'Connor & Forster, 1981; Perea et al., 2005; Vergara-Martínez et al., 2013). Based on this compelling evidence, we can infer that lexical frequency might be an influential factor affecting the magnitude of TC effects in developing children.

An important theoretical implication of the present study is the introduction of a nonsensitive character position encoding assumption to the premise of previous models in Chinese compound word recognition (see Li et al., 2009, for the word segmentation and recognition model; Taft, 2004 for the multilevel interactive model; Zhou & Marslen-Wilson, 1995, for the multilevel cluster representation model). Contrary to previous models, our result demonstrated the flexibility of character position encoding in Chinese compound word reading. The present study not only delineates the flexibility of positional information in learning to read but also reveals the developmental trend of accessing compound words at the whole-word level, providing novel insights about the significant effect of word frequency on whole-word representation. Furthermore, this point is of interest because it can advance current word reading interventions and reading ability assessments of Chinese compound word recognition with respect to whole-word level processing across reading development.

A potential limitation in our study is the relatively small number of stimuli, which might limit the generalizability of our findings. Similarly to previous Chinese compound word studies, we had difficulty in selecting sufficient numbers of paired sets of compound words that differed in a categorical variable (e.g., high- vs. low-frequency words) with good matches on other lexical variables such as the morpheme frequency, morphological family size, number of strokes, and imageability of the first and second constituent characters (see Tse et al., 2017, for a comprehensive discussion). In addition, it is essential to note that all the compound words used in the stimuli sets were semantically transparent instead of semantically opaque. This criterion for stimulus selection prevented the present study from using a larger number of stimuli. In future studies, we can address this issue by including semantically opaque compounds.

Conclusion and future direction

This research aimed to investigate a developmental trend of the TC effect in Chinese compound word processing by examining the kinematics of mouse movements in developing readers. Our findings indicate that developing readers of Chinese exhibit flexibility in character positional encoding and accessing the base word representation while recognizing two-character compound words. Our results are in line with the findings reported by Liu et al. (2014) in their testing of second and third graders. Liu et al. (2014) reported similar results for Chinese elementary school students. Their results indicated that second- and third-grade students show very similar sizes of TC effects in the response times and errors in a lexical-decision task. Unlike their study, the current study did not include proficient readers of Chinese such as high school students or adults. To obtain a clearer developmental trend of the transposition effect in Chinese reading development, the inclusion of proficient readers of Chinese should be further considered in future research.

To date, it should be noted that relatively few studies have investigated how the size of the TC effect changes across grades in the development of Taiwanese children's Chinese reading. Unlike response-based outcomes, such as the lexical-decision task adopted by Liu et al. (2014), we employed a novel experimental paradigm that also produced a similar result for response-time data. Nonetheless, it is difficult, using only outcome-based measures like response-time and accuracy rate, to infer anything about lexical and cognitive processes (e.g., inhibitory control processes) of the TC effect in Chinese children. The response data from previous studies cannot be used to examine whether the TC effect arises in an early representational stage or in a late motor decision stage. In contrast, our mouse-tracking task was a time-sensitive and process-tracking methodology designed to detect the TC effect that occurred before the response was executed or during the time course of the mouse movement. The mouse-tracking lexical-decision paradigm required participants to reject the unselected alternative response from the base word (i.e., the "YES" response option). Therefore, this paradigm not only added to the lexical processing demand but also increased the probability of detecting the inhibition processes underlying the transposition effects via mouse initiation time and movement trajectories.

In summary, the current study, by using mouse movement trajectories, showed that the magnitude of the TC effect does not change across grades in elementary school. A similar finding was obtained by Liu et al. (2014). Our mouse-movement data demonstrated that the TC nonwords had strong effects on the mouse-movement trajectories at the late motor response decision stage in Chinese reading development. Specifically, the results showed that the size of the TC effect varies with the lexical frequency, with stronger TC effects for high-frequency words. The influence of a higher degree of word-likeness of the high-frequency word on the transposition effect is not specific to alphabetic writing systems (e.g., English), as these effects were also observed in a nonalphabetic writing system (e.g., Chinese). This finding contributes to a deeper understanding of the important role played by lexical item in the character position encoding in Chinese reading development. Such a novel finding demonstrates how critical it is to consider issues of lexical properties, such as word frequency, when researchers are evaluating the developmental trajectories of TC effects in Chinese compound word recognition.

Acknowledgments. We sincerely thank Chang-Shou Lin (林彰收) and Chin-Yun Wu (吳錦雲) for their generous and invaluable assistance with the data collection. They are our heroes and inspiration. Without their efforts, this study could never have been conducted. Therefore, the article is dedicated to them.

Notes

1. Including age did not significantly improve the model fit of the data ($p = .07$). Therefore, age was removed from the model.

2. In the present study, the Kenward–Roger approximation (Kenward & Roger, 1997) was used to assess the significance of each fixed effect. The approximation is the default setting for the mixed() function in the afex package. It is based on a modified F test and also estimates the denominator degrees of freedom. It can provide the best control of Type I errors and Type II errors with limited sample sizes (see Judd et al., 2012). The Kenward–Roger approximation requires a model to be fitted with restricted maximum likelihood estimation (Bates & DebRoy, 2004).

References

Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology*, **99**, 245–264. doi:10.1348/000712607X224478

Anderson, S. E., Farmer, T. A., Goldstein, M., Schwade, J., & Spivey, M. (2011). Individual differences in measures of linguistic experience account for variability in the sentence processing skill of five-year-olds. In I. Arnon & E. V. Clark (Eds.). *Experience, variation, and generalization: Learning a first language* (pp. 203–221). Amsterdam: Benjamins.

Andrew, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language*, **35**, 775–800. doi:10.1006/jmla.1996.0040

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, **68**, 279–296. doi:10.1080/17470218.2014.1003949

Bai, C., Cai, S., & Schumacher, P. B. (2011). Reversibility in Chinese word formation influences target identification. *Neuroscience Letters*, **499**, 14–18. doi:10.1016/j.neulet.2011.05.020

Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS ONE*, **7**, e35932. doi:10.1371/journal.pone.0035932

Barca, L., & Pezzulo, G. (2015). Tracking second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS ONE*, **10**, e0116193. doi:10.1371/journal.pone.0116193

Barca, L., Pezzulo, G., Ouellet, M., & Ferrand, L. (2017). Dynamic lexical decisions in French: Evidence for a feedback inconsistency effect. *Acta psychologica*, **180**, 23–32. doi:10.1016/j.actpsy.2017.08.005

Bates, D., & DebRoy, S. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, **91**, 1–17. doi:10.1016/j.jmva.2004.04.013

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

Cargill, S. A., Farmer, T. A., Schwade, J. A., Goldstein, M. H., & Spivey, M. J. (2007). Children's online processing of complex sentences: New evidence from a new technique. *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 143–148), Mahwah, NJ: Erlbaum.

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). *Tracking the acquisition of orthographic skills in developing readers* (pp. 23–24). doi:10.1080/00049530701658626

Castles, A., Davis, C., & Forster, K. I. (2003). Word recognition development in children: Insights from masked priming. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: The state of the art* (pp. 345–360). Hove, UK: Psychology Press.

Castles, A., Davis, C., & Letcher, T. (1999). Neighborhood effects on masked form-priming in developing readers. *Language and Cognitive Processes*, **14**, 201–224. doi:10.1080/016909699386347

Cheng, C. M. (1981). Perception of Chinese characters. *Acta Psychologica Taiwanica*, **23**, 137–153.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, **35**, 983–996. doi:10.1111/j.1551-6709.2010.01164.x

De Rosario-Martínez, H. (2015). phia: Post-Hoc Interaction Analysis. Retrieved from http://CRAN.R-project.org/paclage=phia

De Zeeuw, M., Schreuder, R., & Verhoeven, L. (2015). Lexical processing of nominal compounds in first- and second-language learners across primary grades. *Writing Systems Research*, **7**, 133–156. doi:10.1080/17586801.2014.926806

Erb, C. D. (2018). The developing mind in action: Measuring manual dynamics in childhood. *Journal of Cognition and Development*, **19**, 233–247. doi:10.1080/15248372.2018.1454449

Erb, C. D., Moher, J., Sobel, D. M., & Song, J. H. (2016). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, **152**, 114–126. doi:10.1016/j.cognition.2016.03.015

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, **42**, 226–241. doi:10.3758/BRM.42.1.226

Giraudo, H., & Dal Maso, S. (2016). The salience of complex words and their parts: Which comes first? *Frontiers in Psychology*, **7**, 1778. doi:10.3389/fpsyg.2016.01778

Häikiö, T., Bertram, R., & Hyönä, J. (2011). The development of whole-word representations in compound word processing: Evidence from eye fixation patterns of elementary school children. *Applied Psycholinguistics*, **32**, 533–551. doi:10.1017/S0142716411000208

Hasenäcker, J., & Schroeder, S. (2019). Compound reading in German: Effects of constituent frequency and whole-word frequency in children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **45**, 920–933. doi:10.1037/xlm0000623

Hasenäcker, J., Schröter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **43**, 1093–1108. doi:10.1037/xlm0000353

Hermens, F. (2018). When do arrows start to compete? A developmental mouse-tracking study. *Acta Psychologica*, **182**, 177–188. doi:10.1016/j.actpsy.2017.11.015

Huang, H. S. (2001). *The Chinese character recognition test for school-age children* Taipei, Taiwan: Psychology Publishing.

Joseph, H. S. S. L., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, **42**, 207–222.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, **103**, 54–69. doi:10.1037/a0028347

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983–997. doi:10.2307/2533558

Krueger, B. I., Storkel, H. L., & Minai, U. (2018). The influence of misarticulations on children's word identification and processing. *Journal of Speech, Language, and Hearing Research*, **61**, 820–836. doi:10.1044/2017_JSLHR-S-16-0379

Lau, K.-Y (2012). *Compound word processing: Development and disorder*. Unpublished doctoral dissertation, University of Hong Kong, Hong Kong, China.

Leung, M. T., & Lee, A. W. Y. (2002). *The Hong Kong corpus of primary school Chinese*. Paper presented at the 9th Meeting of the International Clinical Phonetics and Linguistics Association, Hong Kong.

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

Lin, Y. C., Bangert, A. S., & Schwartz, A. I. (2015). The devil is in the details of hand movement: Visualizing transposed-letter effects in bilingual minds. *The Mental Lexicon*, **10**, 364–389. doi:10.1075/ml.10.3.03lin

Lin, Y. C., & Lin, P. Y. (2016). Mouse tracking traces the "Camrbidge Unievrsity" effects immonolingual and bilingual minds. *Acta Psychologica*, **167**, 52–62. doi:10.1016/j.actpsy.2016.04.001

Liu, P. D., Chung, K. K. H., McBride-Chang, C., & Tong, X. (2010). Holistic versus analytic processing: Evidence for a different approach to processing of Chinese at the word and character levels in Chinese children. *Journal of Experimental Child Psychology*, **107**, 466–478. doi:10.1016/j.jecp.2010.06.006

Liu, D., Chung, K. K. H., Zhang, Y., & Lu, Z. (2014). Sensitivity to the positional information of morphemes inside Chinese compound words and its relationship with word reading. *Reading and Writing*, **27**, 431–450. doi:10.1007/s11145-013-9451-6

Lu, L., & Liu, H. S. (1998). *The Peabody Picture Vocabulary Test–Revised in Chinese*. Taipei: Psychological Publishing.

Mattingly, I. G., & Xu, Y. (1994). Word superiority in Chinese. In H-W. Chang, J-T. Huang, C-W. Hue, & O. J. L. Tzeng (Eds.), *Advances in the study of Chinese language processing: Vol I. Selected writing from the 6th international symposium on cognitive aspects of the Chinese language* (pp. 101–111). Taipei: Department of Psychology, National Taiwan University.

Ministry of Education. (2002). *A list of commonly used characters and words for Taiwanese elementary school students*. Taipei, Taiwan: Author.

O'Connor, R. E., & Forster, K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory and Cognition*, **9**, 78–92.

Peng, D., Ding, G., Wang, C., Taft, M., & Zhu, X. (1999). Chinese reversible word's processing: The role of morpheme on word processing. *Acta Psychological Sinica*, **31**, 36–46.

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter confusability effects in masked associative priming. *Memory and Cognition*, **31**, 829–841. doi:10.3758/BF03196438

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, **51,** 231–246. doi:10.1016/j.jml.2004.05.005

Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception and Psychophysics*, **67**, 301–314. doi:10.3758/BF03206493

R Core Team. (2015). *R: A language and environment for statistical computing* (Version 3.2.2). Vienna, Austria: R Foundation for Statistical Computing. http://www.r-project.org/

Santens, S., & Verguts, T. (2011). The size congruity effect: Is bigger always more? *Cognition*, **118**, 94–110. doi:10.1016/j.cognition.2010.10.014

Schroeder, S., & Verrel, J. (2014). Cognitive processing and motor execution in the lexical decision task: A developmental study. *Psychonomic Bulletin & Review*, **21**, 496–504. doi:10.3758/s13423-013-0509-x

Singmann, H., Bolker, B., & Westfall, J. (2015). afex: Analysis of factorial experiments. R package, version 0.14-2.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, **57**, 745–765. doi:10.1080/02724980343000477

Taft, M., Zhu, X, & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language*, **40**, 498–519. doi:10.1006/jmla.1998.2625

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, **69**, 18–35. doi:10.1016/j.jml.2013.02.003

Tse, C. S., Yap, M. J., Chan, Y. L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*, **49**, 1503–1519. doi:10.3758/s13428-016-0810-5

Vergara-Martinez, M., Perea, M., Gómez, P., & Swaab, T. Y. (2013). ERP correlates of letter identity and letter position are modulated by lexical frequency. *Brain and Language*, **125**, 11–27.

Wang, Y., & McBride, C. (2016). Character reading and word reading in Chinese: unique correlates for Chinese kindergarteners. *Applied Psycholinguistics*, **37**, 371–386. doi:10.1017/S014271641500003X

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word beginning letters. *Journal of Experimental Psychology: Human Perception and Performance*, **34**, 1261–1276. doi:10.1037/0096-1523.34.5.1261

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, **10**, 545–600.

Appendix A. Full stimulus sets for the transposed-character and replaced-character conditions

Grade	Base word	Type	Transposed-character	Replaced-character
Grade 2	唱歌	High frequency	歌唱	放路
Grade 2	動物	High frequency	物動	副歡
Grade 2	學習	High frequency	習學	以修
Grade 2	天空	High frequency	空天	忙火
Grade 2	游泳	High frequency	泳游	遊現
Grade 2	朋友	High frequency	友朋	狀太
Grade 2	清楚	High frequency	楚清	黃注
Grade 2	醫生	High frequency	生醫	光事
Grade 2	拍手	Low frequency	手拍	動美
Grade 2	種植	Low frequency	植種	好作
Grade 2	郵票	Low frequency	票郵	舉事
Grade 2	黑雲	Low frequency	雲黑	當本
Grade 2	討論	Low frequency	論討	病助
Grade 2	音樂	Low frequency	樂音	緊後
Grade 2	親切	Low frequency	切親	開情
Grade 2	牙醫	Low frequency	醫牙	居生
Grade 2	外婆	Low frequency	婆外	手習
Grade 4	努力	High frequency	力努	焦怕
Grade 4	欣賞	High frequency	賞欣	及斷
Grade 4	害怕	High frequency	怕害	使喜
Grade 4	勇敢	High frequency	敢勇	鹿條
Grade 4	眼睛	High frequency	睛眼	行魔
Grade 4	熱鬧	High frequency	鬧熱	包求
Grade 4	鄰居	High frequency	居鄰	庭丹
Grade 4	幫忙	High frequency	忙幫	平香
Grade 4	聲音	High frequency	音聲	從色
Grade 4	禮物	High frequency	物禮	士物
Grade 4	告訴	High frequency	訴告	鄰丹
Grade 4	愉快	High frequency	快愉	評快
Grade 4	希望	High frequency	望希	藝正
Grade 4	貪玩	Low frequency	玩貪	優力
Grade 4	討論	Low frequency	論討	益變

(*Continued*)

https://doi.org/10.1017/S0142716420000181 Published online by Cambridge University Press

(*Continued*)

Grade	Base word	Type	Transposed-character	Replaced-character
Grade 4	平等	Low frequency	等平	安位
Grade 4	奇特	Low frequency	特奇	默河
Grade 4	原料	Low frequency	料原	康暖
Grade 4	緊急	Low frequency	緊急	主冷
Grade 4	油燈	Low frequency	燈油	息掃
Grade 4	演唱	Low frequency	唱演	勇書
Grade 4	真相	Low frequency	相真	高誕
Grade 4	文物	Low frequency	物文	相物
Grade 4	收集	Low frequency	集收	製雞
Grade 4	貪吃	Low frequency	吃貪	菜朋
Grade 4	旅遊	Low frequency	遊旅	權唱

Cite this article: Lin, Y.-C. and Lin, P.-Y. (2020). Reading minds in motion: Mouse tracking reveals transposed-character effects in Chinese compound word recognition. *Applied Psycholinguistics* **41**, 727–751. https://doi.org/10.1017/S0142716420000181