Behav Res
DOI 10.3758/s13428-013-0355-9

The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters

Wei Ping Sze · Susan J. Rickard Liow · Melvin J. Yap

Abstract The Chinese language has more native speakers than any other language, but research on the reading of Chinese characters is still not as well-developed as it is for the reading of words in alphabetic languages. Two areas notably lacking are the paucity of megastudies in Chinese and the relatively infrequent use of the lexical decision paradigm to investigate single-character recognition. The Chinese Lexicon Project, described in this article, is a database of lexical decision latencies for 2,500 Chinese single characters in simplified script, collected from a sample of native mainland Chinese (Mandarin) speakers (N = 35). This resource will provide a valuable adjunct to influential mega-databases, such as the English, French, and Dutch Lexicon Projects. Using two separate analyses, some advantages associated with megastudies are exemplified. These include the selection of the strongest measure to represent Chinese character frequency (Cai & Brysbaert's (PLoS ONE 5(6): e10729, 2010) *subtitle contextual diversity* frequency count), and the conducting of virtual studies to replicate and clarify existing findings. The unique morpho-syllabic nature of the Chinese writing system makes it a valuable case study for functional language contrasts. Moreover, this is the first publicly available large-scale repository of behavioral responses pertaining to Chinese language processing (the behavioral dataset is attached to this article, as a supplemental file available for download). For these reasons, the data should be of substantial interest to psychologists, linguists, and other researchers.

Electronic supplementary material The online version of this article (doi:10.3758/s13428-013-0355-9) contains supplementary material, which is available to authorized users.

W. P. Sze (✉) · S. J. Rickard Liow · M. J. Yap
Department of Psychology, National University of Singapore,
Singapore, Singapore
e-mail: weipingsze@gmail.com

Keywords Mandarin · Visual word recognition · Megastudy · Reaction time · Nonalphabetic · Logograph

The main purpose of this article is to present the Chinese Lexicon Project, a repository of lexical decision response latencies for 2,500 Chinese single characters in simplified script. The visual lexical decision task is a basic lexical processing paradigm (see, e.g., Leong, Cheng, & Mulcahy, 1987) in which participants have to discriminate between real Chinese characters and made-up Chinese characters via a buttonpress.

Investigating the processes underlying Chinese character recognition is important for two reasons. First, Chinese is a major world language. Currently spoken by over a billion individuals, the Chinese language is used as a communicative medium by almost one fifth of the world's population (Rogers, 2005). Given the language's wide usage, it is useful, from a practical standpoint, to gain a better understanding of it. Second, as a language, Chinese has unique linguistic properties. From a psycholinguistic standpoint, Chinese characters afford a window into the cognitive processing of a nonalphabetic script. Understanding how Chinese characters are processed therefore has strong theoretical implications, especially when constructing language universal claims (e.g., the universal phonological hypothesis by Perfetti, Zhang, & Berent, 1992) or when making interlanguage comparisons (e.g., Katz & Frost's (1992) orthographic depth hypothesis). Data on the processing of Chinese are able to provide a neat contrast to that for alphabetic languages, from which the bulk of psycholinguistic knowledge is currently based (see Share, 2008, on how the overemphasis on English has shaped a rather Anglocentric view of psycholinguistics). Specifically, the Chinese writing system is morpho-syllabic, which means that each script symbol corresponds simultaneously to a unit of meaning and a single syllable (Hoosain, 1991, pp. 12–13). Thus, unlike, say, Spanish or English, where there is a quasi-regular relationship between spelling and sound, the lack of correspondence between script and sound at the phoneme level in Chinese means that one usually has to memorize the pronunciation for a character in order to sound it out correctly.

In addition, there are other linguistic properties inherent in Chinese that are distinct from those found in the Indo-European alphabetic languages. In terms of orthography, Chinese characters are logographic and are composed of strokes, which, in turn, combine to form different components (these sometimes map directly onto radicals; see Yin & Rohsenow, 1994, for more information about characters). In terms of phonology, speakers of Chinese rely on lexical tones to help distinguish between characters, another feature less notable in alphabetic languages (see Yip, 2002, pp. 171–196, for an overview of tones in Chinese). A better specification of how these variables affect language processing will allow researchers richer insights into the interplay between orthography, phonology, and semantics (for relevant discussions, see H.-C. Chen & Zhou, 1999; Li, Tan, Bates, & Tzeng, 2006, pp. 1–9).

Despite these interesting contrasts, research on Chinese lexical processing is relatively impoverished, as compared with work on word recognition in English and other alphabetic languages. As yet, there is no widely available database of behavioral data for a large set of Chinese characters. In contrast, the English Lexicon Project (Balota et al., 2007) contains lexical decision and speeded pronunciation data for over 40,000 English words, and there are now similar databases for French (Ferrand et al., 2010), Dutch (Keuleers, Diependaele, & Brysbaert, 2010), Malay (Yap, Rickard Liow, Jalil, & Faizal, 2010), and British English (Keuleers, Lacey, Rastle, & Brysbaert, 2012). Collectively, these databases reflect the megastudy approach (Balota, Yap, Hutchison, & Cortese, 2012), for which researchers have collected behavioral data for large sets of words (typically, at least a thousand items). While small-scale factorial experiments are undoubtedly useful, there are several methodological advantages associated with the megastudy approach that make it an important complement to the factorial approach (see Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Balota et al., 2012). Two major advantages are listed below:

1. Mega-datasets are useful test beds to empirically evaluate competing metrics for the same measure. Brysbaert and New (2009), for example, utilized the latencies from the English Lexicon Project to provide evidence for the superiority of the subtitles-based word frequency counts over more established English word frequency measures, such as the Kučera and Francis (1967) norms or the CELEX norms (Baayen, Piepenbrock, & van Rijn, 1993).
2. Mega-datasets provide a comprehensive sampling of the lexicon. One consequence of this large sampling is that they allow experimenters to conduct virtual studies. By simply selecting the appropriate target items for different experimental conditions, researchers can explore a range of novel research questions, without the effort of collecting new data. This makes the mega-dataset an invaluable resource for researchers to test new hypotheses. Additionally, virtual studies can also help ascertain whether existing findings can be replicated, especially when unusual data patterns are obtained.

Given these two advantages, it seems worthwhile to extend the megastudy approach to the Chinese language. There are three major motivations for this extension.

First, to our knowledge, there are no freely accessible large-scale behavioral databases for any nonalphabetic languages, including Chinese. Unfettered access to such data provides researchers with a quick way to test specific experimental hypotheses pertaining to the cognitive processing of nonalphabetic scripts. This situation contrasts sharply with the alphabetic languages, such as English, French, Dutch, or Malay, where rich repositories of latency data have been published and made publicly available (Balota et al., 2007; Ferrand et al., 2010; Keuleers et al., 2010; Yap et al., 2010). The lexical decision data provided here will help fill this crucial gap.

Second, there has never been a megastudy published on the lexical decision for single Chinese characters. While ours is not the first Chinese megastudy,[1] it is the first comprehensive effort based on the lexical decision paradigm. Lexical decision performance has provided helpful insights into word recognition in other languages (e.g., Rubenstein, Garfield, & Millikan, 1970, and Stone, Vanhoy, & Van Orden, 1997, for English; Bonin, Chalard, Méot, & Fayol, 2001, for French) and in the reading of Chinese double-character words (e.g., B. Chen & Peng, 2001; Myers, Huang, & Wang, 2006; Tsai et al., 2009). It is thus unfortunate that the lexical decision task is not as often utilized by researchers to explore the recognition of single Chinese characters. A Web of Science database search revealed that since the mid-1980s, there have been around 35 studies using the lexical decision task to investigate Chinese single-character recognition, in comparison with 59 studies based on the speeded naming task.[2] There are no good reasons against the use of this task to examine single characters. The disparity is most likely a consequence of the

---

[1] A mega-naming study was previously conducted by Liu, Shu, and Li (2007). Unfortunately, their naming latencies were not released for public access.

[2] The Web of Science search was conducted on July 2, 2012. Results from the search can be generic, so each of the articles generated by the search was checked manually to remove the irrelevant titles.

tremendous labor involved in creating plausible pseudo-Chinese character foils for the lexical decision task.

Finally, as was previously mentioned, Chinese has distinctive psycholinguistic properties. The writing system's crucial lack of grapheme–phoneme correspondence and morpho-syllabic structure are critical as researchers seek insights from nonalphabetic languages to advance what we already know from the alphabetic languages.

In the rest of the article, we will provide a fuller description of our dataset, followed by a demonstration of two potential uses for these data. One important contribution of megastudies is to facilitate empirical adjudication between competing metrics. For this reason, we will be evaluating different frequency norms currently available for simplified Chinese. Selecting an optimal frequency norm in applied research is important because this variable is such a fundamental predictor of Chinese character recognition performance (e.g., Liu, Shu, & Li, 2007; Seidenberg, 1985; You, Chen, & Dunlap, 2009). We will also report a series of virtual experiments in which our data are employed to run virtual replications of some published studies within the field (also see Keuleers et al., 2012). The extent to which our megastudy replicates findings from these factorial experiments will be a good gauge of its validity and generalizability.

Method

Participants

Forty-six native speakers of Mandarin from mainland China signed up for the study. Eleven were excluded due to poor performance either on the screening tasks (described below) or on the lexical decision task (<85 % accuracy, a criterion also adopted by Keuleers et al., 2010).[3] The final sample size was 35 (21 females, 14 males; mean age = 21.8 years, *SD* = 4.14). All participants had normal or corrected-to-normal vision. They were reimbursed for their participation.

Materials and apparatus

Language proficiency screeners

Two language tests were used to screen the participants. They were (1) the Chinese Author Recognition Test and (2) a cloze passage screener taken from the HSK Chinese Proficiency Test (Hànyǔ Shuǐpíng Kǎoshì, 汉语水平考试). Language screening was necessary because we were interested in recruiting proficient readers of the language, whose data would help benchmark skilled visual recognition of Chinese characters.

Performance on the Chinese Author Recognition Test (ART) gauges print exposure—that is, how well-read a person is. The original ART was developed by Stanovich and West (1989) in English, as an objective assessment of the volume of reading a person does. Empirically, the test has been found to predict reading comprehension and spelling ability, after accounting for vocabulary and general abilities (Cunningham & Stanovich, 1997; Stanovich & Cunningham, 1993). ART's utility had also been replicated in the Chinese language (Lee & Krashen, 1996).

To facilitate this study, an updated Chinese version of the paper-and-pencil ART was created following the principles of Stanovich and West (1989). Basically, participants were asked to check the names of the popular authors found in a list of 76 names (half of which were decoys). The test was normed using a separate sample of mainland Chinese students ($N = 29$). The reliability of the items was relatively high (Cronbach's $\alpha = .80$). The participants in the norming exercise yielded an average of 11.4, so cutoff was placed at 12. This means that, as part of the study's qualifying requirement, a participant must score no less than 12 in the ART segment. Negative grading was applied, as was the case for the original ART. The average score obtained by those who qualified was 17.7 (*SD* = 5.56).

A second screener composed of cloze passages was also used. Cloze passages can be administered quickly, and they examine important linguistic skills, such as writing ability, semantic retrieval, and reading comprehension. The cloze passages were taken from the HSK Chinese Proficiency Test (http://www.hsk.org.cn/english/Default.aspx). HSK is a standardized test designed by the Beijing Language and Culture University and is approved by the Ministry of Education in China. It assesses the Chinese proficiency of nonnatives and thus, in many ways, is analogous to the TOEFL (Test of English as a Foreign Language). According to the HSK norms (i.e., for nonnatives), the range of scores corresponding to an "A" grade for the cloze passage section is from 78 % to 100 %. Because our target participants are skilled readers of Chinese, the cutoff was raised to 90 %. There are 50 items in this screener, and each item is awarded a mark. The average score obtained by those who qualified was 46.8 (*SD* = 1.44).

Stimuli

The stimuli consisted of 2,500 commonly used single Chinese characters and 2,500 pseudocharacters. The 2,500

---

[3] The numerical breakdown of participants excluded on the basis of poor performance on the screening tasks or lexical decision task is as follow: 1 participant was eliminated on the basis of The Chinese Author Recognition Test, 2 were eliminated on the basis of The HSK Chinese Proficiency Test, and 8 participants scored less than 85% accuracy on the lexical decision task and thus, their data were discarded.

characters were randomly selected from the Lu (1989, 1992) teaching corpus using the Research Randomizer (Urbaniak & Plous, 2011). These selected characters were also verified against the "List of Commonly Used Characters" released by the mainland Chinese State Language Commission and News Bureau (1988). This was done to further ensure that the selected characters were not foreign to the mainland Chinese participants.

These 2,500 genuine Chinese characters were first coded for their composition—that is, the breakdown of the components and their positions within each character. For instance, 煌 is made of the semantic component 火 and the phonetic component 皇, and they are positioned on the left and right, respectively. *The Modern Chinese Dictionary* (Institute of Linguistics in the Chinese Academy of Social Sciences, 2008), *A Compendium of Chinese Characters* (Gu, 2007), and a resource on Chinese character etymology (说文解字考证; Dong, 2005) were consulted during the coding. Characters were then randomly paired up, paving way for the creation of pseudocharacters. The Research Randomizer was used to facilitate this random pairing-up (Urbaniak & Plous, 2011).

The 2,500 pseudocharacters were created by switching semantic components (部首, bù shǒu) within each pair. Care was also taken to ensure that the position of the component was maintained during swapping. This meant that the semantic component initially found on the left would be swapped with another semantic component also located on the left. For instance, the semantic components of 煌 and 狂 were exchanged to form two pseudocharacters, in which 火 was now paired with 王 and 犭 with 皇. This procedure controls for structural legality, while maintaining the component parts and number of strokes across characters and pseudocharacters. Each pseudocharacter was verified against four sources to determine that it did not exist within the Chinese lexicon. The four sources were (1) the *Modern Chinese Dictionary* (The Institute of Linguistics in the Chinese Academy of Social Sciences, 2008), (2) a resource on Chinese character etymology (Dong, 2005), (3) Xin Hua online dictionary (which contains the ancient Chinese corpus; http://xh.5156edu.com), and (4) nciku Chinese dictionary (an online repository; www.nciku.com).

Procedure

Each participant was tested in three sessions on separate days. Each session lasted no more than 2 h. At the start of the first session, both language screening tasks were administered. Those who qualified proceeded to the lexical decision task.

The entire lexical decision task consisted of 5,000 trials. The trials were divided into three blocks (1,700, 1,700, and 1,600 trials, respectively), with equal numbers of characters and pseudocharacters within each block. The blocks were counterbalanced using a Latin-square. The trials within each block were also randomized by the E-Prime program anew for each participant. There was an optional 2-min break after every 100 trials. To acquaint the participants, there were 40 practice trials at the start of each session, the data from which were not analyzed. None of the items used in practice appeared in the experimental trials.

During the lexical decision task, each participant sat in front of a computer, which had E-Prime 1.2 installed (Schneider, Eschman, & Zuccolotto, 2002). They were instructed to pay close attention to the display screen, read each character silently, and make corresponding keyboard presses using the index fingers of the respective hands (the right [”] button was labeled "Character (字)" and the left [A] button, "Pseudocharacter (非字)"). They were encouraged to respond quickly, but not at the expense of accuracy. All instructions were printed on-screen in Chinese.

The sequence for each lexical decision trial was as follows. First, a fixation cross was displayed at the center of the monitor for 500 ms, followed by a 120-ms blank screen, and finally, the target was presented. All the real characters and pseudocharacters were printed in the 36-point Song font. The target would remain online until a response was made. For each incorrect trial, a "ding" sound was played directly after the response.

Results

The results are based on the response latencies from the character trials. Pseudocharacter trials were not analyzed. Reaction times from inaccurate responses were also excluded. The overall performance based on the 35 participants was strong, each scoring well above the 85 % accuracy threshold. Their mean accuracy rates for characters and pseudocharacters were 95.23 % (*SD* = 1.65 %) and 95.14 % (*SD* = 1.93 %), respectively. Responses slower than 3,000 ms or faster than 200 ms were considered extreme and discarded, after which the mean and standard deviation were calculated for each individual. Response latencies either above or below 2.5 standard deviations from the respective individual mean were then removed. In total, 7.84 % of the latencies were trimmed. The eventual mean response time for the correct character trials was found to be 601.70 ms (*SD* = 80.24 ms).

The reaction time data are published in the attached Excel file (Chinese Lexicon Project Sze et al.xlsx; available online from the section titled "Electronic Supplementary Materials" found below the article) for noncommercial use by researchers. The Excel file contains the following columns:

Character:	the Chinese character in simplified script
Acc:	mean accuracy

Ntrials:	number of participants whose trials were sufficiently reliable to provide the latencies for that item (maximum being 35)
RT:	mean response time for the item, computed across participants
SE:	standard error of the response time for each character
SD:	Standard deviation of the response time for each character
Z(RT):	Mean of the standardized response time. In line with the procedure described by Faust, Balota, Spieler, and Ferraro (1999), all raw response times were transformed into *z* scores for each participant, before averaging across the participants for each character to yield the individual item's *Z*(RT).

Information in these columns is also reproduced in both csv (encoded in Unicode UTF-8) and pdf file formats. The pdf is provided as an extra resource, in case users experience trouble opening the Chinese characters in the other formats, due to font loading or other software installation problems.

Applying the data from the Chinese Lexicon Project

The lexical decision latencies collected are valuable for addressing different research questions. In the next section, we illustrate how the data can be fruitfully applied to two research issues in visual character recognition research.

Analysis 1: selecting an optimal Chinese character frequency measure

Character frequency has consistently been shown to be the strongest predictor of Chinese character recognition times (e.g., mega-naming results by Liu et al., 2007). Consequently, character frequency is an important variable for researchers to control in experimental design and statistical analyses. It is therefore important to identify which frequency count serves as the predictor of character recognition performance.

There are a number of Chinese character frequency norms available in the literature. However, their quality could vary because of differences in their corpora (Burgess & Livesay, 1998). Wang (2001), Feng (2002), and Liu (2009) can be consulted for an overview on existing Chinese corpora. Here, we review seven of the norms that could be publicly accessed (there are norms whose access is restricted; also, only corpora based on simplified script are considered in this study). The characteristics of these seven norms and their corpora are summarized in Table 1. To date, the most popular measure used by mainland researchers is the *Dictionary of Modern Chinese Frequency* (Language Teaching and Research Institute of Beijing Language and Culture University, 1986). A site search on Acta Psychologica Sinica (心理学报; http://journal.psych.ac.cn/xuebao/en/dqml.asp), the flagship journal for psychological science in mainland China, revealed that there were at least 12 articles citing the use of the *Dictionary of Modern Chinese Frequency*, as compared with two counts for CCL(PKU) and zero for the rest, from July 2010 to May 2012.

An evaluation of some of these Chinese frequency measures was previously conducted by Cai and Brysbaert (2010) to validate their new subtitle frequency measure. They reported that their subtitle index was "slightly better than the other measures" (p. 5). In analysis 1, we attempt to extend their findings. First, Cai and Brysbaert evaluated the character frequency measures using behavioral data from speeded pronunciation (in this case, Liu et al.'s (2007) mega-naming data were used). For English, it is well-established that frequency effects are stronger in lexical decision than in speeded pronunciation (Balota et al., 2004; Scarborough, Cortese, & Scarborough, 1977), possibly because of lexical decision's emphasis on familiarity-based information, which is associated with the frequency of occurrence (Balota & Chumbley, 1984). Speeded pronunciation, on the other hand, emphasizes the computation of phonology and is, therefore, largely influenced by the stimuli's phonological characteristics (Cortese, 1998; Reynolds & Besner, 2006). Thus, lexical decision performance should, in principle, be a more sensitive measure for discriminating between different frequency norms.

Second, Cai and Brysbaert (2010) compared three measures (four when they counted subtitles' raw and contextual diversity frequencies separately). The other two measures studied were the LCSMCS and CCL(PKU). Omission of the highly popular *Dictionary of Modern Chinese Frequency* measure is surprising, given how ubiquitously it is cited—a detail acknowledged by Cai and Brysbaert themselves.

Lastly, since the publication of Cai and Brysbaert's (2010) paper, an influential new measure has emerged. The "cncorpus" norm (Institute of Applied Linguistics, 2009, 2010; www.cncorpus.org) is a state-sanctioned frequency measure based on a national corpus compiled by the State Language Commission, the government agency in charge of Chinese language reforms. It seemed worth examining how well other frequency measures hold up against this official norm.

Evaluating the seven frequency measures

To enhance the reliability of this assessment, we first excluded items that did not achieve at least a mean accuracy rate of 70 % (66.67 %, being 1 standard deviation above chance). Out of the full set of 2,500 characters, 95 % (2,379 characters) fulfilled this requirement and were retained for further analyses. Then, in an effort to be comprehensive, all seven measures in Table 1 (or eight when subtitles' raw and

Table 1 Frequency measures of Chinese characters in simplified script

Frequency measures
Books
(1) *Dictionary of Modern Chinese Frequency* (现代汉语频率词典) (1986)
Author: Language Teaching and Research Institute of Beijing Language and Culture University (北京语言学院语言教学研究所)
Corpus size: 1.8 million characters
Materials included in the corpus could be classified into (1) political and current affairs (*People's Daily*, writings by politicians like Mao Zedong, Zhou Enlai, Ye Zhaoying, etc.), (2) scientific works (periodicals like *Beijing Science* and *Technology News*, etc.), (3) theatrical works and daily conversations (Cao Yu's "Sun Rise," selections from Hou Bao Lin's crosstalk, 10 excerpts from people's daily conversation etc.), (4) novels and short stories (e.g., excerpts from Mao Dun's "Midnight," Ding Ling's "The Sun Shines on the River Sang Gan," and Yang Mo's "Song of Youth," etc., and all readings from the Chinese Language textbooks, from primary to high school)
(2) *A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners*
Authors: Xiao, Rayson, & McEnery (2009) (an online version can be obtained from the first author through www.corpus4u.org)
Corpus size: 73 million characters
Built upon the Lancaster Corpus of Mandarin Chinese and UCLA Written Corpus. Includes these four registers: (a) spoken (daily conversations, telephone calls, radio broadcasts etc.), (b) news (Xin Hua agency's newswire texts in 1995, *People's Daily* 1998 and 2000), (c) fiction, and (d) nonfiction
Online corpora compiled by research institutes
(3) Modern Character Frequency List by cncorpus (www.cncorpus.org), referred to as cncorpus in the main text
Author: Institute of Applied Linguistics in the Ministry of Education (教育部语言文字应用研究所计算语言学研究室)
Corpus size: 1.1 billion characters and growing. Pledged to add about 3 million new characters annually
Frequency count was based on the State Language Commission's Modern Chinese Corpus (国家语委现代汉语语料库)
Boasts of a comprehensive sampling of diverse materials (print/online) in almost every discipline since 1919
(4) Center of Chinese Linguistics' (Peking University) Character Frequency List, referred to as CCL(PKU) in the main text
Author: Center of Chinese Linguistics in Peking University (北京大学汉语语言学研究中心)
Corpus Size: 307 million characters in the Modern Chinese Corpus (现代汉语语料库)
Materials include major periodicals and newspapers, as well as literary works. Details are found in http://ccl.pku.edu.cn:8080/ccl_corpus/
(5) Language Corpus System of Modern Chinese Study (LCSMCS) (现代汉语研究语料库查询系统; http://www.dwhyyjzx.com/cgi-bin/yuliao/)
Author: Center for Studies of Chinese as a Second Language (北京语言文化大学对外汉语研究中心)
Corpus size: 660 million characters according to Liu, Shu, & Li (2007)
Part of this corpus is also found in the Chinese Single-Character Word Database (www.personal.psu.edu/pul8/psylin_norm/psychnorms.html)
Online corpora compiled by individuals
(6) Character Frequency List of Modern Chinese
Author: Da, J. (2004)
Corpus size: 193 million characters
Harvested electronic texts and digitized hard copies from 16 Web sites between 1997 to 2003 (newspaper/medical/museum/literary, etc.) (see http://lingua.mtsu.edu/chinese-computing/docs/report/ for details)
(7) Chinese Character Frequency Based on Film Subtitles (SUBTLEX-CHR)
Authors: Cai and Brysbaert (2010)
Corpus size: 46.8 million characters; 6,243 contexts
Based on Chinese subtitles from films and television shows. Subtitle texts were mined from DVDs and the two biggest Chinese Web sites supplying subtitles in the simplified script. Two character frequency measures are derived—namely, (a) subtitle (raw) frequency (i.e., how many times a character actually occurs (token count) and (b) subtitle (contextual diversity) frequency (i.e., the number of films and television shows a character occurs in)

contextual diversity frequencies are counted separately) were examined. Except for LCSMCS, CCL(PKU), and subtitle frequency (raw and contextual diversity), none of the others have ever been empirically tested. Out of the 2,379 eligible characters, only 1,273 (53.51 %) had a frequency count specified in each of the seven frequency norms. To keep the analyses comparable, only latencies from these 1,273 items will be considered in this section.

Each measurement count was first log-transformed. We then computed the proportion of variance in lexical decision response times accounted for by each frequency measure. The proportions are presented in Table 2.

The results demonstrated that subtitle (contextual diversity) accounted for the most variance, but was followed closely by several other measures, before the national norm, cncorpus. Importantly, the most popular frequency of

Table 2 Percentage of variance accounted by each frequency norm for 1,273 characters

Measure	Variance (%)
Dictionary of Modern Chinese frequency (1986)	12.61
Modern Chinese corpus character frequency list by cncorpus	26.34
Subtitle (raw) character frequency	27.62
CCL(PKU) character frequency list	27.83
Character Frequency List of Modern Chinese by Da, J. (2004)	28.20
A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners	29.27
Language Corpus System of Modern Chinese Study (LCSMCS)	29.47
Subtitle (contextual diversity) character frequency	30.69

Data accessed on 31 May 2012

choice, the *Dictionary of Modern Chinese Frequency*, fared worst by a large margin. It explained only less than half of the variance its weakest competitor could account for (12.61 ÷ 26.34 = 0.48), a difference that was also significant ($p < .001$).

To examine whether there were any reliable differences in terms of the variance explained by the top three frequency measures, we ran a series of hierarchical regressions. When LCSMCS was tested against *A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners*, the ΔR^2 was small at .008 but significant ($p < .001$). Similarly, when subtitle (contextual diversity) was tested against *A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners* and LCSMCS, the ΔR^2s were modest at .033 and .032, respectively, but the differences were significant ($ps < .001$). Thus, statistically, Cai and Brysbaert's (2010) subtitle (contextual diversity) explained significantly more variance than its closest rivals.

Implications

The results in Table 2 caution against the use of the *Dictionary of Modern Chinese Frequency*. Its underlying corpus is rather small (1.8 million characters) and was compiled back in the 1980s; thus, it may lack contemporary relevance. Unfortunately, this frequency dictionary is still heavily utilized by researchers, especially in mainland China. Reliance on an inferior frequency norm will lead to weak experimental manipulation, since the experimental matching of stimuli will not be reliable. Poor frequency manipulation could even lead to spurious outcomes (Brysbaert & New, 2009; Gernsbacher, 1984).

Subtitle (contextual diversity) turned out to be the strongest measure. This analysis based on character lexical decision latencies provides additional support for Cai and Brysbaert's (2010) initial validation using naming latencies. It also demonstrates the superiority of operationalizing frequency using contextual diversity, instead of the raw numerical count. This contextual diversity metric calculates the number of films and television shows a character appears in. The results are consistent with the findings obtained in English (for theoretical explanations on why contextual diversity is better, see Adelman & Brown, 2008; Adelman, Brown, & Quesada, 2006; Brysbaert & New, 2009).

Finally, it is interesting to observe that the government-sanctioned cncorpus, while delivering a modest result, was outperformed by other measures (e.g., the *Character Frequency List of Modern Chinese* by Da, 2004). A close inspection of Table 1 offers some clues why this might be so. (1) The cncorpus has over a billion characters (Institute of Applied Linguistics, 2009), easily topping the charts in terms of corpus size. (2) Its sampling is extremely extensive. It sampled from almost every discipline and covered almost any printed/online text since 1919. These two characteristics fit Brysbaert and New's (2009) analysis in English, on how excessively large corpus sizes could yield diminishing returns. After partitioning the British National Corpus into various sizes, these researchers found that a corpus size of 16–30 million suffices for a reliable frequency norm. Beyond 30 million, there might be diminishing returns when infrequent texts were included, inflating character counts for scarce items. This might explain why the result for the cncorpus was relatively poor: It became the victim of its own size.[4]

Analysis 2: replicating previous findings with virtual studies

Another major advantage associated with large-scale repositories of behavioral data is that it allows virtual studies to be carried out. In this second analysis, we attempt to replicate previous lexical decision findings within the single-character recognition literature. Through these virtual experiments, we are able to ascertain the validity and generalizability of the data in the Chinese Lexicon Project.

As was mentioned in the introduction, there are relatively few published studies on single-character recognition that have employed the lexical decision task. Furthermore, not many studies provide the specific stimuli used. Within these constraints, we managed to identify three papers on lexical decision with their complete stimuli appended. They are (1) Leong et al. (1987, Experiment 2), (2) Peng, Deng, and Chen (2003, Experiment 1), and (3) B. Chen, Dent, You, and Wu (2009, Experiment 3).

Leong, Cheng, and Mulcahy (1987) examined whether there was a relationship between the number of strokes in a

---

[4] We should point out that cncorpus was not created solely for psycholinguistic research. As a national corpus, it probably serves other functions—for example, providing historical linguists information on the evolution of character use and so forth. One recommendation would be for the corpus to include an option for users to select, say, which period and type of text to begin computing character/word statistics from, thus reducing dead weight.

Table 3 Visual lexical decision findings in original and virtual studies based on the same set of stimuli that are common in both Chinese Lexicon Project and the respective original study

	Effect	Examples of Stimuli	Number of Stimuli (Original/Replication)	RT in Original Study (ms)	RT in Virtual Study (ms)
Leong, Cheng, & Mulcahy (1987)	Few strokes	和没扒吠呐	23/15	1,045	571
	Many strokes	就解慷塌寥	23/9	1,139	602
	Effect of stroke			94 **	31 **
Peng, Deng, & Chen (2003)	High frequency, Multiple meanings	别高深论批	17/14	449	553
	High frequency, Single meaning	很你最建有	17/13	457	551
	Low frequency, Multiple meanings	绝周派落信	17/11	446	539
	Low frequency, Single meaning	该够距抓江	17/11	492	558
	Frequency × polysemy			38 ***	21 †
Chen, Dent, You, & Wu (2009)	Early AoA	枪护扩亮奖	36/38	585	562
	Late AoA	纵观眠寓驰	36/40	595	578
	Effect of AoA			10 *	16 **

* $p < .05$ for F_1 and F_2

** $p < .01$ for F_1

*** $p < .01$ for F_1 and F_2

† $p < .07$ for F_1

character and the time taken to recognize it. Across two studies (speeded pronunciation and lexical decision), both skilled and less skilled Chinese readers took a significantly longer time to recognize characters with more strokes. Correspondingly, in our virtual replication with the same stimuli, we should also observe a similar main effect of stroke count. Peng et al. (2003), on the other hand, sought to investigate whether frequency interacted with effects of polysemy. Interactions between the two were previously reported in other languages and in Chinese words (double-characters), but the researchers wanted to examine whether the same phenomenon could be found at the level of single characters. Peng et al.'s data did reveal an interaction, such that the participants were significantly faster for characters with multiple meanings than for those with a single meaning, but this differentiation was present only for the low-frequency characters. This interaction should be reproduced in our virtual experiment.

Finally, B. Chen et al. (2009) conducted three studies (a tachistoscopic task, a visual duration threshold task, and a lexical decision task) to evaluate whether the age-of-acquisition (AoA) effect is present during visual character processing. An AoA main effect was found in each of the tasks. Likewise, in our virtual replication, we should also expect an AoA main effect, with later-acquired characters yielding significantly longer latencies.

Stimuli and results

The same characters employed in the original studies were selected as stimuli in the respective replications. However, not all of the original characters could be located in the Chinese Lexicon Project. The number of original characters for each study that could be extracted from our dataset is stated in Table 3. It should be noted that characters in the traditional script were used in the original Leong et al.'s (1987) study. However, characters can be identical across traditional and simplified scripts. We only selected such identical characters in the replication study, which explains the slight discrepancy in the number of stimuli between the original Leong et al. study and its replication.[5, 6]

All three studies were within participants in design. ANOVA analyses parallel to the original research were conducted in the respective replications, and the results are displayed in Table 3. To facilitate comparisons, Table 3 also

---

[5] In the virtual replication of Leong et al. (1987), 数 (數 in traditional script) was included in the stimuli as a character with many strokes. The inclusion of 数 does not violate the stroke manipulation, since its simplified form has 13 strokes, which is above the cutoff of "many strokes" (Leong et al.'s cutoff is placed at 12). The traditional 數 has 15 strokes. In any case, we also ran an additional series of analyses that excluded 数. The same pattern of findings was elicited [response time: $F_1(1, 34) = 5.26$, $MSE = 1{,}985.36$, $p < .03$, $\eta_{\text{partial}} = .13$].

[6] B. Chen et al. (2009) created two sets of stimuli for their three experiments (Experiment 1 [tachistoscopic task] used stimuli set 1, while Experiments 2 and 3 [visual duration threshold and lexical decision tasks] used stimuli set 2). Both sets of stimuli were created on the basis of the same design and requirements. These two sets of stimuli were thus combined for the drawing of our stimuli. Out of the 128 unique characters found in the combined set (61 early AoA, 67 late AoA; there were repetitions in the two stimuli sets), 78 of them are present in the Chinese Lexicon Project (38 early AoA, 40 late AoA). The value 78 is rather close to the original number of stimuli used in Chen et al.'s lexical decision experiment (72 characters were chosen as stimuli—i.e., 36 early AoA, 36 late AoA).

includes the original descriptive statistics and main findings published by the three original papers.

From the table, it can be seen that the data sampled from the Chinese Lexicon Project replicated the published findings in a relatively faithful manner. All the critical trends and significant effects were found. These include a replicated main effect of stroke for Leong et al. (1987), an AoA main effect for B. Chen et al. (2009), and the frequency × polysemy interaction for Peng et al. (2003), such that responses for low-frequency characters were differentiated by the characters' number of meanings, $F_1(1, 34) = 7.00$, $MSE = 937.08$, $p = .012$, $\eta^2_{\text{partial}} = .17$; $F_2(1, 20) = 1.15$, $MSE = 1822.36$, n.s., but no such distinction was present for the high-frequency characters, $F_1(1, 34) = 0.04$, $MSE = 1{,}295.92$, n.s.; $F_2(1, 25) = 0.02$, $MSE = 1{,}683.53$, n.s.).

It should perhaps be pointed out that the original participants in Leong et al. (1987) appeared to take a slightly longer time to perform the tasks than did our participants. This is probably because the original 1987 data were collated from both skilled and less skilled Chinese readers. The participants in the Chinese Lexicon Project, however, were highly proficient Mandarin readers. According to Leong et al., if only the skilled readers were considered, the descriptive statistics for few and many strokes should be lowered to 872 ms ($SD = 146$) and 946 ms ($SD = 203$), respectively. These latencies would be more comparable to the values found in the replication. Nonetheless, our overall ability to replicate findings across three different lexical variables, each previously obtained by different researchers using independent samples and separate stimuli sets, is strong evidence that the data from the Chinese Lexicon Project do reflect processing by typical readers. This generalizability is noteworthy in spite of the relatively strict standards applied during language screening.

Conclusion

In this article, we introduced the Chinese Lexicon Project, a repository of lexical decision latencies collected for 2,500 single characters in simplified script. To demonstrate some of its potential usages, seven competing frequency metrics were evaluated, and three virtual simulations were conducted. We choose to commence with single characters, since they are the fundamental building blocks of the Chinese lexicon: All Chinese words, beyond monosyllabic ones, are essentially compounds of single characters (Liu, Wang, & Zhou, 2009, pp. 62–78; Sun, 2006, pp. 45–74). Additionally, out of the different dialectal varieties of Chinese, we begin with a sample of native Mandarin speakers. This is because Mandarin or Standard Chinese is the official language of mainland China, and it is widely used and internationally spoken (Ostler, 2008, estimated that there are around 1,055 million Mandarin speakers and 760 million for English). Currently, there is no similar large-scale dataset that is easily accessible for researchers and scholars interested in the Chinese language (or nonalphabetic languages for that matter) to utilize. Combined with the scarcity of empirical work involving Chinese single-character lexical decision, this project should make a valuable empirical and methodological contribution to the literature.

References

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

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database on CD-ROM*. Philadephia, PA: Linguistic Data Consortium.

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*(3), 340–357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology. General, 133*(2), 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*(3), 445–459.

Balota, D. A., Yap, M. J., Hutchison, K.A., & Cortese, M. J. (2012). Megastudies: Large scale analysis of lexical processes. In James S. Adelman (Series and Vol. Ed). *Visual word recognition, Volume 1: Models and methods, orthography and phonology* (pp. 90–115). Hove, East Sussex: Psychological Press.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2001). Age-of-acquisition and word frequency in the lexical decision task: Further evidence from the French language. *Current Psychology of Cognition, 20*(6), 401–443.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kucera and Francis. *Behavior Research Methods, Instruments, & Computers, 30*(2), 272–277.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One, 5*(6), e10729. doi:10.1371/journal.pone.0010729

Chen, B., Dent, K., You, W., & Wu, G. (2009). Age of acquisition affects early orthographic processing during Chinese character recognition. *Acta Psychologica, 130*(3), 196–203.

Chen, B., & Peng, D. (2001). 汉语双字多义词的识别优势效应 [The effects of polysemy in two-character word identification]. *Acta Psychologica Sinica, 33*(4), 300–304.

Chen, H.-C., & Zhou, X. (1999). Processing East Asian languages: An introduction. *Language & Cognitive Processes, 14*(5/6), 425–428.

Cortese, M. J. (1998). Revisiting serial position effects in reading. *Journal of Memory and Language, 39*(4), 652–665.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945.

Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In P. Zhang, T. Xie, & J. Xu (Eds.), *Proceedings of the 4th International Conference on New Technologies in Teaching and Learning Chinese: The studies on the theory and methodology of the digitized Chinese teaching to foreigners* (pp. 501–511). Beijing: The Tsinghua University Press.

Dong, L.-C. (2005). 说文解字考证 [An investigation of Chinese characters' etymology]. Beijing: Writer's Publishing House.

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*(6), 777–799.

Feng, Z. (2002). 中国语料库研究的历史与现状 [Evolution and present situation of corpus research in China]. *Journal of Chinese Language and Computing, 12*(1), 43–62.

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*(2), 488–496.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology. General, 113*(2), 256–281.

Gu, J.-P. (2007). 字解: 字形图解字典 [A compendium of Chinese characters]. Singapore: Chinese Heritage Lodge.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Institute of Applied Linguistics. (2009). 国家语委现代汉语语料库介绍 [Introduction to the Modern Chinese corpus (cncorpus) by the State Language Commission]. Retrieved March 29, 2011, from Chinese Linguistic Data web site: www.cncorpus.org

Institute of Applied Linguistics. (2010). 现代汉语语料库汉字频率表 [Modern Chinese corpus character frequency list]. Retrieved March 29, 2011, from Chinese Linguistic Data web site: www.cncorpus.org

Institute of Linguistics in the Chinese Academy of Social Sciences. (2008). 现代汉语词典 [Modern Chinese dictionary] (5th ed.). Beijing: The Commercial Press.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam: Elsevier North-Holland.

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. doi: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*(1), 287–304.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

Language Teaching and Research Institute of Beijing Language and Culture University. (1986). 现代汉语频率词典 [Dictionary of modern Chinese frequency]. Beijing: Beijing Language and Culture University Press.

Lee, S.-Y., & Krashen, S. (1996). Free voluntary reading and writing competence in Taiwanese high school students. *Perceptual and Motor Skills, 83,* 687–690.

Leong, C. K., Cheng, P.-W., & Mulcahy, R. (1987). Automatic processing of morphemic orthography by mature readers. *Language and Speech, 30*(2), 181–196.

Li, P., Tan, L. H., Bates, E., & Tzeng, O. J. L. (2006). Introduction: New frontiers in Chinese psycholinguistics. In P. Li (Series and Vol. Ed.), L.H. Tan, E. Bates, & O. J. L. Tzeng (Vol. Eds.), *Handbook of East Asian psycholinguistics: Vol. 1. Chinese (pp. 1–9)*. Cambridge, UK: Cambridge University Press.

Liu, Y. (2009). 汉语词汇研究统计述评 [A review of Chinese vocabulary statistic studies]. *Chinese Language Learning, 30*(1), 62–69.

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods, 39*(2), 192–198.

Liu, Y., Wang, R. D., & Zhou, H. (2009). 现代汉语概论（留学生版） [Modern Chinese: An overview]. Shanghai: Shanghai Educational Publishing House.

Lu, S. C. (1989). 字词频率词典（以拼音为序）: 新加坡《小学华文教材》 [Frequency dictionary of Chinese characters, words and phrases used in Singapore primary school textbooks]. Singapore: Center of Research for Chinese, National University of Singapore.

Lu, S. C. (1992). 字词频率词典（以拼音为序）: 新加坡《中学华文教材》 [Frequency dictionary of Chinese characters, words and phrases used in Singapore secondary school textbooks]. Singapore: Center of Research for Chinese, National University of Singapore.

Myers, J., Huang, Y.-C., & Wang, W. (2006). Frequency effects in the processing of Chinese inflection. *Journal of Memory and Language, 54*(3), 300–323.

Ostler, N. (2008). World languages. In P. K. Austin (Ed.), *1000 languages: The worldwide history of living and lost tongues* (pp. 10–34). UK: Thames & Hudson.

Peng, D., Deng, Y., & Chen, B. (2003). 汉语多义单字词的识别优势效应 [The polysemy effect in Chinese one-character word identification]. *Acta Psychologica Sinica, 35*(5), 569–575.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a 'universal' phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227–248). Amsterdam: Elsevier North-Holland.

Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology. Human Perception and Performance, 32*(6), 1303–1323.

Rogers, H. (2005). *Writing systems: A linguistic approach*. Malden, MA: Blackwell Publishing.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior, 9*(5), 487–494.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology. Human Perception and Performance, 3*(1), 1–17.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime (Version 1.2) [Computer software]*. Pittsburgh: Psychology Software Tools Inc.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*(1), 1–30.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*(4), 584–615.

Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology, 85*(2), 211–229.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*(4), 402–433.

State Language Commission & News Bureau. (1988). 现代汉语通用字表 [List of Commonly Used Characters]. Retrieved June 30, 2011, from http://www.china-language.gov.cn/wenziguifan/shanghi/014c.htm

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*(3), 337–359.

Sun, C. (2006). *Chinese: A linguistic introduction*. New York: Cambridge University Press.

Tsai, P.-S., Yu, B. H.-Y., Lee, C.-Y., Tzeng, O. J. L., Hung, D. L., & Wu, D. H. (2009). An event-related potential study of the concreteness effect between Chinese nouns and verbs. *Brain Research, 1253,* 149–160.

Urbaniak, G. C., & Plous, S. (2011). Research Randomizer (Version 3.0) [Computer software]. Retrieved on January 1, 2011, from http://www.randomizer.org/

Wang, J. (2001). Recent progress in corpus linguistics in China. *International Journal of Corpus Linguistics, 6*(2), 281–304.

Xiao, R., Rayson, P. A., & McEnery, T. (2009). *Frequency dictionary of Mandarin Chinese: Core vocabulary for learners*. London: 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*(4), 992–1003.

Yin, B., & Rohsenow, J. S. (1994). *Modern Chinese characters*. Beijing: Sinolingua.

Yip, M. (2002). *Tone*. New York: Cambridge University Press.

You, W., Chen, B., & Dunlap, S. (2009). Frequency trajectory effects in Chinese character recognition: Evidence for the arbitrary mapping hypothesis. *Cognition, 110*(1), 39–50.