Behavior Research Methods (2023) 55:4382–4402
https://doi.org/10.3758/s13428-022-02022-z

The Chinese Lexicon Project II: A megastudy of speeded naming performance for 25,000+ traditional Chinese two-character words

Chi-Shing Tse[1,2] · Yuen-Lai Chan[1] · Melvin J. Yap[3] · Ho Chung Tsang[1]

Accepted: 6 November 2022 / Published online: 28 November 2022

Abstract
Using a megastudy approach, (Tse et al., 2017 *Behavior Research Methods*, 49, 1503–1519) established a large-scale repository of lexical variables and lexical decision responses for more than 25,000 traditional Chinese two-character words. In the current study, we expand their database by collecting norms for speeded naming reaction times (RTs) and accuracy rates, and compiling more lexical variables (e.g., phonological consistency and semantic neighborhood size). Following Tse et al.'s procedure, about 33 college-aged native Cantonese speakers in Hong Kong read aloud each word. We conducted item-level regression analyses to test the relative predictive power of orthographic variables (e.g., stroke count), phonological variables (e.g., phonological consistency), and semantic variables (e.g., semantic transparency) in naming performance. We also compared the effects of lexical variables on naming performance and Tse et al.'s lexical decision performance to examine the extent to which effects are task-specific or task-general. Freely accessible to the research community, this resource provides a valuable addition to other influential mega-databases, such as the English Lexicon Project (Balota et al., 2004 *Journal of Experimental Psychology: General*, 133, 283–316), and furthers our understanding of Chinese word recognition processes.

Keywords Chinese · Two-character word · Naming · Megastudy · Visual word recognition

In Chinese, characters are monosyllabic morphemes that are often combined to form a two-character word. According to the Institute of Language Teaching and Research (1986), 73.6% of modern Chinese words are two-character words (see Li et al., 2015, for more discussion), suggesting that the majority of Chinese words consist of two characters (Myers, 2006; Packard, 2000). Given this important feature of Chinese, Tse et al.'s (2017) megastudy focused on two-character words when developing a Chinese database of normed lexical decision (i.e., deciding whether a two-character string forms a Chinese word, e.g., 朋友friend, or a nonword, e.g., 形忌). In the present study, we expanded this database by norming speeded naming (i.e., reading aloud a Chinese word) performance, which has been considered one of the gold standards for assessing skilled readers' word recognition abilities (e.g., Yap & Rickard Liow, 2016) and for developing computational models of lexical processing (e.g., Coltheart et al., 2001). Moreover, we compiled more lexical variables (e.g., phonological consistency) and evaluated the role of orthographic, phonological, and semantic variables in naming versus Tse et al.'s lexical decision performance. In our speeded naming task, Cantonese, rather than Mandarin, was used, as it is the native dialect of our participant population (i.e., Hong Kong, as also in our previous megastudy, Tse et al., 2017). Before elaborating on the details of our study, we first briefly introduce the megastudy approach and then selectively review naming studies of Chinese characters/words to motivate the use of that approach to test the predictions presented at the end of the Introduction section.

Megastudy approach in psycholinguistics

Using the megastudy approach in psycholinguistic studies, researchers compile the values of lexical variables for a very large pool of words and collect the normative

✉ Chi-Shing Tse
cstse@cuhk.edu.hk

1 Department of Educational Psychology, The Chinese University of Hong Kong, Hong Kong, China

2 Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong, New Territories, China

3 Department of Psychology, National University of Singapore, Singapore, Singapore

reaction time (RT) and accuracy rates for these words in a lexical processing task. In contrast to the conventional factorial-design approach in which stimuli are selected based on a small set of criteria, the megastudy approach allows the language to define the stimuli (see Balota et al., 2013, and Tse et al., 2017, for reviews of problems associated with factorial-design experiments and how they can be addressed by the megastudy approach). After this, researchers conduct item-level multiple regression analyses on the normed data to test for the influence of lexical variables on word processing. The databases developed with this megastudy approach, often termed "Lexicon Project," have been reported in American English (Balota et al., 2007), British English (Keuleers et al., 2012), single-character Chinese (Chang et al., 2016; Liu et al., 2007; Sze et al., 2014), two-character Chinese (Tsang et al., 2018; Tse et al., 2017), French (Ferrand et al., 2010), Dutch (Keuleers et al., 2010), Italian (Barca et al., 2002), and Malay (Yap et al., 2010). The item-level analyses help identify the proportion of unique variance of performance that lexical variables, whether continuous or categorical, explain after statistically controlling for the effect of other variables, and reveal their relative contribution in lexical processing (e.g., Sze et al., 2015). These analyses can also address other research questions, such as replicating theoretically important lexico-semantic effects (Balota et al., 2013), evaluating the impact of novel lexical variables (Kang et al., 2011; Yarkoni et al., 2008), assessing computational models (Perry et al., 2007), and revealing the role of individual differences in word recognition (Lim et al., 2020; Yap et al., 2012).

Tse et al. (2017) used the megastudy approach to establish a lexical decision database of RT and accuracy rates for more than 25,000 Chinese two-character words. Each of 594 native Cantonese-speaking participants made lexical decisions for about 1400 words and 1400 nonwords, presented over three sessions. Each word was normed based on about 33 responses. Item-level regression analyses were conducted to determine the influence of theoretically important lexical characteristics on lexical decision performance. Making use of these data, Tse and Yap (2018) compared the role of orthographic, phonological, and semantic variables in predicting lexical decision performance and found some theoretically driven interaction effects (e.g., word frequency × character frequency) that are not readily explainable by any model of Chinese word recognition. Lim et al. (2020) further demonstrated that the effects of some lexical variables (e.g., word frequency) on lexical decision could be moderated by participants' lexical processing fluency. All these demonstrate the usefulness of a megastudy database for investigating the effects of lexical variables on performance in a lexical processing task.

Why investigate speeded naming in Chinese?

In the speeded naming task, participants read aloud a visually presented word as quickly and as accurately as they can. In the present study, we extended Tse et al.'s (2017) database by collecting participants' normed naming responses for their large pool of 25,000+ two-character Chinese words. We also conducted analyses on both the normed lexical decision data and the naming data in Tse et al.'s database to test whether their reported lexical effects might reflect general word recognition processes or merely idiosyncratic task demands.

Lexical effect on speeded naming performance As few studies have examined the speeded naming of two-character Chinese words, in the following review we also refer to the work reported on the speeded naming of single Chinese characters. For orthographic variables, Yu and Cao (1992, see also Liu et al., 2007) found that high-frequency characters (e.g., 花 flower) were named faster than low-frequency characters (e.g., 埅 compliance). Gao et al. (2016) reported that high-frequency words (e.g., 多謝 thank you) were named faster than low-frequency words (e.g., 睿哲 divinely wise). Number of strokes is often used to indicate visual complexity (e.g., Xing et al., 2004), wherein characters with more strokes are considered visually more complex (e.g., 人 human vs. 鬱 depressed). Shen and Zhu (1994, see also Leong et al., 1987; Liu et al., 2007) demonstrated that characters with fewer strokes were named faster than those with more strokes.

For phonological variables, Hue et al. (1992, see also Liu et al., 2007) showed that phonologically regular characters (a character is pronounced the same as its phonetic radical, e.g., 叮ding1[1] and 丁ding1) were named faster than irregular characters (a character and its phonetic radical do not share the pronunciation, e.g., 模mou4 and 莫mok6). Orthography-to-phonology mapping consistency refers to whether the mapping is systematic between a character's phonetic radical and its pronunciation. For example, 換 wun6 is a highly consistent character, as all other characters sharing 奐 as their phonetic radical are also pronounced as wun6, e.g., 喚 and 渙. In contrast, 板baan2 is less consistent, because characters with 反 as their phonetic radicals are pronounced differently, from 皈gwai1 to 飯faan6. Yang et al. (2009) showed that characters consistent with orthography-to-phonology mapping (e.g., 渙thaw out) were named faster

[1] All phonological transcriptions/syllables are Cantonese pronunciations based on the Multi-function Chinese Character Database (see https://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/ or https://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/ for more details).

than inconsistent characters (e.g., 飯rice). Phonological consistency refers to whether a character has one (consistent, e.g., 骨*gwat1*) or more than one (inconsistent, e.g., 重*zung6* and *cung*5) pronunciation (Tan & Perfetti, 1999). Leong and Cheng (2003) found that words with consistent characters (e.g., 骨 in 骨頭 bone) were named faster than those with inconsistent characters (e.g., 重 in 體重 weight), whether the character was the first or second in the word. Finally, characters that share pronunciations with more other characters (i.e., high homophone density, e.g., 支) were named faster than those that share pronunciations with fewer other characters (i.e., low homophone density, e.g., 鏡) (e.g., Ziegler et al., 2000, but see the null effect of homophone density in Chang et al., 2016).

For semantic variables, characters that combine more readily with another character to form a word (i.e., larger neighborhood size, e.g., 花園 garden, 花盆 flowerpot, 花店 florist, 花叢 flower bush, 花市 flower market, 花椒 wild pepper, 花蜜 honey, 開花 to blossom, 花錢 to spend money, 花粉 pollen, 鮮花 fresh flower, etc., for 花) were named faster than those not as likely (i.e., smaller neighborhood size, e.g., only 鳳凰 phoenix for 凰) (e.g., Chang et al., 2016; Liu et al., 2007, see Baayen et al., 2006 for similar findings in English). Peng et al. (2003) reported that characters with more meanings (e.g., 行) were named faster than those with fewer meanings (e.g., 廚). Semantic transparency refers to whether words are semantically related (transparent, e.g., 黑板black-board [blackboard]) or unrelated to their characters (opaque, e.g., 東西east–west [thing]). It can be quantified by participants' ratings for the semantic relatedness between a word and its characters (e.g., Tse et al., 2017). Although words with more semantically transparent characters yielded faster lexical decision performance than those with less semantically transparent characters (e.g., Myers et al., 2004; Tse et al., 2017), to our knowledge this variable has not been examined in the two-character Chinese word naming literature. In other languages like English, transparent words were reported to yield better lexical decision and naming performance than opaque words (e.g., Kim et al., 2019).

Most of the previous findings in Chinese character/word speeded naming have been based on factorial-design experiments, wherein a relatively small set of words constrained to be matched on various lexical characteristics were used. It is important to test whether these effects of lexical variables occur when a much larger pool of words is included in the megastudy approach. Assuming that the findings of naming single characters can be generalized to naming two-character words, we expect to replicate most, if not all, of the above findings.

The role of character variables in word processing There has been debate on whether Chinese words are accessed holistically or analytically (by combining the characters) in the literature. Packard (1999) has argued that the access is via the whole word, as the large number of homophones and extent of semantic ambiguity in Chinese might make character processing too costly in terms of time and effort. However, this holistic view might be incompatible with findings that character variables, such as character frequency, in addition to word variables (e.g., word frequency), account for variance in lexical decision performance (e.g., Tse et al., 2017, see also the evidence of priming studies, e.g., Tsang & Chen, 2014). Moreover, Tse et al. showed that the effects of lexical variables on lexical decision performance were quite similar whether the characteristic was associated with the first or second character, suggesting that word processing may be parallel, with both characters being processed simultaneously. That said, it is not clear whether this parallel processing also occurs in naming. Specifically, the speeded naming task requires participants to read aloud the first and then second characters, and the voice key is triggered to record naming RT once they begin to pronounce the first character (i.e., speech onset). Unlike in English, where participants have to process the entire letter string to correctly pronounce the word, participants may not need to completely process the whole Chinese word before naming the first character of the word.[2] If so, we expect that character variables, especially those of the first character, would be more predictive of naming performance than word variables, suggesting that the holistic versus analytic nature of Chinese word processing depends on task demands. To our knowledge, this issue has not been addressed in factorial-design experiments. The megastudy approach with multiple regression analyses could compare the predictive power of the lexical variables of the first and second characters so as to shed light on this issue.

The role of orthographic, phonological, and semantic variables in word processing In contrast to the English language, an alphabetic writing system, Chinese language possesses a logographic system with far more opaque orthography; that is, there is no regular mapping between spelling and sound. While the phonetic radical of about 80% of Chinese characters provides pronunciation hints (Hoosain, 1992), this depends on orthography-to-phonology mapping consistency (i.e., whether a character's pronunciation is shared with other

[2] It is possible that when the word consists of at least one phonologically inconsistent character (i.e., character that can have more than one pronunciation), people might then need to process both characters in order to assign the correct pronunciation of that phonologically inconsistent character. For example, 重can be pronounced as cung4, cung5, or zung6. When pronouncing 重要, 重has to be pronounced as zung6. Thus, people need to process both characters 重and 要of the word 重要, in order to assign correct pronunciations of 重 (i.e., zung6, but not cung4 or cung5).

Chinese characters possessing the same phonetic radical). Hence, it is possible that phonology plays a less influential role in Chinese, compared with English, word naming. On the other hand, Perfetti and Tan (1999) argued that Chinese word recognition results from a convergence of orthography, phonology, and semantics. Because orthography-to-phonology mapping (often in one-to-one relationship, e.g., 表 pronounced as biu2) is more reliable than orthography-to-semantics mapping (often in one-to-many relationship, e.g., 表 can mean watch, express, surface, or meter), phonology is the privileged constituent in word recognition. Inconsistent with this view, based on Tse et al.'s (2017) normed lexical decision data, Tse and Yap (2018, see also Sze et al., 2015) reported that orthographic and semantic variables, respectively, accounted for more variance than phonological variables. This suggests that phonological activation, compared with semantic activation, is less influential in a task which does not explicitly require the generation of word phonology and instead emphasizes the processing of stimulus familiarity (e.g., Balota & Chumbley, 1984). Evidence supporting the role of phonological information in Chinese word recognition has often been based on the speeded naming task (e.g., Perfetti & Tan, 1999), which places more emphasis on the generation of phonology. By identifying the proportion of variance in naming performance accounted for by orthographic, phonological, and semantic variables, we expect that phonological variables play a much larger role than semantic variables in naming performance.

In the few studies that have explored task-specific differences between Chinese lexical decision and naming (e.g., Gao et al., 2016; Li et al., 2017), word frequency was reported to produce a larger effect in the lexical decision task. Given the demands of speeded naming, when activating the whole word, participants might also sequentially activate the phonological information for the first and then the second character (e.g., Li et al.). This contrasts with the lexical decision task, in which participants can respond only by processing the whole word. The first-character-focused naming process might somehow play a role in reducing the word frequency effect in Chinese lexical processing. This is consistent with the view that, relative to naming, lexical decision emphasizes more frequency-based information in discriminating between word and nonword (e.g., Balota & Chumbley, 1984). In English, studies reported task-specific effects in lexical decision and naming (e.g., Andrews & Heathcote, 2001; Balota & Chumbley, 1985; Balota et al., 2004). For instance, there was an influence of semantic variables in lexical decision (e.g., Yap & Balota, 2009), but semantic effects in naming were smaller (as compared with lexical decision, e.g., Balota et al., 2004) or restricted to low-frequency irregular words (e.g., Cortese et al., 1997), so semantic variables might have a greater influence on lexical decision than on naming.

Most of the findings in Chinese have been based on factorial-design experiments and should thus be tested for generalizability using the megastudy approach (see, e.g., Balota et al., 2004, for similar analyses in English). By conducting the same analyses on the current naming data and Tse et al.'s lexical decision data, we compare the influences of these variables between the two tasks. We expect phonological variables to play a larger role in naming than in lexical decision, and semantic variables to play a larger role in lexical decision than in naming. For example, the first-character frequency effect would be larger in naming than in lexical decision, whereas the word frequency effect would be larger in lexical decision than in naming.

Present study

To recapitulate, we expanded Tse et al.'s (2017) database by norming speeded naming measures for more than 25,000 two-character Chinese words and compiling more lexical variables to test the role of orthographic, phonological, and semantic variables (see Table 1) in naming performance and to compare those with lexical decision performance in the same item-level multiple regression analyses. We have the following predictions.

(1) We replicate the lexical effects (e.g., word frequency and phonological consistency) of previous speeded naming studies in Chinese, even though most of them were based on single-character naming factorial-design experiments.
(2) Given that participants serially activate individual character phonology when they begin to pronounce the two-character words, character variables (e.g., first-character frequency) will be more predictive of naming performance than word variables (e.g., word frequency). The first-character lexical variables will also be more predictive of naming performance than the second-character lexical variables.
(3) By comparing lexical decision (Tse et al., 2017) and naming based on the proportion of variance explained in multiple regression analyses, we expect that phonological variables play a larger role than semantic variables in naming, and semantic variables play a larger role than phonological variables in lexical decision (i.e., similar to Tse & Yap's, 2018, findings, based on Tse et al.'s, 2017, normed lexical decision data).

Table 1 Variables listed in the Excel .xlsx and Unicode .csv files

Column	Variable name	Definition
1	Word_Trad	Chinese word in traditional characters
2	Word_Sim	Chinese word in simplified characters
3	Ntrials	Number of participants whose trials were sufficiently reliable to provide the RT for that item (maximum being 33)
4	Acc	Mean accuracy rate for each word computed across participants
5	RT	Mean RT for each word computed across participants
6	RT_SD	Standard deviation of the RT for each word
7	zRT	Mean of the standardized RT for each word computed across participants
8–9	Pronounciation_1 and Pronounciation_2	Phonological transcriptions/syllables (with lexical tone) of the first and second characters are Cantonese Pronunciations based on the Multi-function Chinese Character Database (see https://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/ or https://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/ for more details)
10–22	Initial_phoneme_feature_1, …, Initial_phoneme_feature_13	The 13 dichotomous variables to code the initial phoneme of the first character on 13 features, in the following order: affricative, alveolar, bilabial, labiodentals, dental, fricative, glottal, liquid, stop, nasal, palatal, velar, and voiced (e.g., Spieler & Balota, 1997).
23–24	Stroke_1 and Stroke_2	Number of strokes of the first and second characters based on a pocket dictionary (Que, 2008)
25–26	Phoneme_1 and Phoneme_2	Number of phonemes of the first and second characters
27–28	Character_frequency_1 and Character_frequency_2	Character frequency of the first and second characters based on Cai and Brysbaert's (2010) contextual diversity subtitle character frequency count
29	Word_frequency	Word frequency based on Cai and Brysbaert's (2010) contextual diversity subtitle word frequency
30–31	Regularity_1_dummy_contrast_1 and Regularity_2_dummy_contrast_1	Whether the character is pronounced the same as (i.e., regular) or different from (i.e., irregular) its phonetic radical, e.g., 叮 is regular, as it shares the same pronunciation as 丁. For "change of tone," the character shares the same syllable but not the same tone with its phonetic radical (e.g., regularity of 倩cing4 vs. 青cing1). Regular = 1, irregular = −0.5, change of tone = −0.5
32–33	Regularity_1_dummy_contrast_2 and Regularity_2_dummy_contrast_2	See above. Regular = 0, irregular = 0.5, change of tone = −0.5
34–35	OP_Consistency_1 and OP_Consistency_2	Orthography-to-phonology mapping consistency: the extent to which the mapping is systematic between a character's phonetic radical and its pronunciation. For example, 換 wun6 is a character with high consistency, because all other characters sharing 奐 as their phonetic radical are also pronounced as wun6, e.g., 喚 and 渙. In contrast, 板 baan2 is less consistent, because characters with 反 as their phonetic radicals are pronounced differently, from 皈 gwai1 and 飯 faan6.
36–37	Phono_consistency_1 and Phono_consistency_2	Phonological consistency: whether a character has one (phonologically consistent) or more than one (phonologically inconsistent) pronunciation (as defined by Leong & Cheng, 2003)
38–39	Homophone_1 and Homophone_2	Homophone density of the first and second character: number of characters that share the same pronunciation provided in the Multi-function Chinese Character Database (http://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/)
40–41	Neighborhood_1 and Neighborhood_2	Neighborhood size of the first and second character: number of two-character words sharing the same character, regardless of the character position (Tsai et al., 2006)
42–43	Meaning_1 and Meaning_2	Number of meanings of the first and second characters: raw count of distinct conceptual representations a character has, i.e., how polysemous a character is (Tsang et al., 2018)
44–45	Transparency_1 and Transparency_2	Semantic transparency between first/second characters and word based on standardized means of participants' semantic relatedness ratings for the first/second characters in respect to the word

Method

To ensure comparability, the participants, material, design, and procedure followed those in our lexical decision megastudy (Tse et al., 2017).

Participants

Two hundred and ninety-seven native Cantonese-speaking students from the Chinese University of Hong Kong (CUHK), with self-reported normal or corrected-to-normal vision, participated in our study. Data from 19 additional participants were replaced due to their failure to participate in all experimental sessions. Participants were asked to report their gender (191 female), age ($M = 19.73$, $SD = 1.43$), amount of time they read Chinese materials per week ($M = 210.43$ minutes, $SD = 460.82$), cumulative grade-point average ($M = 3.23$, $SD = .34$),[3] self-rated knowledge in seven-point scales for traditional characters ($M = 5.63$, $SD = .88$) and spoken Cantonese ($M = 5.99$, $SD = .90$), and overall grades and sub-grades in Chinese language university entrance exam (overall: $M = 4.64$, $SD = 1.22$; reading comprehension: $M = 4.82$, $SD = 1.52$; writing: $M = 4.28$, $SD = 1.56$; listening and integrated skills: $M = 5.19$, $SD = 1.41$; speaking: $M = 3.88$, $SD = 1.39$).[4] We did not impose any restriction on participants' Chinese proficiency, as substantial variability in proficiency may allow an exploration of individual differences in language processing (e.g., Lim et al., 2020, see also Yap et al., 2012). Each participant was paid 300 HKD (~38 USD) for their voluntary participation.

Materials and design

We include the words that were normed for lexical decision performance in the Chinese Lexicon Project (see Tse et al., 2017, for the selection procedure). Stimuli were all traditional Chinese characters typically used in Hong Kong and should thus be familiar to participants. The lexical variables are listed in Table 1 (see Tse et al., 2017s, for how semantic transparency was normed). All values of these variables are listed in the .xlsx/.csv file available at: https://osf.io/vwnps. Using Tse et al.'s procedure (see also the English Lexicon Project of Balota et al., 2007), we divided 25,281 two-character Chinese words into nine lists of 2809 words each, with a restriction that the proportion of the words with a specific first character in each subset roughly reflected those in the whole word pool. We collected data from 297 participants (i.e., 9 lists × 33 participants) to obtain 33 naming responses for each word. This number of observations per word is comparable to the mean of other normed naming databases of other languages (e.g., 25 in Balota et al.; > 25 in Ferrand et al., 2010; 31 in Goh et al., 2020; 39 in Keuleers et al., 2010; 39 in Keuleers et al., 2012) and allows us to obtain stable estimates of each word's mean RTs. Each participant read aloud 936–937 words in each of the three sessions and thus in total 2809 words. The presentation order of word sets across the three sessions were counterbalanced across participants.

Procedure

Informed consent was obtained at the beginning of the study. Each participant was tested in a quiet cubicle in three sessions separated by no more than a week. PC-compatible computers with E-Prime 2.0 (Schneider et al., 2001) were used to display the stimuli and collect RT and accuracy data. (The participants' naming duration was also recorded, but we do not consider them in the current work.) Stimuli were in white and visually presented on a black background, one at a time, at the center of the screen. In each session, 936–937 words were divided into four blocks of 234–235 words. At the end of each block, participants received a self-paced rest break, so there were three rest breaks in each session. Participants took about 55 min to complete each session. The presentation order of the four blocks within each session were randomized for each participant. Within each block, the presentation order of the words was also randomized for each participant.

Each trial started with a 500-ms fixation point *** at the center of the screen, followed by a 120-ms blank-screen interstimulus interval. A two-character Chinese word in font size 36 and font type 標楷體 DFKai-SB was presented at the same location as the fixation point until the participant made a response. The physical size of the word was about 2.5 cm in height by 5 cm in width. Given that participants sat 55 cm away from the screen, the visual angle was about 1.56°. To ensure that the current data reflected the young adults' naming responses to the script that was most familiar to them, we presented all stimuli in traditional characters, following Tse et al. (2017). The participant was instructed to read aloud the word in Cantonese (i.e., their native language) as quickly and as accurately as possible to trigger the voice key connected via the Chronos response box to the

[3] The mean and SD of cumulative grade-point average (max = 4) are based on 233 participants who were not first-year students and/or able to provide this information.

[4] All but one of our participants (i.e., $N = 296$) took the Hong Kong Diploma of Secondary Education Examination (HKDSE) for university entrance, in which their performance is classified by a seven-point scale (5**, 5*, 5, 4, 3, 2, and 1, with the last two being the failing grades). We converted their top two grades 5** and 5* into 7 and 6, respectively, for their overall grades and sub-grades in the Chinese language university entrance exam. The overall grades were based on 294 participants, as two participants did not report that. The speaking sub-grades were based on 194 participants, because this subtest was canceled in 2020 and 2021 due to the COVID-19 pandemic.

computer. The Chronset algorithm (Roux, Armstrong, & Carreiras, 2016) was used to ensure the precision of onset detection. The word remained on the screen for an additional 500 ms after voice onset and then disappeared. If no response was detected in 4000 ms, the word disappeared and then a verbal signal 太慢!Too Slow! appeared for 500 ms. At the end of each trial, a blank screen was displayed for 1000 ms, to serve as an intertrial interval. Following Balota et al.'s (2007) procedure, before the 1000-ms intertrial interval, participants were instructed to manually code, in a self-paced manner, their own responses as correct pronunciation, mispronunciation, or microphone error. Participants were informed about the importance and use of these coding options. At the start of a session, participants were presented 10 practice words to become familiar with the task. These words did not appear as experimental words in that session.

Results

A phonetic bias might occur for an articulatory reason (some phonemes take more time to initiate) or acoustic reason (some phonemes take more time for the voice key to detect) (e.g., Liu et al., 2007). Following previous studies (e.g., Balota et al., 2004), we controlled for these variables in our analyses by coding the following 13 characteristics of the initial phoneme of the first characters as 1 (presence) or 0 (absence): affricative, alveolar, bilabial, labiodentals, dental, fricative, glottal, liquid, stop, nasal, palatal, velar, and voiced.

Only responses for experimental words were analyzed. Based on 297 participants, the mean accuracy rate (i.e., percentage of correct responses after excluding trials with microphone error) was 96.40% ($SD = 2.41\%$). Following previous studies (e.g., Sze et al., 2014; Tse et al., 2017), remaining responses faster than 200 ms or slower than 3000 ms were excluded (about 0.41% of the trials). The mean and *SD* were then computed for each participant's word responses. Any correct word response 2.5 *SD* above or below their mean was regarded as an outlier and excluded (about 3.31% of the trials). The mean RT of the trimmed correct word trials was 659.53 ms ($SD = 186.71$ ms). As suggested by Faust et al. (1999) and following previous megastudies (e.g., Balota et al., 2007; Ferrand et al., 2010; Ferrand et al., 2018; Keuleers et al., 2012; Sze et al., 2014; Tse et al., 2017; Yap et al., 2010), these raw RTs were transformed into *z* scores for each participant, before averaging across participants for each word to yield individual word zRT values. This standardization controls for differences in overall RT and variability between participants, which then helps to reduce noise in the data. Previous works reported that zRT was more reliable than raw RT (e.g., Ferrand et al., 2010). The level of significance was set at .05.

To determine the reliability of our dependent measures (RT, zRT, and accuracy rates), we followed Tse et al. (2017) and computed the split-half correlation, with a correction for length (i.e., about 33 observations per word) with the Spearman-Brown formula $(2 \times r)/(1 + r)$. This reflects the proportion of variance in a variable that can be explained. The corrected correlations (r_{corr}) between the dependent measures computed on the first half ($N = 16$) of participants who saw the word and the dependent measures computed on the second half ($N = 17$) of participants who saw the word were .82, .88, and .86 for RT, zRT, and accuracy, respectively. The finding that the reliability of zRTs is higher than the reliability of the raw RTs is in line with Faust et al.'s (1999) view that taking away differences in overall RT and variability between participants may remove noise from the data and does not artificially reduce the variability of the words.

Following Tse et al. (2017), we excluded words that yielded lower than 70% accuracy rates (about 2.31%). We ran item-level multiple regression analyses analogous to those that Tse et al. did for the lexical decision data, with mean zRT, averaged across participants, as the dependent variable. For the accuracy data,[5] since the values are bounded (from 0 to 100) and violate the assumptions of linear regression models (i.e., without boundaries), we first transformed each word's mean accuracy rates to logit (i.e., unbounded score) and then ran the same item-level multiple regression analysis as that for mean zRT, using the binomial family under the generalized linear modeling function in R 4.1.3. Note that McFadden's pseudo R^2 values were computed for these accuracy analyses, which could not be directly compared with the R^2 values for zRT. The values of the lexical variables as predictors were all mean-centered in all regression analyses (Jaccard & Turrisi, 2003). Table 1 presents all variables listed in the Excel .xlsx and Unicode .csv files, which are available at: https://osf.io/vwnps (Tse_et_al.xlsx and Tse_et_al.csv). Following previous megastudies (e.g., Tse et al., 2017), we entered the lexical variables in the following order: characteristics of initial phoneme of the first character (step 1), orthographic variables (number of strokes, the frequency of the first and second characters, and word frequency) (step 2), phonological variables (number of phonemes, homophone density, phonological consistency, phonological regularity,[6] and orthography-to-phonology mapping consistency of the first and second characters) (step 3), and semantic variables (neighborhood size, number of meanings, and semantic transparency of the first and second characters) (step 4). After taking into account the missing values of some words, the following analyses were based on 19,888 words (about 78.67% of 25,281 words in total).

[5] We thank an anonymous reviewer for this suggestion.

[6] Phonological regularity is a three-level variable, so it is coded as two dummy variables (regular = 1 0, irregular = −0.5 0.5, change of tone = −0.5 to 0.5; i.e., the character shares the same syllable but not the same tone with its phonetic radical, e.g., regularity of 情cing4 vs. 青cing1).

Table 2 Descriptive statistics of the lexical variables and two behavioral measures (zRT and accuracy) involved in item-level regression analyses

	Mean	*SD*	Range
Naming accuracy	.98	.04	.70–1.00
Naming zRT	−.036	.438	−1.02 to 2.54
Number of strokes C1	10.61	4.43	1–30
Number of strokes C2	10.62	4.48	1–33
Log frequency C1	3.36	.49	.00–3.80
Log frequency C2	3.41	.47	.00–3.80
Log frequency (word)	1.58	.81	.00–3.80
Number of phonemes C1	3.35	.72	1–6
Number of phonemes C2	3.32	.73	2–6
Homophone density C1	17.61	15.75	0–106
Homophone density C2	17.67	15.62	0–106
Phonological consistency C1	.59	.49	0–1
Phonological consistency C2	.58	.49	0–1
Phonological regularity C1 (dummy contrast 1)	−.29	.52	−0.5 to 1
Phonological regularity C1 (dummy contrast 2)	.33	.32	−0.5 to 0.5
Phonological regularity C2 (dummy contrast 1)	−.27	.54	−0.5 to 1
Phonological regularity C2 (dummy contrast 2)	.33	.32	−0.5 to 0.5
Orthography-to-phonology mapping consistency C1	.16	.37	0–1
Orthography-to-phonology mapping consistency C2	.17	.37	0–1
Neighborhood size C1	38.15	33.88	1–229
Neighborhood size C2	44.43	40.46	1–229
Number of meanings C1	5.16	3.27	1–21
Number of meanings C2	5.28	3.27	1–21
Semantic transparency C1	−.0089	.49	−2.14 to 1.61
Semantic transparency C2	−.0069	.51	−2.24 to 1.81

N = 19,888 (i.e., words with values in all available lexical variables). C1 = first character. C2 = second character. The three-level variable, phonological regularity, is coded as two dummy contrasts (regular = 1 0, irregular = −0.5 0.5, change of tone = −0.5 0.5)

Analyses of naming performance

Tables 2 and 3 report the descriptive statistics of and inter-correlational matrices among these lexical variables and the two behavioral measures (zRT and accuracy). There was no multicollinearity problem for orthographic, phonological, and semantic variables, as indicated by their overall moderate- to low inter-correlations (see Table 3). The variance inflation factors were also generally low (< 1.94) in all these variables.[7] The betas (standardized regression coefficients) of these lexical variables when they were first entered in the regression models and the R^2 change (i.e., change in proportion of variance accounted for in zRT) or pseudo R^2 change in accuracy at each step are reported in Table 4. The total proportions of variance accounted for were 51.6% and 19.1% for zRT and accuracy, respectively.

Variance accounted for in naming zRT and accuracy First, characteristics of the initial phoneme of the first character accounted for 10.3% and 0.2% of the variance in naming zRT and accuracy, respectively, suggesting that phonological coding at the first character's phoneme level is a significant predictor of naming zRT, but much less so naming accuracy. This reflects that the coding of onsets may primarily influence response speed due to voice key sensitivity and articulation, rather than response accuracy. Thus, they were statistically controlled in the following analyses. Second, phonological variables accounted for much less variance in naming zRT than orthographic variables (2.1% vs. 37.9%) and slightly more variance in naming zRT than semantic variables (2.1% vs. 1.3%). A similar pattern was observed in naming accuracy. The modest amount of variance accounted

[7] While some of the characteristics of the initial phoneme of the first character show relatively high variance inflation factors, after excluding those with higher than 10 in the analyses (i.e., affricative, alveolar, fricative, stop, and voiced), we obtained similar findings for all orthographic, phonological, and semantic variables reported in the main text. The only exception was that the second character's phonological regularity dummy contrast (0 = regular, 0.5 = irregular, −0.5 = change of tone) approached significance (beta = .012), suggesting that zRT was slower when the second character was phonologically irregular (vs. change of tone). In short, we do not consider that the high variance inflation ratio in five of the first character's initial phoneme variables distorted the pattern of major findings.

Table 3 Inter-correlational matrices among these non-centered lexical variables and the two behavioral measures (zRT and accuracy) involved in item-level regression analyses ($N = 19{,}888$)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	Accuracy	1																								
2	zRT	−.490**	1																							
3	Number of strokes C1	−.082**	.251**	1																						
4	Number of strokes C2	−.058**	.139**	.019**	1																					
5	Log frequency C1	.307**	−.499**	−.262**	−.019**	1																				
6	Log frequency C2	.266**	−.318**	−.010	−.267**	.151**	1																			
7	Log frequency (word)	.194**	−.330**	−.057**	−.070**	.245**	.228**	1																		
8	Number of phonemes C1	.004	−.112**	.074**	.002	−.002	−.004	−.016*	1																	
9	Number of phonemes C2	−.018*	.029**	.010	.116**	.003	−.063**	−.042**	.021**	1																
10	Homophone density C1	.016*	.081**	.023**	−.001	−.090**	−.013	−.010	−.151**	−.002	1															
11	Homophone density C2	.008	.010	−.008	−.024**	−.026**	−.037**	−.011	−.016*	−.217**	.025**	1														
12	Phonological consistency C1	.041**	.060**	.062**	.018**	−.099**	−.031**	−.017*	−.095**	−.003	.064**	.010	1													
13	Phonological consistency C2	.045**	.025**	.003	.091**	−.010	−.117**	−.043**	.000	−.024**	−.004	.066**	.022**	1												
14	O-to-P mapping consistency C1	.001	.006	.022**	.005	−.035**	−.006	−.013	.015*	.006	.057**	.000	.027**	.004	1											
15	O-to-P mapping consistency C2	−.002	.014*	.001	−.017*	−.011	−.009	−.001	.000	.009	.008	.076**	.010	.028**	.006	1										
16	Phonological regularity (contrast 1) C1	−.039**	.092**	.201**	.022**	−.205**	−.050**	−.029**	−.017*	.003	.197**	.006	.071**	−.003	.129**	.006	1									
17	Phonological regularity (contrast 2) C1	.030**	−.078**	−.256**	−.010	.158**	.019**	.024**	−.037**	−.001	−.062**	−.003	−.019**	.005	−.038**	.002	−.423**	1								

Table 3 (continued)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
18	Phonological regularity (contrast 1) C2	−.037**	.069**	.007	.230**	−.042**	−.219**	−.043**	.004	.016*	.007	.159**	.014*	.112**	.005	.114**	.029**	−.018*	1							
19	Phonological regularity (contrast 2) C2	.018*	−.042**	−.002	−.263**	.021**	.153**	.033**	−.013	−.075**	−.001	−.015*	−.006	−.048**	−.011	−.027**	−.016*	.024**	−.436**	1						
20	Neighborhood size C1	.156**	−.348**	−.346**	.002	.562**	.040**	.108**	.062**	.022**	−.124**	.000	−.097**	.009	.001	−.002	−.188**	.185**	.000	.005	1					
21	Neighborhood size C2	.092**	−.201**	.006	−.385**	.017*	.538**	.111**	.001	−.091**	−.016*	.005	−.015*	−.137**	−.001	−.015*	−.016*	.003	−.190**	.193**	−.004	1				
22	Number of meanings C1	.115**	−.295**	−.242**	−.004	.477**	.045**	.093**	.087**	.010	−.111**	−.024**	−.230**	−.008	.003	−.005	−.169**	.169**	−.020**	.017*	.570**	.009	1			
23	Number of meanings C2	.080**	−.164**	−.015*	−.249**	.045**	.472**	.101**	.010	−.022**	−.010	−.052**	−.025**	−.251**	.002	−.006	−.020**	.013	−.178**	.171**	.008	.567**	.040**	1		
24	Semantic transparency C1	−.056**	.161**	.207**	−.028**	−.238**	−.010	−.016*	−.003	.011	.036**	.013	.092**	−.009	.010	.019**	.073**	−.089**	.017*	.001	−.223**	.069**	−.330**	.031**	1	
25	Semantic transparency C2	−.060**	.146**	.031**	.214**	−.076**	−.271**	−.121**	.025**	.074**	.014*	−.053**	.017*	.104**	.008	.004	.021**	−.002	.062**	−.103**	−.056**	−.327**	−.031**	−.365**	.069**	1

*$p < .05$ (two-tailed). **$p < .01$ (two-tailed). C1 = first character. C2 = second character. O-to-P = orthography-to-phonology

Table 4 Standardized regression coefficients (beta values) on naming zRT and accuracy rate ($N = 19{,}888$)

	zRT	Accuracy
First character's initial phoneme - affricative	0.182**	0.121
First character's initial phoneme - alveolar	NA1	0.113
First character's initial phoneme - bilabial	−0.032	0.198
First character's initial phoneme - labiodentals	−0.069**	0.004
First character's initial phoneme - dental	−0.005	−0.006
First character's initial phoneme - fricative	0.240**	0.174
First character's initial phoneme - glottal	−0.151**	NA2
First character's initial phoneme - liquid	−0.013	0.129
First character's initial phoneme - stop	−0.050	0.157
First character's initial phoneme - nasal	−0.053**	0.151
First character's initial phoneme - palatal	−0.011	0.477
First character's initial phoneme - velar	0.052**	0.132
First character's initial phoneme - voiced	−0.072	−0.138
Step 1 - R^2 change	0.103**	0.002
Number of strokes C1	0.152**	−0.011
Number of strokes C2	0.064**	< 0.001
Log frequency C1	−0.398**	0.642**
Log frequency C2	−0.207**	0.558**
Log frequency (word)	−0.177**	0.289**
Step 2 - R^2 change	0.379**	0.170
Number of phonemes C1	−0.143**	0.060
Number of phonemes C2	0.004	0.003
Homophone density C1	0.020**	0.005
Homophone density C2	−0.007	0.002
Phonological consistency C1 (1 = consistent)	−0.035**	0.266**
Phonological consistency C2 (1 = consistent)	−0.017**	0.272**
Orthography-to-phonology mapping consistency C1 (1 = consistent)	0.003	0.038
Orthography-to-phonology mapping consistency C2 (1 = consistent)	0.007	−0.018
Phonological regularity C1 (1 = regular, −0.5 = irregular, −0.5 = change of tone)	−0.014*	0.053
Phonological regularity C1 (0 = regular, 0.5 = irregular, −0.5 = change of tone)	0.017**	−0.034
Phonological regularity C2 (1 = regular, −0.5 = irregular, −0.5 = change of tone)	−0.005	0.034
Phonological regularity C2 (0 = regular, 0.5 = irregular, −0.5 = change of tone)	0.010	−0.079
Step 3 - R^2 change	0.021**	0.017
Neighborhood size C1	−0.110**	0.003
Neighborhood size C2	−0.068**	−0.001
Number of meanings C1	−0.016*	−0.006
Number of meanings C2	0.003	−0.005
Semantic transparency C1	0.028**	−0.047
Semantic transparency C2	0.008	−0.003
Step 4 - R^2 change	0.013**	0.002

*$p < .05$ (two-tailed). **$p < .01$ (two-tailed). C1 = first character. C2 = second character. NA1 = this variable was excluded due to multicollinearity problem. NA2 = the regression coefficient could not be estimated due to the problem of singularities. The R^2 change in accuracy analyses was based on pseudo R^2 values, so their statistical significance could not be determined, and they cannot be compared with the R^2 values estimated in zRT analyses

for by phonological variables in Chinese naming relative to English naming (e.g., Balota et al., 2004) could be due to the deeper orthography in the Chinese language relative to the English language.

Predictions 1 and 2: Effect of lexical variables on naming zRT and accuracy We test whether our findings replicate the lexical effects of previous speeded naming studies in Chinese, most of which were based on single-character naming factorial-design experiments. Our findings were generally consistent with those reported in previous works, although there are some exceptions, thereby being partially congruent with our Prediction 1.

For orthographic variables, naming responses were faster for words comprising characters with fewer strokes (e.g., 人) than those with more strokes (e.g., 體), consistent with previous findings (e.g., Shen & Zhu, 1994). The naming responses were faster and more accurate for higher-frequency words (e.g., 多謝 thank you), relative to lower-frequency words (e.g., 睿哲 divinely wise), and for those with higher-frequency characters (e.g., 花 flower), relative to those with lower-frequency characters (e.g., 埅 compliance). These findings are also consistent with earlier studies (e.g., Gao et al., 2016; Yu & Cao, 1992). Consistent with our Prediction 2, character variables (e.g., first-character frequency) were more predictive of naming performance, as reflected by larger regression coefficients (see Table 4), than word variables (e.g., word frequency). In addition, the effects of number of strokes and character frequency were larger for the first character than the second character.

For phonological variables, naming responses were faster and more accurate when the first or second character was phonologically consistent (e.g., 骨 is pronounced as gwat1 in all words with it, such as骨頭gwat1 tau4 [bone]), relative to phonologically inconsistent (e.g., 重 is pronounced as cung5 in 體重tai2 cung5 [weight], zung6 in 重要zung6 jiu3 [important], or cung4 in 重申cung4 san1 [reiterate]). This effect was in line with previous findings (e.g., Leong & Cheng, 2003), although the effect was similar[8] for the first and second characters. The naming responses were faster when the first character contained more phonemes (e.g., 框 kwaang1), relative to fewer phonemes (e.g., 之 zi1), and when the first character was phonologically regular (i.e., pronounced the same as its phonetic radical) than when it shared the same syllable but not the same tone with its phonetic radical (i.e., change of tone) or was phonologically irregular (i.e., pronounced differently from its phonetic radical). The naming responses were slower when the first character was phonologically irregular (vs. change of tone). The phonological regularity effect was in line with Hue et al. (1992) and it occurred in the first, but not the second, character. The naming responses were slower when the first character was higher in homophone density (e.g., 支zi1, which is also the pronunciation of 祇, 諮, 孜, 蜘, 吱, 芝, 茲, 肢, 滋, 脂, 姿, 枝, 資, 知, 之, and 恣...) than lower in homophone density (e.g., 鏡 geng3, a pronunciation is not shared with other characters), contradicting the positive or null effect of homophone density reported in previous studies (e.g., Chang et al., 2016; Ziegler et al., 2000). The null effect of orthography-to-phonology mapping consistency was inconsistent with the facilitatory effect reported by Yang et al. (2009). Other than phonological consistency, all significant effects of phonological variables on naming were larger for the first character than the second character, partially consistent with our Prediction 2.

For semantic variables, naming responses were faster when the first character was associated with larger neighborhood size (e.g., 花 flower, relative to smaller neighborhood size, e.g., 凰 phoenix) or more meanings (e.g., 行, relative to fewer meaning, e.g., 廚). These are consistent with previous findings (e.g., Chang et al., 2016; Peng et al., 2003). In contrast, naming responses were slower when the first character was more transparent in meaning. All these effects occurred only for the first character or were larger for the first character than the second character, consistent with our Prediction 2.

Prediction 3: Comparison between lexical decision and naming To test our Prediction 3, we compare the effect of lexical variables on naming versus lexical decision by combining Tse et al.'s (2017) normed lexical decision data with the current naming data. After excluding words that yielded lower than 70% accuracy in lexical decision or naming data, we ran analyses similar to those we did for naming data based on 18,736 words (about 74.11% of 25,281 words in total). The betas of these lexical variables when they were first entered in the regression models and the R^2 change for zRT and pseudo R^2 change for accuracy at each step are reported in Table 5. The total proportions of variance accounted for were 36.5% and 20.1% for lexical decision zRT and accuracy, respectively, and 51.8% and 19.1% for naming zRT and accuracy, respectively.

Variance accounted for in lexical decision and naming zRT and accuracy Consistent with previous naming studies in other languages (e.g., Yap & Balota, 2009), the characteristics of the initial phoneme of the first character accounted for more variance in naming zRT than in lexical decision zRT (10.9% vs. 0.2%, respectively) but were quite similar in naming accuracy (0.3%) and lexical decision accuracy (0.1%).

To compare the variance explained by orthographic, phonological, and semantic variables in lexical decision and naming, orthographic variables accounted for slightly more variance in naming zRT (37.6%) than lexical decision zRT (34.9%), but the difference was the opposite for accuracy (16.9% vs. 18.7%, respectively). Consistent with our Prediction 3, phonological variables accounted for slightly more variance in naming zRT/accuracy than lexical decision zRT/accuracy (2.1%/1.6% vs. 0.3%/0.1%, respectively).

[8] The larger effect of phonological consistency in zRT for the first character, beta = −.035, than the second character, beta = −.017, was somehow counteracted by the slightly larger effect in accuracy for the second character, beta = .272, than the first character, beta = .266.

Table 5 Standardized regression coefficients (beta values) on lexical decision and naming zRT and accuracy rate ($N = 18,736$)

	Lexical Decision		Naming	
	zRT	Accuracy	zRT	Accuracy
First character's initial phoneme - affricative	0.006	−0.101	0.191**	0.150
First character's initial phoneme - alveolar	NA1	0.049	NA1	0.115
First character's initial phoneme - bilabial	−0.009	0.010	−0.039*	0.256
First character's initial phoneme - labiodentals	0.018*	0.005	−0.070**	0.026
First character's initial phoneme - dental	−0.024	< 0.001	−0.013	0.016
First character's initial phoneme - fricative	−0.020	−0.098	0.246**	0.223
First character's initial phoneme - glottal	0.026**	NA2	−0.153**	NA2
First character's initial phoneme - liquid	0.011	−0.106	−0.015	0.123
First character's initial phoneme - stop	0.045	−0.117	−0.044	0.192
First character's initial phoneme - nasal	−0.021	−0.032	−0.060**	0.150
First character's initial phoneme - palatal	−0.034	0.026	−0.021	0.561
First character's initial phoneme - velar	−0.038*	0.127	0.049**	0.150
First character's initial phoneme - voiced	0.052	−0.087	−0.061	−0.140
Step 1 - R^2 change	0.002**	0.001	0.109**	0.003
Number of strokes C1	0.027**	0.011	0.156**	−0.013
Number of strokes C2	0.026**	0.009	0.066**	0.003
Log frequency C1	−0.067**	0.028	−0.405**	0.675**
Log frequency C2	−0.065**	0.003	−0.210**	0.575**
Log frequency (word)	−0.545**	0.625**	−0.151**	0.236**
Step 2 - R^2 change	0.349**	0.187	0.376**	0.169
Number of phonemes C1	−0.021**	0.022	−0.144**	0.062
Number of phonemes C2	−0.002	−0.002	0.005	−0.004
Homophone density C1	−0.019**	< 0.001	0.020**	0.005
Homophone density C2	−0.015*	< 0.001	−0.008	0.002
Phonological consistency C1 (1 = consistent)	−0.029**	0.043	−0.029**	0.242*
Phonological consistency C2 (1 = consistent)	−0.027**	0.053	−0.017**	0.260*
Orthography-to-phonology mapping consistency C1 (1 = consistent)	0.003	−0.003	0.003	0.034
Orthography-to-phonology mapping consistency C2 (1 = consistent)	0.010	0.003	0.005	−0.018
Phonological regularity C1 (1 = regular, −0.5 = irregular, −0.5 = change of tone)	0.005	< 0.001	−0.016**	0.074
Phonological regularity C1 (0 = regular, 0.5 = irregular, −0.5 = change of tone)	0.007	0.023	0.017**	−0.042
Phonological regularity C2 (1 = regular, −0.5 = irregular, −0.5 = change of tone)	−0.001	0.002	−0.003	0.024
Phonological regularity C2 (0 = regular, 0.5 = irregular, −0.5 = change of tone)	0.021**	−0.069	0.010	−0.107
Step 3 - R^2 change	0.003**	0.001	0.021**	0.016
Neighborhood size C1	−0.060**	0.001	−0.101**	0.003
Neighborhood size C2	−0.044**	< 0.001	−0.068**	−0.001
Number of meanings C1	0.014	−0.002	−0.015*	−0.001
Number of meanings C2	0.037**	−0.008	0.003	−0.001
Semantic transparency C1	−0.065**	0.185**	0.034**	−0.092
Semantic transparency C2	−0.063**	0.162*	0.017**	−0.026
Step 4 - R^2 change	0.011**	0.012	0.012**	0.003

*$p < .05$ (two-tailed). **$p < .01$ (two-tailed). C1 = first character. C2 = second character. NA1 = this variable was excluded due to multicollinearity problem. NA2 = the regression coefficient could not be estimated due to the problem of singularities. The R^2 change in accuracy analyses was based on pseudo R^2 values, so their statistical significance could not be determined, and they cannot be compared with the R^2 values estimated in zRT analyses

Semantic variables accounted for a similar proportion of variance in naming and lexical decision zRT (about 1.1–1.2%), although they accounted for slightly more variance in lexical decision accuracy than naming accuracy (1.2% vs. 0.3%, respectively). The latter finding was in line with our Prediction 3 and consistent with the emphasis of lexical decision on semantic information for word/nonword discrimination. Nevertheless, it is important to note that the differences in the contribution of orthographic, phonological, and semantic variables between lexical decision and naming were smaller in Chinese word processing than English word processing (e.g., Yap & Balota, 2009), all in all suggesting that similar word recognition processes/mechanisms might be used to process two-characters Chinese words in the lexical decision and naming tasks. This was attributed to the deep orthography of Chinese language. The overall variance patterns of naming and lexical decision findings based on this subset ($N = 18{,}736$) were the same as those reported above and in Tse and Yap (2018), respectively, although the word pools were not perfectly overlapping across studies, and transformed accuracy rates (in logits), instead of raw accuracy rate as in Tse and Yap, were used in the current study.

Effect of lexical variables on lexical decision and naming zRT and accuracy The overall pattern of naming results was similar to those reported above, except that now both semantically transparent first and second characters, not just the semantically transparent first character, predicted slower naming zRT. Still consistent with our Prediction 2, the effect of semantic transparency was larger for the first character than for the second character. For the lexical variables shared between the two studies, the pattern of lexical decision findings was similar to that reported in Tse and Yap (2018), except that the number of strokes, character frequency, and phonological consistency no longer predicted lexical decision accuracy. It is noteworthy that, unlike Tse and Yap's results, there was no speed–accuracy trade-off for the effect of number of strokes on lexical decision performance in the current study.[9]

The comparison of the effects of lexical variables on lexical decision and naming showed the following findings (see Table 5).

First, we found that the character with more strokes slowed both lexical decision and naming zRT, although the effect was larger on naming than on lexical decision. Similarly, the facilitatory effect of character frequency was also larger on naming than on lexical decision. This effect was also larger in the first character than in the second character for naming, but the effect was of similar magnitude between characters for lexical decision. In contrast, the facilitatory effect of word frequency was larger on lexical decision than on naming. Similar results regarding word frequency were obtained in the English Lexicon Project (Balota et al., 2004).

Second, the facilitatory effect of the first character with more phonemes was larger for naming zRT than for lexical decision zRT. While the homophone density of the first and second characters sped up lexical decision responses, the homophone density of the first character, but not second character, slowed naming responses. The facilitatory effect of phonological consistency on accuracy was larger in naming than in lexical decision, although the effect was similar in size for lexical decision and naming zRT. There was a facilitatory effect of first-character phonological regularity on naming, but not on lexical decision. For the character that shared the same syllable but different tone with its phonetic radical (i.e., "change of tone"), we obtained faster naming zRT for these "change-of-tone" first characters (e.g., 胖bun6, which shares the same syllable but not the same tone with its phonetic radical 半bun3) than phonologically irregular first characters (e.g., 耀jiu6, which does not share pronunciation with its phonetic radical 翟zaak6). The lexical decision zRT was also faster for "change-of-tone" second character than phonologically irregular second character.

Finally, the facilitatory effect of first and second character's neighborhood size was larger in naming zRT than in lexical decision zRT.[10] The words with more first-character meaning sped up naming responses, whereas those with more second-character meaning slowed down lexical decision responses. Whereas the first and second character's semantic transparency facilitated lexical decision zRT and accuracy, they slowed naming zRT (especially the first character's one).

Discussion

The purpose of the current study was twofold. First, using a megastudy approach, we expanded Tse et al.'s (2017) database, which was developed for lexical decision

[9] These discrepancies were attributed to different dependent variables (logit-transformed accuracy rate in the current study versus raw accuracy rate in Tse & Yap, 2018) used in the analyses. Another difference was the inclusion of characteristics of the initial phoneme of the first character in the analyses. We reran another set of regression analyses, which was the same as reported in the main text but excluding the characteristics of the initial phoneme of the first character. The results of those analyses were qualitatively the same as those that included characteristics of the initial phoneme of the first character, except that lexical decision responses were slower when the first character consisted of a greater number of meanings.

[10] Given the moderate correlation between neighborhood size and character frequency (+.562 and +.538, see Table 3), characters with more neighbors are often of high character frequency, so one could argue that neighborhood size might not be a pure semantic variable. In the current analyses, we found that first-character and second-character frequency and first-character and second-character neighborhood size accounted for unique variance in naming performance. Nonetheless, in future research it will be important to examine the interaction between neighborhood size and character frequency and test whether these two variables influence the common stage of Chinese word processing during naming, based on Sternberg's (1967) additive factors logic (see Yap et al., 2008, for an example).

performance, by norming speeded naming RT and accuracy rates for more than 25,000 traditional Chinese two-character words. We also compiled more lexical variables (e.g., phonological consistency and semantic neighborhood size). Second, we performed item-level multiple regression analyses to test the relative predictive power of orthographic variables (e.g., stroke count), phonological variables (e.g., phonological consistency), and semantic variables (e.g., semantic transparency) in naming and compared these with those in lexical decision, based on Tse et al.'s normed data, to determine whether specific lexical effects are task-specific or task-general. We had three major predictions. Prediction 1: We expected to replicate the lexical effects of previous Chinese speeded naming studies. Prediction 2: Given the serial nature of the word pronunciation, we expected that the character variables, especially the first-character variable, would be more predictive of naming performance than word variables. Prediction 3: In the comparison for the proportion variance explained by phonological and semantic variables between lexical decision and naming, we expected phonological variables to account for larger variance than semantic variables in naming, but expected the opposite pattern to occur in lexical decision. In the following, we discuss the findings in response to the above predictions as well as their general implications for Chinese lexical processing.

Prediction 1: Replication of benchmark findings in speeded naming

The current naming data generally replicate the standard lexical effects in Chinese naming literature. Naming performance, as indicated by faster zRT and/or higher accuracy, was facilitated as a function of word frequency (e.g., Gao et al., 2016), both characters' character frequency (e.g., Yu & Cao, 1992), phonological consistency (e.g., Leong & Cheng, 2003), and neighborhood size (e.g., Chang et al., 2016), and the first character's number of phonemes, number of meanings, and phonological regularity (e.g., Hue et al., 1992). In contrast, naming performance was inhibited as a function of number of strokes for both characters (e.g., Shen & Zhu, 1994) and as a function of homophone density and semantic transparency for the first character. The inhibitory effect of homophone density was inconsistent with that reported in previous studies (e.g., Chang et al., 2016; Ziegler et al., 2000). Also, unlike Yang et al. (2009), orthography-to-phonology mapping consistency did not predict naming performance.

It is noteworthy that most of the previous findings on Chinese speeded naming were based on single characters; that is, participants read aloud the single characters shown on the screen, rather than two-character words as in the present study. Psycholinguistic effects might not necessarily be generalizable from single-character studies to multiple-character studies. When the character appears in a word context, the lexical characteristics of word and/or another character might moderate the processing of the character being pronounced. For example, naming a character in a high-frequency word might be less affected by its lexical characteristics than when the same character appears in a low-frequency word, as word familiarity might facilitate the activation of the whole word and in turn its characters. On the other hand, given the sequential (i.e., first, then second character) processing implicated in word naming, the influence of the lexical characteristics of the second character might be preempted by that of the lexical characteristics of the first character and the whole word. These possibilities could be further investigated by comparing the lexical effects on lexical decision and naming in the isolated context versus word context. Moreover, in the regression analyses, the interaction terms could be added to test the moderating role of a specific word's/character's variables. Nevertheless, the results of the lexical effects in our normed naming data provide useful empirical benchmarks for the Chinese word processing literature.

Prediction 2: Predictive power of character versus word variables and the first-character versus second-character variables in speeded naming

By comparing the differences in the contribution of the lexical effect of the words and/or the first and second characters on speeded naming, we may determine whether the pronunciation of the Chinese words is activated holistically or through the combination of the characters, and whether the influence of the first character is stronger than that of the second character. For the effects of the character's versus the word's lexical variable, we compare the character and word variables at the same (orthographic) level, that is, the effect of character frequency versus the effect of word frequency. Consistent with our Prediction 2, the effect of word frequency was smaller than the effect of character frequency, especially the first character's frequency, in both zRT and accuracy (see Table 4). For the effects of the first character's versus second character's lexical variable, we found that the significant lexical effects, with the exception of phonological consistency, of the first character on naming performance were stronger than those of the second character (see the regression coefficients in Table 4), again in line with our Prediction 2. These suggest that when participants read aloud a word, they first processed the first character, followed by the second character, rather than processing them simultaneously. Given that the voice key was triggered to record naming RT once participants began to pronounce the first character (i.e., speech onset), their naming responses may be more influenced by the lexical characteristics of the first character than those of the second character and the word. This result is consistent with the analytic view (e.g., Tse et al., 2017), but not the holistic view (e.g., Packard, 1999) of Chinese word processing. At the very least, these findings show that the holistic versus analytic nature of word processing is flexible and depends on task demands. This view is similar to word processing in English in that the meaning of a word is simultaneously and independently accessed via direct retrieval of a whole-word representation, which is more sensitive to word frequency, and a decomposition-then-composition process of constituent (character) representations, which is more sensitive to constituent variables (e.g., Baayen et al., 1997; Libben & Jarema, 2007). Other character-level and word-level lexical variables (e.g., word imageability) should be considered in future research to further test the holistic versus analytic nature of Chinese word processing.

Prediction 3: Predictive power of phonological versus semantic variables in speeded naming versus lexical decision

In the Chinese language, the orthography-to-phonology mapping is nearly deterministic (i.e., often in a one-to-one relationship, e.g., 表pronounced as biu2), whereas the orthography-to-semantics mapping is under-deterministic (i.e., often in a one-to-many relationship, e.g., 表can mean watch, express, surface, or meter). Perfetti and Tan (1998, 1999) argued that the one-to-one relation can be more quickly established in Chinese word recognition, so phonology should play a stronger role than semantics in Chinese word processing (see also Tan & Perfetti, 1999, for a model of visual recognition of two-character Chinese words). Tse and Yap (2018) analyzed Tse et al.'s (2017) normed lexical decision data and showed that orthographic and semantic variables, respectively, accounted for more variance than phonological variables, although some could point out that the lexical decision task does not require the generation of word phonology. Based on the R^2 or pseudo R^2 changes in each step (see Tables 4 and 5), orthographic variables (e.g., character and word frequency) were associated with the strongest predictive power on naming performance, followed by phonological variables; semantic variables accounted for the least variance in our analyses. Phonological variables did account for more variance than semantic variables in predicting naming performance, consistent with Tan and Perfetti's view. However, the proportion of variance accounted for was still low for phonological variables (smaller than 2.5%) even when the naming task explicitly tapped character phonology, suggesting that the contribution of word phonology might not be as large as what Tan and Perfetti proposed.

Few studies have directly compared the effect of lexical variables on lexical decision versus naming for two-character Chinese words. Consistent with these previous findings (e.g., Gao et al., 2016; Li et al., 2017), we found that word frequency was a stronger predictor for lexical decision performance (−.545) than naming performance (−.151, see Table 5). This is compatible with the view that lexical decision emphasizes more frequency information, an index of whole-word familiarity, when participants discriminate words from nonwords (e.g., Balota & Chumbley, 1984). On the other hand, because participants read aloud the two-character word beginning from the first character, the first-character-focused naming process might reduce the influence of the word frequency effect in Chinese lexical processing. In fact, when comparing the effects of the first versus second character's lexical variables on lexical decision and naming performance (see Table 5), we did find that the predictive power was stronger for the first character's variable than the second character's variable in naming zRT (e.g., −.405 vs. −.210 in character frequency, −.144 vs. .005 in the number of phonemes, and .034 vs. .017 in semantic transparency). In contrast, the difference in predictive power of the first versus second character's variable was not as large in lexical decision (e.g., −.067 vs. −.065 in character frequency, −.021 vs. −.002 in the number of phonemes, and −.065 vs. −.063 in semantic transparency). The significant predictive power of character variables showed that words are processed not merely holistically (as reflected by word variables), but also analytically (as reflected by character variables). The different patterns of predictive power of the first versus second character's orthographic, phonological, and semantic variables in the lexical decision and naming tasks suggest that the influence of first versus second character's variables depends heavily on the task demand, regardless of the lexical characteristics of the characters/words.

The importance of task demands could be observed in the relative contribution of the phonological and semantic variables on naming and lexical decision performance. While orthographic variables accounted for much more variance than semantic and phonological variables in both lexical decision and naming performance, semantic variables accounted for more variance than phonological variables in lexical decision, and phonological variables accounted for more variance than semantic variables in naming. While these are consistent with our Prediction 3, semantic variables accounted for just slightly more variance in lexical decision accuracy (.012) than in naming accuracy (.003) but similar variance when the zRT measure was considered (.011 vs. .012) (see Table 5). This suggests that phonological variables clearly contributed more to naming performance than to lexical decision performance, but the contribution of semantic variables was quite similar in the performance of the two tasks. This was in contrast to those obtained in

English, where the influence of semantic variables was more salient in lexical decision than in naming (e.g., Chang et al., 2019; Cortese & Khanna, 2007; Cortese et al., 1997; Yap & Balota, 2009). Semantic variables might have more influence on Chinese naming than on English naming because more than half of Chinese characters have no systematic orthography-to-phonology mappings (Chang et al., 2016), such that readers might rely on semantics when pronouncing Chinese characters (see also Chang & Lee, 2018).

Finally, it is noteworthy that there was a marked contrast in the role of semantic transparency in lexical decision versus naming performance. While semantic transparency facilitated lexical decision, the same variable, especially for the first character, had an inhibitory effect on naming. Following the general dual-route model of compound word processing (e.g., Libben & Jarema, 2007), the meaning of a word can be accessed via direct retrieval from semantic memory or via decomposition, with the meanings of its characters first activated and then combined to obtain the word meaning. The outputs from direct retrieval and decomposition routes are similar for transparent words (e.g., 花園garden). In contrast, for opaque words at least one of the characters is unrelated to the whole word, so there is a conflict between the two routes (e.g., combined meaning of the two unrelated characters, 花flower and 生grow versus meaning of the word, 花生peanut). In lexical decision, this conflict may trigger participants to do post-lexical checking to confirm the word lexicality (see, e.g., Balota & Chumbley, 1984), thereby slowing the word recognition process for opaque, relative to transparent, words and producing the semantic transparency effect (see, e.g., Kim et al., 2019). In naming, there was no need for post-lexical checking, as participants read aloud the word character-by-character, so the conflict in the meaning of whole word and its characters likely did not influence the naming responses. That being said, participants named opaque words faster than transparent words. This negative semantic transparency effect was unexpected and should be further examined in future research.

Comparison with Chang and Lee's (2020) and Sun et al.'s (2018) findings

Before concluding the current study, two recent studies using the megastudy approach to examine the influence of lexical variables on lexical decision and naming of Chinese characters/words are worth discussing in the context of our current findings.

Using 3314 traditional Chinese single characters as their stimuli, Chang and Lee (2020) examined the predictive power of age of acquisition, as well as semantic variables, in naming and lexical decision of traditional Chinese characters. Contrary to the findings in English (e.g., Yap & Balota, 2009), they found that semantic variables had higher predictive power in naming than in lexical decision. These findings were also incongruent with our current findings that semantic variables had similar predictive power in naming zRT and lexical decision zRT. In fact, semantic variables even accounted for slightly more variance in lexical decision accuracy than in naming accuracy. However, it is noteworthy that phonological and semantic variables were defined differently between the two studies. Indeed, there were only two lexical variables in common in Chang and Lee and the current study.[11] Character frequency, defined as an orthographic variable in the current study but a semantic variable in Chang and Lee, was more predictive of naming than lexical decision in both studies. Moreover, the facilitatory effect of character frequency was also consistently obtained in the two studies. On the other hand, we found that number of strokes was more predictive of naming than lexical decision, contrary to the absence of such task differences reported by Chang and Lee.

Apart from little overlap in the lexical variables in the analyses, there were a number of stimulus and procedural differences that might complicate the direct comparison of the findings of the two studies. First, whereas Chang and Lee (2020) used single characters as stimuli, we used two-character words. Second, their participants read aloud the characters in Mandarin, instead of Cantonese as in the current study. Third, phonological variables were quantified based on different dialects (Mandarin and Cantonese) in the two studies. Fourth, as mentioned above, most of the lexical variables were not shared between the two studies. For instance, concreteness, imageability, and age of acquisition were included in Chang and Lee but not in the current study, whereas semantic transparency and neighborhood size were included in the current study but not in Chang and Lee. Hence, future research can be directed at determining whether some of these differences contributed to the above discrepancies in the findings in the two studies.

Unlike Chang and Lee (2020), which focuses on the comparison between lexical decision and naming of single

[11] Phonological consistency was defined at the sublexical level in Chang and Lee (2020)—that is, the number of friends (characters sharing the same phonetic radical and pronunciation) divided by the total number of characters sharing the same phonetic radical—whereas in the current study, we define that as whether a character has one (phonologically consistent) or more than one (phonologically inconsistent) pronunciation. For phonological regularity, Chang and Lee defined the characters with unpronounceable phonetic radical as a separate group, whereas in the current study they were defined as irregular characters. They did not separate the characters that shared the same syllable but not the same tone with their phonetic radical (i.e., "change of tone" characters), as in the current research. Finally, although Chang and Lee defined semantic ambiguity as the number of meanings of a character, they quantified that based on subjective ratings rather than raw count of distinct conceptual representations of a character, as in the current study. Hence, we do not compare the findings for these variables across the two studies.

traditional characters, Sun, Hendrix, Ma, and Baayen (2018) included single-character, two-character, three-character, and four-character words in their Chinese lexical database for simplified Chinese and performed analyses on existing lexical decision and naming data in other megastudies (e.g., naming data in Sun, 2016, as cited in Sun et al., 2018 and lexical decision data in Tse et al., 2017 and Tsang et al., 2018). It is noteworthy that the lexical variables in Sun et al. were not the same as those in the current study, as they were based on Mandarin dialect and simplified-script characters, in contrast to Cantonese dialect and traditional-script characters in the current study. Given the difference in script, Sun et al. restricted the evaluation of their measures to the subset of the two-character words in Tse et al. (2017), for which the written form is identical in simplified and traditional Chinese, such that the number of words was reduced from 25,281 to 8005. This limits the number of words in their analyses. Sun et al.'s naming data were provided by a single participant and based on 25,935 two-character words. In the following, we compare the current findings with Sun et al.'s results on single characters and two-character words. We also consider their single-character data, as we did with Chang and Lee and other studies that did not use the megastudy approach.

For single-character lexical decision data, in line with the current study, Sun et al. (2018) reported the facilitatory effect of character frequency and inhibitory effect of number of strokes. For two-character lexical decision data, consistent with the current study, Sun et al. also found the facilitatory effect of word frequency and character frequency, with the former effect being larger than the latter. The effects of first-character and second-character frequency were also similar in magnitude. Although Sun et al. did not obtain a significant effect of stroke count, this could be due to the restricted set of stimuli that was used in their study and/or that the effect size of this variable was rather small, as indeed reflected by their relatively small regression coefficients in our analyses (see Table 5).

For single-character naming data, in line with the current findings, Sun et al. (2018) found a facilitatory effect of character frequency and inhibitory effect of number of strokes. For two-character naming data, Sun et al. also found a facilitatory effect of character frequency and word frequency. However, the effects of character frequency (beta = −.025 for the first character and −.012 for the second character) were similar to or even smaller than the effect of word frequency (beta = −.021), in contrast to the current study in which the effect of first-character or second-character frequency was larger than the effect of word frequency (see Tables 4 and 5). However, the finding that the effect of first-character frequency was stronger than that of second-character frequency was in line with our current findings. Consistent with the current study, Sun et al. also found that the naming responses were slowed with the number of strokes, with the effect being larger for the first characters than the second characters.[12]

Unlike other studies (including the current one), Sun et al. (2018) used information-theoretic measures to examine how the uncertainty of characters at the word level (e.g., entropy and conditional probability) influenced lexical decision and naming performance. They obtained a facilitatory effect of first-character entropy, suggesting that the greater uncertainty about the second character given the first character could trigger faster lexical decision and naming responses. Although this lexical decision finding was in contrast to the findings in the English language (e.g., Schmidtke et al., 2016; Hendrix et al., 2017), it highlights the importance of taking the combinatorial properties of characters into account when investigating lexical processing above the character level in future studies.

Conclusion and future directions

Motivated by the megastudy approach (e.g., Balota et al., 2013), the current extension of the Chinese Lexicon Project to speeded naming data represents an important addition to the different existing lexicon projects (e.g., English, Dutch, French, and Malay) being rapidly developed across the world and in line with the current research zeitgeist. Given that there are more native speakers of Chinese than any other language and that two-character words are the most common type of word encountered in Chinese reading, a two-character Chinese word speeded naming database can significantly contribute to research in psycholinguistics and other research domains using Chinese word stimuli. In the .xlsx/.csv files available at: https://osf.io/vwnps, we make the item-level data freely accessible to the research community, which allows researchers to search for naming RT, zRT, and accuracy rate for words and descriptive statistics of lexical variables (Table 1) for characters and words. This database serves as a critical resource of lexical characteristics and

[12] The orthography-to-phonology mapping measures in Sun et al. (2018) were not defined in the same way as in the current study, e.g., phonological consistency and orthography-to-phonology mapping consistency). Specifically, Sun et al. defined the "friend of a character" as an occurrence of the same character–pronunciation mapping in a different word. It might be a bit similar to the neighborhood size in the current study, although we count the occurrence of the character in different words, regardless of whether the character was pronounced similarly or differently. For instance, 重要, 重量, and 重陽 are all counted toward the neighborhood size, even though 重 is pronounced differently in these three words. Nevertheless, similar to the facilitatory effect of the first-character friends on single-character lexical decision and naming data of Sun et al. (i.e., faster responses for those with more first-character friends), we found that words with larger first-character neighborhood size yielded faster responses than those with smaller first-character neighborhood size.

behavioral measures for future research. Researchers could perform item-level regression analyses to test the higher-order interactions among lexical variables (e.g., word frequency × character frequency interaction, to determine whether character frequency effect might be particularly salient for certain levels of word frequency), the evidence of which might be mixed in previous factorial-design experiments. This may address some theoretical issues in Chinese word processing, such as whether words and their characters are represented at the same level in the Chinese mental lexicon, whether words are accessed as a whole or via their characters, and the extent to which phonological and semantic variables interact to influence Chinese naming, which will have implications on the models of Chinese word recognition (see, e.g., Tse & Yap, 2018, for an example based on Tse et al.'s, 2017, normed lexical decision data).

Researchers could add other variables to our dataset to make it more comprehensive. For instance, previous studies showed that characters with higher phonological frequency (i.e., those with higher cumulative character frequency for their homophone mates) were named more slowly than those with lower phonological frequency (e.g., Ziegler et al., 2000, but see Chen et al., 2009). Chang and Lee (2020) showed that, after controlling for semantic variables, age of acquisition accounted for more variance in naming than in lexical decision for single characters. In addition, this variable interacted with phonological consistency in predicting character naming. Huang et al. (2006) showed that characters with higher summed frequency of all words in which a character occurred were named more slowly than those with lower summed frequency. It is important to test whether all these findings could be generalized in the current dataset (i.e., a large pool of two-character words with traditional script and Cantonese-speaking participants). The present large-scale data for various lexical variables and speeded naming performance could also facilitate the development of computational models of Chinese lexical processing (e.g., Shuai & Malins, 2017; Yang et al., 2009), with their computer simulation providing greater clarity and transparency than traditional descriptive models. Apart from addressing theoretical questions related to word recognition models, the current data, together with those developed by Tse et al. (2017) for lexical decision, could be regarded as normative data for traditional Chinese word recognition by native Cantonese-speaking university students and utilized to build Chinese proficiency tests. Using the findings from the French Lexicon Project (Ferrand et al., 2010), Brysbaert (2013) selected French words and nonwords varying in difficulty level and constructed a short French language proficiency test (LEXTALE-FR), which yielded good psychometric properties and can be used to assess French proficiency for both native and second language learners. Researchers could explore the possibility of developing a Chinese proficiency test with receptive (word recognition) and production (word pronunciation) components, based on lexical decision and speeded naming data, respectively. The large word pool in these databases would make it possible to construct multiple forms of tests that involved different sets of words and nonwords with similar difficulty levels, such that readers, be they first or second language learners, could be assessed without repeating the same set of words.

Open Practices Statement

The data are available at: https://osf.io/vwnps, and none of the experiment was preregistered.

Acknowledgments This research was supported by the General Research Fund (#14600019) awarded by the Hong Kong Research Grants Council. We thank Chloe Chow for her help in the data collection.

References

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 514–540.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language, 37*, 94–117.

Baayen, R. H., Feldman, L., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 53*, 496–512.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340–357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language, 24*, 89–106.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2013). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition, Models and methods, orthography and phonology* (Vol. *1*, pp. 90–115). Psychology Press.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers, 34*, 424–434.

Brysbaert, M. (2013). LEXTALE-FR: A fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica, 53*, 23–37.

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

Chang, Y.-N., Hsu, C.-H., Tsai, J.-L., Chen, C.-L., & Lee, C.-Y. (2016). A psycholinguistic database for traditional Chinese character naming. *Behavior Research Methods, 48*, 112–122.

Chang, Y.-N., & Lee, C.-Y. (2018). Semantic ambiguity effects on traditional Chinese character naming: A corpus-based approach. *Behavior Research Methods, 50*, 2292–2304.

Chang, Y.-N., & Lee, C.-Y. (2020). Age of acquisition effects on traditional Chinese character naming and lexical decision. *Psychonomic Bulletin & Review, 27*, 1317–1324.

Chang, Y.-N., Monaghan, P., & Welbourne, S. (2019). A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language, 108*, 104025.

Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes, 24*, 967–982.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60*, 1072–1082.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4*, 226–231.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777–799.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., ... & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42*, 488–496.

Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., et al. (2018). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods, 50*, 1285–1307.

Gao, X.-Y., Li, M.-F., Chou, T.-L., & Wu, J.-T. (2016). Comparing the frequency effect between the lexical decision and naming tasks in Chinese. *Journal of Visual Experiment, 110*, e53815.

Goh, W. D., Yap, M. J., & Chee, Q. W. (2020). The Auditory English Lexicon Project: A multi talker, multi-region psycholinguistic database of 10,170 spoken words and nonwords. *Behavior Research Methods, 52*, 2202–2231.

Hendrix, P., Bolger, P., & Baayen, R. H. (2017). Distinct ERP signatures of word frequency, phrase frequency and prototypicality in speech production. *Journal of Experimental Psychology: Language, Memory, and Cognition, 3*, 128–149.

Hoosain, R. (1992). Psychological reality of the word in Chinese. In H. C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 111–130). Elsevier Science.

Huang, H. W., Lee, C. Y., Tsai, J. L., Lee, C. L., Hung, D. L., & Tzeng, O. J. L. (2006). Orthographic neighborhood effects in reading Chinese two-character words. *Neuroreport, 17*, 1061–1065.

Hue, C.-W., Chen, H.-C., & Tzeng, O. J.-L. (1992). Recognition processes in character naming. *Advances in Psychology, 90*, 93–107.

Institute of Language Teaching and Research. (1986). *A frequency dictionary of Modern Chinese*. Beijing Language Institute Press.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression*. Sage.

Kang, S. H. K., Yap, M. J., Tse, C.-S., & Kurby, C. A. (2011). Semantic size does not matter: "Bigger" words are not recognised faster. *Quarterly Journal of Experimental Psychology, 64*, 1041–1047.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology, 1*, 174. https://doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*, 287–304.

Kim, S. Y., Yap, M. J., & Goh, W. D. (2019). The role of semantic transparency in visual word recognition of compound words: A megastudy approach. *Behavior Research Methods, 51*, 2722–2732.

Leong, C.-K., & Cheng, P.-W. (2003). Consistency effects on lexical decision and naming of two-character Chinese words. *Reading and Writing, 16*, 455–474.

Leong, C.-K., Cheng, P.-W., & Mulcahy, R. (1987). Automatic processing of morphemic orthography by mature readers. *Language and Speech, 30*, 181–196.

Li, M.-F., Gao, X.-Y., Chou, T.-L., & Wu, J.-T. (2017). Neighborhood frequency effect in Chinese word recognition: Evidence from naming and lexical decision. *Journal of Psycholinguistic Research, 46*, 227–245.

Li, X., Zang, C., Liversedge, S. P., & Pollatsek, A. (2015). The role of words in Chinese reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 232–244). Oxford University Press.

Libben, G., & Jarema, G. (2007). *The representation and processing of compound words*. Oxford University Press.

Lim, Y.-H., Yap, M. J., & Tse, C.-S. (2020). Individual differences in Cantonese Chinese word recognition: Insights from the Chinese Lexicon Project. *Quarterly Journal of Experimental Psychology, 73*, 504–518.

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods, 39*, 192–198.

Myers, J. (2006). Processing Chinese compounds: A survey of the literature. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 169–196). Oxford University Press.

Myers, J., Libben, G., & Derwing, B. (2004). *The nature of transparency effects in Chinese compound processing*. In *Poster presented at the Fourth International Conference on the Mental Lexicon*.

Packard, J. L. (1999). Lexical access in Chinese speech comprehension and production. *Brain and Language, 68*, 89–94.

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Peng, D., Deng, Y., & Chen, B. (2003). The polysemy effect in Chinese one-character word identification. *Acta Psychologica Sinica, 35*, 569–575.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 101–118.

Perfetti, C. A., & Tan, L. H. (1999). The constituency model of Chinese word identification. In A. Inhoff, H.-C. Chen, & J. Wang (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 115–134). Erlbaum.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*, 273–315.

Que, D. L. (2008). *Longman Chinese Dictionary* (3rd ed.). Longman.

Roux, F., Armstrong, B., & Carreiras, M. (2016). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods, 49*, 1864–1881.

Schmidtke, D., Kuperman, V., Gagne, C., & Spalding, T. (2016). Competition between conceptual relations affects compound recognition: The role of entropy. *Psychonomic Bulletin & Review, 23*, 556–570.

Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-prime user's guide*. Psychology Software Tools.

Shen, L. M., & Zhu, X. (1994). Investigating stroke count and character frequency effects in the visual recognition of characters. *Psychological Science (in Chinese), 17*, 245–247.

Shuai, L., & Malins, J. G. (2017). Encoding lexical tones in jTRACE: A simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods, 49*, 230–241.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*, 411–416.

Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behavior Research Methods*, *50*, 2606–2629.

Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods, 46*, 263–273.

Sze, W. P., Yap, M. J., & Rickard Liow, S. J. (2015). The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis. *Quarterly Journal of Experimental Psychology, 68*, 1541–1570.

Tan, L. H., & Perfetti, C. A. (1999). Phonological activation in visual identification of Chinese two-character words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 382–393.

Tsai, J.-L., Lee, C.-Y., Lin, Y.-C., Tzeng, O. J.-L., & Hung, D. L. (2006). Neighborhood size effects of Chinese words in lexical decision and reading. *Language and Linguistics, 7*, 659–675.

Tsang, Y.-K., & Chen, H.-C. (2014). Activation of morphemic meanings in processing opaque words. *Psychonomic Bulletin and Review, 21*, 1281–1286.

Tsang, Y.-K., Huang, J., Lui, M., Xue, M., Chan, Y.-W. F., Wang, S., & Chen, H.-C. (2018). MELD-SCH. A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods, 50*, 1763–1777.

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.

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

Xing, H., Shu, H., & Li, P. (2004). The acquisition of Chinese characters: Corpus analyses and connectionist simulations. *Journal of Cognitive Science, 5*, 1–49.

Yang, J. F., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory & Language, 61*, 238–257.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory & Language, 60*, 502–529.

Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 495–513.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 53–79.

Yap, M. J., & Rickard Liow, S. J. (2016). Processing the written word. In V. Cook & D. Ryan (Eds.), *Routledge handbook of English writing system* (pp. 453–469). Routledge.

Yap, M. J., Rickard Liow, S. J., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods, 42*, 992–1003.

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979.

Yu, B., & Cao, H. (1992). A new exploration on the effect of stroke number in the identification of Chinese characters. *Acta Psychologica Sinica, 24*(2), 10–16.

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: the phonological frequency effect in written Chinese. *Psychological Science, 11*, 234–238.