Psycholinguistic Norms for 3,783 Two-Character Words in Simplified Chinese

Dangui Song¹ and Degao Li²

Abstract
Over 70% of the more than 56,000 most frequently used words in simplified Chinese are two-character words (2C-words). The present study collected data on subtitle frequency, number of strokes, number of meanings, familiarity, concreteness, imageability, age of acquisition, subjective frequency, subjective number of meanings (sNOMs), compositionality, emotional experience rating, sensory experience rating (SER), and semantic transparency (ST) for 3,783 commonly used 2C-words. Correlative patterns were identified between the 13 variables, all of which, with the exception of imageability, SER, and sNOM, were significantly predictive of the changes in participants’ RTs in LDs on the target words. In conclusion, skilled readers’ awareness of 2C-words’ features such as concreteness, imageability, compositionality, sNOM, and ST is closely associated with their semantic perception of the constituent characters.

Keywords
Chinese two-character words, normative features, association, semantic perception, constituent characters

Introduction
Normative descriptions of words or expressions are available in many languages, including Chinese (Chang et al., 2016), Dutch (Hubers et al., 2019), English (Juhasz et al., 2015), French (Bonin et al., 2015), German (Schulte Im Walde & Borgwaldt, 2015), Hebrew (Yael et al., 2015), Italian (Barca et al., 2002), Malay (Yap et al., 2010), Russian (Grigoriev & Oshhepkov, 2013), Spanish (Guausch et al., 2016), and Swedish (Blomberg & Öberg, 2015). Probably due to the complexity of vocabulary per se, this kind of study is typically conducted on samples of words or expressions in a language. The present study attempts to provide a normative description for 3,783 Chinese two-character words (2C-words).

Chinese 2C-Words
In Chinese, the basic writing units are Chinese characters, almost all of which are meaningful in their own right. Many Chinese characters can be used as one-character words (1C-words), but most Chinese characters can join with one or more other Chinese characters to form words that are more than one Chinese character in length. Indeed, 72% and 22.3% of the 56,008 frequently used words (State Language Affairs Commission, 2008) are 2C-words and words of more than two Chinese characters, respectively.

On the other hand, what is most prominent with 2C-words, for example, seems to be their semantic relationships with the constituent characters (Lv, 1979). A corpus-based study suggests that the constituent characters in a 2C-word have a probability of over .80 to mean the same as when they are in isolation (Yuan & Huang, 1998). According to Ge (2018), the meaning of a 2C-word can be semantically related to its constituent characters in one of six ways: The meaning of a 2C-word can be the simple composition of its constituent characters, a 2C-word’s meaning can be inferred from its constituent characters, a 2C-word largely means the same as one of its constituent characters, a 2C-word is simply the semantic repetition of one Chinese character, a 2C-word has nothing to do with its constituent characters in meaning and a 2C-word’s constituent characters themselves are not meaningful at all.

On the other hand, as will be briefly reviewed in the next sub-section, Chinese characters, 2C-words, and words of more than two Chinese characters have normative features. Thus, in addition to the degrees of semantic overlap, there might be different strengths of association between a 2C-word and its constituent characters in features. For example, high-frequency 2C-words are likely to be composed of high-frequency Chinese characters, and low-frequency 2C-words of low-frequency Chinese characters.

¹Zhejiang University, Hangzhou, China
²Qufu Normal University, Qufu, China

Corresponding Author:
Degao Li, College of Chinese Language and Literature, Qufu Normal University, No. 57, Jingxuan Road, Qufu City, Shandong Province 273165, China.
Email: li-degao@163.com
The possibility of complex relationships between 2C-words and their constituent characters in semantic overlap and in feature association may lead to special correlations between the words’ normative features. However, much of this concern remains to be settled at the present time.

Normative Studies in Chinese

Many studies have been conducted on the normative features of Chinese characters, 2C-words, and expressions of more than two Chinese characters in simplified and traditional Chinese (e.g., Chang et al., 2016; Li et al., 2016; Liu et al., 2007; Sun et al., 2018; Sze et al., 2015; Tsang et al., 2018; Tse & Yap, 2018; Tse et al., 2017; Wang et al., 2008, 2020; Yao et al., 2017; Yee, 2017). For example, Liu et al. (2007) investigated 15 features of 2,423 simplified Chinese characters. They suggested that for skilled readers, the reaction times (RTs) in naming may be predicted by cumulative frequency, phonological frequency, age of acquisition (AOA), number of word formations, number of components, number of strokes (NOSs), familiarity, concreteness, imageability, and regularity of Chinese characters. Similarly, Chang et al. (2016) examined 10 features of 3,314 traditional Chinese characters. The initial phonemes, semantic ambiguity, familiarity, frequency, consistency, and regularity of Chinese characters were significant predictors of naming RTs for skilled readers, but the semantic combinability, homophone density, phonetic combinability, and NOS were not.

Wang et al. (2008) examined five features (valence, arousal, dominance, appulsion, and familiarity) of 1,500 emotional 2C-words in simplified Chinese and showed that valence, arousal, dominance, and familiarity communicated 98% of the emotional information of these words. Yao et al. (2017) further studied the relationships between the affective (valence and arousal) and semantic (concreteness, imageability, context availability, and familiarity) attributes of emotional 2C-words. Consistent with Guasch et al. (2016), they found that valence and arousal were not predictive of RTs in lexical decisions (LDs) for skilled readers. Adopting the mega-study methodology, Tse et al. (2017) conducted a project on more than 25,000 2C-words in traditional Chinese and concluded that frequency was the strongest predictor of LD RTs for skilled readers. Tse and Yap (2018) also examined the constituent characters’ frequencies, NOS, phonological consistency, homographic density, and semantic transparency (ST) of 18,983 2C-words in traditional Chinese. The orthographic and semantic variables were more important than the phonological variables of the constituent characters in their influences on the recognition of 2C-words.

It is noteworthy that Tsang et al. (2018) conducted an elegant mega-study on 1,020 1C-words, 10,022 2C-words, 949 three-character words, and 587 four-character words. They showed that objective indexes such as word length and frequency of words of more than one Chinese character and cumulative character frequency, number of meanings (NOMs), NOSs, and number of pronunciations of Chinese characters were significantly predictive of the changes in LD RTs. Furthermore, they revealed a U-shape relationship between word-length and LD RTs. Similarly, Sun et al. (2018) conducted a lexical database study on 3,913 1C-words, 34,233 2C-words, 7,143 three-character words, and 3,355 four-character words in a naming task as well as an LD task.

Chinese characters, 2C-words, and expressions of more than two Chinese characters all have normative features that are similar to words or fixed expressions in other languages such as English. Meanwhile, however, there are studies that seem to have revealed special aspects of Chinese expressions. In addition to the features themselves such as NOS and phonology-related attributes that are particular of Chinese characters, for example, the U-shape correlation between word length and LD RT (Tsang et al., 2018) and the complex relations between Chinese four-character-idioms’ features and internal structures (Li et al., 2016) are promising to inspire studies into Chinese lexical processing. Considering the possible relationships between 2C-words and their constituent characters, one may argue that there should be particular aspects of 2C-words’ normative features. Studies in support of this argument will be of significance since most commonly used Chinese words are 2C-words.

The Present Study

Words are the basic units in the mental lexicon (Dronjic, 2011) and in sentence reading (Bai et al., 2008) by Chinese skilled readers, but a 2C-word can be perceived both as a whole entity and according its constituent characters in its early stage of recognition (Du et al., 2013; Liu et al., 2010; Peng et al., 1999; Tong et al., 2014; Wang et al., 2017; Yang, 2013). Indeed, several models (e.g., Peng et al., 1999; Taft & Zhu, 1997; Zhou & Marslen-Wilson, 2009) have been formulated on 2C-word recognition, which potentially indicate that the recognition of a 2C-word can be influenced by the character-level as well as by the word-level features. According to the multilevel interactive-activation framework (Taft & Zhu, 1997), for example, skilled readers are supposed to implement information processing at the levels of strokes, components, constituent characters, and words in 2C-word recognition. Similarly, the Inter- and Intra-Connection Model (Peng et al., 1999) and the Lexical Representation Framework for Chinese Compound Word (Zhou & Marslen-Wilson, 2009) assume inter-relationships between the representations for 2C-words and their constituent characters. In fact, it is repeatedly indicated that feature manipulations at the word level and at the level of constituent characters do seem promising to yield new insight into the mechanism of 2C-word recognition (Huang & Lee, 2018; Miwa et al., 2014; Sun et al., 2018).

As reviewed in the previous sub-section, frequency, familiarity, NOM, AOA, imageability, and concreteness are among those that are often examined to indicate the
normative characteristics of expressions consisting of one or more Chinese characters. Considering the relevant findings of both normative and psychological studies, an examination of NOS, compositionality, emotional experience rating (EER), sensory experience rating (SER), subjective frequency (sF), ST, and subjective number of meanings (sNOMs) would also be of a significant value.

NOS reflects the visual complexity of 2C-words. The efficiency of recognition of Chinese characters (Li et al., 2014; Liversedge et al., 2014; Sze et al., 2015) and Japanese kanjis (Tamaoka & Kiyama, 2013) tends to be influenced by NOS. SF is often measured by asking participants to estimate word usage frequencies on a Likert scale (Desrochers et al., 2010). It is a powerful predictor of participant performance in lexical processing (e.g., Balota et al., 2004; Chen & Dong, 2019). Indeed, normative descriptions of sF are available for words in alphabetic languages (Balota et al., 2004; Desrochers et al., 2010; Gonthier et al., 2009). Similarly, sNOM is measured by asking participants to rate the NOMs of a word on a Likert scale. It is an assessment of how many meanings that the users believe a word to have, and it can also be adopted as a variable in cognitive tasks (Chang et al., 2016).

Compositionality is often used in studies on words of more than one morpheme, and is defined as the degree to which the constituent morphemes contribute to the overall meaning of the whole expression (Cięslicka, 2013). In addition to experimental studies (Cięslicka, 2013), compositionality has been regarded as a variable in normative studies on complex expressions (Dasgupta et al., 2016; Li et al., 2016). It was taken into consideration in the present study because nearly all 2C-words can be regarded as composed of two morphemes of Chinese characters.

EER refers to the degree of ease with which a word arouses the user’s emotional experiences (Newcombe et al., 2012), and SER represents the degree of ease with which a word evokes sensory and perceptual experiences in the users (Juhasz et al., 2011). For example, “fountain” has a low EER score, whereas “relief” has a high score; “bag” scores low for SER, whereas “thirst” scores high. Both EER (Newcombe et al., 2012) and SER (Juhasz et al., 2011) have been shown to influence participants’ performance in cognitive tasks and are potentially important in cognitive processing according to theories of embodied cognition (Barsalou, 2008; Goldman & de Vignemont, 2009). In addition, normative data on SER are available in English (Juhasz et al., 2011) and French (Bonin et al., 2015). ST means the degree to which the constituents are semantically related to the word itself (Tse et al., 2017). It indicates the degree to which the word can be understood literally through its constituent(s), and it has been proved to be significantly influential for word recognition.

In summary, NOS is a particular feature of Chinese characters, but all the other 12 ones (frequency, NOM, familiarity, AOA, sF, sNOMs, concreteness, ST, imageability, compositionality, EER, and SER) have been commonly examined in many languages. One may expect to achieve similar results to those of previous studies, concerning the correlations between the 13 features and the features’ capability of predicting word recognition in LDs. Considering the possibility of complex relationships between 2C-words and their constituent characters, however, we propose two questions: What correlations between these 13 features might be special of 2C-words? Which of these features will be particular of 2C-words in predicting word recognition in an LD task?

Methods

This section consists of three sub-sections: word selection, feature scores collection and the implementation of an LD task. We expected to achieve an insight into the effects of these features on participants’ LD performance and thus have a deeper understanding of the inter-relationships between these normative characteristics. This study was approved by the Ethics Committee of Zhejiang University.

Word Selection

To determine those that were representative of the most frequently used 2C-words in simplified Chinese, we first turned to Modern Chinese Frequency Dictionary (Language Teaching and Research Institute of Beijing Language and Culture University, 1986) and HSK Chinese Proficiency Test Vocabulary Guideline (A Dictionary of Chinese Usage: 8,000 words) (Chinese Proficiency Oral Test Center of Beijing Language and Culture University, 2009). Modern Chinese Frequency Dictionary is a reference of word frequency for 8,000 of the most commonly used words in the 1980s, while the 8,822 words in HSK Chinese Proficiency Test Vocabulary Guideline (A Dictionary of Chinese Usage: 8,000 words), which are ranked into four categories according to usage frequency, are presumably the most common in use at the present time. The initially selected target words were the 4,197 2C-words that appear in both of these two resources and the 183 2C-words that fall into the first two categories in HSK Chinese Proficiency Test Vocabulary Guideline (A Dictionary of Chinese Usage: 8,000 words) but are not listed in Modern Chinese Frequency Dictionary. All these 4,380 words are compound words and belong to the word classes of nouns, verbs, or adjectives according to Modern Chinese Standard Dictionary (Li, 2014). By a compound word we mean a 2C-word that consists of two Chinese characters each of which is meaningful in its own right according to Ge (2018).

A word that is frequently used as a response in a word-association task should be more common in use than a word that is not often used as a response in such a task. Thus, as a second step to determine the target words, we subsequently conducted a task of single-word association with the initially selected words as the stimuli. We randomly divided them into 63 groups. Each group consisted of 69 to 75 words,
Table 1. Questionnaire Instructions.

| Feature | Instructions |
|---------|--------------|
| AOA     | For each word, please indicate when you first saw, heard, or used it by selecting one of the seven numbers next to it. Numbers 1, 2, 3, 4, 5, and 6 refer to when you were a student of grades 3, 5, 8, 9, 11, and 12, respectively, and 7 indicates when you were enrolled as a university student. The bigger the number, the later you think you acquired it. |
| Compositionality | For each word, please decide the extent to which both the constituent characters contribute to the overall meaning of the word as a whole by selecting one of the seven numbers next to it. The bigger the number, the more you think the constituent characters contribute to the overall meaning of the word. |
| Concreteness | For each word, please decide the extent to which it makes you think of specific entities or actions by selecting one of the seven numbers next to it. The bigger the number, the more easily you associate it with specific things or actions. |
| EER     | For each word, please rate the degree to which it evokes your emotional experiences (happiness, anger, sadness, pain, or fear) by selecting one of the seven numbers next to it. The bigger the number, the more easily your emotional experiences may be evoked. |
| Familiarity | For each word, please decide the extent to which you are familiar with it by selecting one of the seven numbers next to it. The bigger the number, the more familiar it is to you. |
| Imageability | For each word, please rate the degree to which it arouses mental images by selecting one of the seven numbers next to it. The bigger the number, the more familiar it is to you. |
| SER     | For each word, please decide the degree to which it arouses your sensory experiences (taste, touch, sight, sound, or smell) by selecting one of the seven numbers next to it. The bigger the number, the easier your sensory experiences may be aroused. |
| sF      | For each word, please rate the extent to which it is frequently used in your daily life by selecting one of the seven numbers next to it. The bigger the number, the more frequently you think it is used in everyday life. |
| ST      | For each word, please decide the extent to which it may be understood literally through its constituent character(s) by selecting one of the seven numbers next to it. The bigger the number, the more easily it may be understood through its constituent characters. |
| sNOM    | For each word, please indicate the number of meanings it has by selecting one of the three numbers next to it. Numbers 1, 2, and 3 indicate that it has one, two, and three or more meanings, respectively. |

which were printed on a single piece of paper. We obtained 50 copies of the 63 sheets of words, and 3,150 college students (1,727 females, 1,423 males; \( M = 20.3 \) years, age range: 18–25 years) participated in the survey. Each student received a list of stimuli and was required to write down the first word that they thought of as the response to each stimulus. In the end, 3,796 of the stimulus words were selected that also appeared at least twice among the responses.

**Feature Scores Collection**

**Objective features.** The NOM and NOS scores were obtained from *Modern Chinese Polysemous Words Dictionary* (Yuan, 2001) and *Xinhua Dictionary* (http://zidian.cibiao.com/), respectively. Following Tse et al. (2017), we took the score for each word from the corpus of contextual diversity subtitle frequency (Cai & Brysbaert, 2010) as the representative of the word’s objective frequency (referred to as subTF, hereafter).

**Subjective features.** Adults might be unable to remember the exact learning age for a specific word (Morrison et al., 1997), and they are likely to have acquired most words in written forms as the result of school education. Thus, similar to previous studies (Barca et al., 2002; Cortese & Khanna, 2008; Liu et al., 2007; Schock et al., 2012), we required participants to report in which period of school life they thought they learned the words. The 3,796 words were randomly divided into 44 groups, with each group consisting of 86 or 87 words. A seven-point Likert scale of AOA was designed for each group of words, which were listed on a single sheet with the instructions (see Table 1) printed at the top of the paper. Similarly, 44 scales were created to measure compositionality, concreteness, EER, familiarity, imageability, SER, ST, and sF.

According to *Modern Chinese Polysemous Words Dictionary* (Yuan, 2001), 62% and 96% of the 3,796 words have one meaning and not over three meanings, respectively. Thus, they were divided into 29 groups at random, with each group consisting of 130 to 133 words, and a three-point Likert scale of sNOM was designed for each group of words. In the end, we created 425 single-sheet scales.

It appeared acceptable to obtain around 15 scores in the evaluation of an attributive feature of the materials in a word-recognition study in Chinese (Yan et al., 2014; Zhang et al., 2013). Thus, we made 17 copies of the 425 single-sheet scales. The 7,225 sheets were randomly sent to 7,225 students from seven universities in China: 1,800 from Zhejiang University (945 females, 855 males; \( M = 21.7 \) years, age range: 18–24 years), 500 from Ningbo Dahongying University (236 females, 264 males; \( M = 22.2 \) years, age range: 18–25 years), 265 from South China Agricultural University (130 females, 135 males; \( M = 22.1 \) years, age range: 18–25 years).
range: 18–25 years), 1,160 from Ningbo University (562 females, 598 males; \(M = 21.4\) years, age range: 18–26 years), 300 from Guangdong Peizheng College (143 females, 157 males; \(M = 21.5\) years, age range: 18–23 years), 200 from Southwest University for Nationalities (110 females, 90 males; \(M = 21.6\) years, age range: 18–24 years), and 3,000 from China University of Petroleum (1,465 females, 1,535 males; \(M = 21.8\) years, age range: 18–26 years).

The rating task was conducted collectively in classroom settings with each sheet delivered to one student during the break times of classroom courses. The teachers were the experimenters who explained the instructions. Students were required to carefully read the instructions before starting to rate the stimuli. The raters did not participate in the above-mentioned task of single-word association.

The LD Task

Participants. One hundred and seventy college students (113 females, 57 males; \(M = 22.0\) years, age range: 18–28 years) were recruited on the campus of a university by means of a flyer advertisement. They were all native speakers of Chinese and had normal or corrected-to-normal vision. After the experiment, they each received RMB 30 (USD 4.5) as a reward. Participants gave informed consent in written form in accordance with the Declaration of Helsinki.

Materials. The 3,796 words were randomly divided into six groups of 632 to 636 words. For every group, we first transposed the constituent characters of each word to make the same number of pre-non-words. Then, we exchanged the left constituent character of a pre-non-word with the left constituent character of another pre-non-word and created the same number of two-character non-words as the word items in the group. The non-words as clusters of two Chinese characters did not exist as wholes in Modern Chinese Dictionary (Institute of Linguistics, Chinese Academy of Social Science, 1988). The first group of words was further divided into three sub-groups of roughly equal size, as was the corresponding non-word group. Every sub-group of words was mixed randomly with the same number of non-words to create one block of trials. The three blocks of trials were considered as materials group 1. Similarly, materials groups 2, 3, 4, 5, and 6 were created in the same manner.

Procedure. Six programs were designed for the six materials groups using DMDX (Forster & Forster, 2003). In a computer room, each program was run on five personal computers. No more than 30 students at a time collectively participated in the experiment, so that each of the six programs was run by roughly the same number of participants. Participants pressed the space bar to initiate a trial: after the red cross of fixation was shown for 500 ms on the center of the screen, the mask (\(\text{ fixation}\)) appeared at the position where the red cross of fixation had disappeared, and remained there for 200 ms. The target stimulus was subsequently displayed on the center of the screen. Participants made lexical decisions on the target. Half participants were required to press the keys “Z” and “/,” and the other half were required to press the keys “/” and “Z” if the target was a word and a non-word in Chinese, respectively. The target remained on the screen for 2,500 ms, or until a key press was received. The inter-trial interval was 1,000 ms. Prior to the experiment, the experimenter orally delivered the instructions, which were also presented at the beginning of the programs. Participants were told not to start the programs before they had carefully read and understood the instructions. In each program, there were 12, 4, and 4 trials of practice prior to the first, second, and third blocks of the experimental trials, respectively. Eventually, 28 or 29 participants were assigned to each of the six programs, and 28 or 29 responses were collected for each of the 3,796 2C-words.

Results

Six of the originally determined words were removed because they are not included in Cai and Brysbaert (2010). With reference to Liu et al. (2007), seven more words were removed because participants’ LD error rates on them were higher than .50. Therefore, the results consisted of the data for the remaining 3,783 2C-words. Also, as suggested in Liu et al. (2007), two participants’ LD data were excluded since their error rates were higher than 20%. The remaining 168 participants achieved an overall accuracy of .96. After the incorrect responses were removed, the LD RTs shorter than 200 ms or 3 SD above the overall average were excluded. Finally, 5.7% of the data was discarded. Table 2 summarizes the results.

The lowest Cronbach’s Alpha was over .96, suggesting a high reliability of the 10 subjective measurements. One might note that the mean RT on the 2C-words was 750 ms in Tsang et al. (2018) but was 598 ms in the present study. There are at least two reasons to account for the nearly 150 ms difference in LD RTs. Firstly, as indicated in 2.1, the participants in the present study were from the same campus, while those in Tsang et al. (2018) appeared to be rather heterogeneous in background. A large degree of diversity among participants could result in a big increase in the averaged RT in LDs. Indeed, the SD in Tsang et al. (2018) (\(SD = 91\) ms) appears to be much larger than the SD in the present study (\(SD = 51\) ms).

We conducted a Pearson-correlation analysis and a principal component analysis (PCA) for the feature scores and a regression analysis with the feature scores and the LD RTs as the predictors and the dependent variable, respectively.
To avoid the influence of distributional skewedness, the subTF and NOM scores were log-transformed before the implementation of the Pearson-correlation analysis. As indicated in Table 3, 64 of the 78 pair-wise correlations between the 13 variables were significant ($p < .05$ or $< .01$), for instance, the positive correlations between NOM and sNOM, between subTF and sF, between SF and EER, between familiarity and sF, between Compositionality and NOS, SER, and ST, between concreteness and ST and SER, between EER and SER, between SER and imageability and between SER and concreteness, ST, and EER, and the negative correlations between SubTF and AOA, between AOA and sF and ST, between familiarity and AOA, between concreteness and sNOM, between SER and AOA and between ST and AOA and sNOM.

We also revealed a few new correlations between the normative features in comparison with previous studies: (a) ST is positively correlated with compositionality and concreteness but negatively correlated with AOA and sNOM; (b) subTF is positively correlated with EER but negatively correlated with AOA; (c) SER is positively correlated with EER, compositionality, concreteness, and ST; (d) compositionality is negatively correlated with NOM, and concreteness is negatively correlated with sNOM.

**PCA Results**

The PCA suggested that the 13 variables can be represented by four principal components (see Table 4; KMO = 0.741; $\chi^2 = 8,263.549$, $df = 78$, $p < .001$). The first (heavily loaded with compositionality and NOS) component accounted 18.98%, 14.77%, 10.71%, and 10.17% of the variance among the 13 variables, respectively.

**Regression Analysis Results**

Similar to previous studies (e.g., Liu et al., 2007), a multiple regression analysis was conducted with the 13 variables as the predictors and the RT as the dependent variable. In order to avoid the influence of subject variability, the analysis was conducted on the $z$-scored RTs. The result showed that 10 features (subTF, NOM, EER, familiarity, AOA, concreteness, ST, compositionality, sF, and NOS) were significantly predictive of the LD RTs. However, the other three features (Imageability, SER, and sNOM) were not significantly predictive of the LD RTs. Moreover, we conducted the same analysis for the 2,510 2C-words common in the present study and in Tsang et al. (2018) and achieved the same pattern of results (see Table 5) ($F(13, 2,496) = 72.074$, $p < .001$, $R^2 = .273$).

**Discussion**

Both results that were similar to those of previous studies and results that appeared to be particular of 2C-words were obtained, concerning the correlations between the 13 features and the features’ predictability of LD RTs.

**Similarities to Previous Studies**

**Pearson-correlations.** As indicated in Table 3, the more meanings a 2C-word has in the dictionary, the more meanings the users have for it in everyday life. Similar to the case of English disyllabic words (Schock et al., 2012), 2C-words with...
Table 3. Correlation Coefficients Matrix.

|                | subTF | sF   | NOS | ST  | Familiarity | Concreteness | NOM | Compositionality | EER  | AOA  | Imageability | SER   | sNOM |
|----------------|-------|------|-----|-----|-------------|---------------|-----|------------------|------|------|--------------|-------|------|
| sF             | 0.519** | -0.058** | 0.111** | 0.353** | -0.070** | 0.189** | 0.046** | 0.246** | -0.317** | 0.040* | 0.117** | -0.489** |
| NOS            | 0.172** | 0.03  | 0.369** | 0.500** | 0.125** | 0.098** | 0.050** | 0.389** | 0.090** | 0.167** | 0.064** | -0.421** |
| ST             | 0.03   | 0.209** | 0.012 | 0.461** | 0.131** | 0.084** | 0.128** | 0.101** | 0.336** | 0.026  | 0.001   | 0.152** |
| Familiarity    | 0.309** | -0.004 | 0.046** | -0.131** | 0.088** | 0.268** | 0.110** | 0.055** | 0.325** | 0.002  | 0.001   | -0.192** |
| Concreteness   | 0.012  | 0.065** | 0.131** | 0.036** | 0.074** | 0.045** | 0.148** | 0.026  | 0.459** | 0.026  | 0.065** | -0.192** |
| NOM            | 0.036** | -0.123** | -0.131** | -0.059** | -0.128** | 0.074** | -0.001  | 0.026  | 0.045** | 0.002  | 0.065** | -0.084** |
| Compositionality | 0.046** | 0.074** | 0.101** | 0.055** | 0.128** | 0.088** | 0.045** | 0.001  | 0.459** | 0.002  | 0.065** | -0.126** |
| EER            | 0.110** | 0.268** | 0.101** | 0.268** | 0.074** | 0.045** | 0.148** | 0.026  | 0.459** | 0.002  | 0.065** | -0.126** |
| AOA            | 0.045** | 0.074** | 0.101** | 0.128** | 0.088** | 0.045** | 0.148** | 0.026  | 0.459** | 0.002  | 0.065** | -0.126** |
| Imageability   | 0.072** | 0.074** | 0.101** | 0.128** | 0.088** | 0.045** | 0.148** | 0.026  | 0.459** | 0.002  | 0.065** | -0.126** |
| SER            | 0.026  | 0.054** | 0.262** | 0.123** | 0.072** | 0.045** | 0.148** | 0.026  | 0.459** | 0.002  | 0.065** | -0.126** |
| sNOM           | -0.213** | -0.084** | -0.126** | -0.126** | -0.084** | -0.126** | -0.084** | 0.022  | 0.222** | 0.305** | 0.008  | -0.114** |

*p < .05. **p < .01.
high subTFs are likely to be acquired early in life. The more frequently a 2C-word is used in everyday life, the more likely it is to evoke emotional experiences in the users. The earlier a 2C-word is learned, the more frequently it is used in everyday life and the more semantically transparent it tends to be. Similar to the case of Chinese idioms (Li et al., 2016), the more familiar 2C-words are to the users, the more commonly they are used in everyday life and the earlier they are likely to be acquired.

The more often a 2C-word may be considered to be a composition of its individual characters, the more visually complicated it is, the more likely it is to arouse user sensory experiences and the more likely it can be understood literally through its constituent characters. The more concrete a 2C-word is in meaning, the more semantically transparent it is and the easier it is to evoke user sensory experiences. However, concrete 2C-words tend to have fewer meanings for the users in their daily lives. Similar to the case of Chinese characters (Liu et al., 2007), the more concrete a 2C-word is, the more likely it is to evoke mental images. However, this correlation between concreteness and imageability ($r = .101$) is seemingly weaker than in Liu et al. (2007; $r = .796$), which will be explained in the next sub-section.

A 2C-word that easily arouses emotional experiences in users also tends to evoke sensory experiences. Consistent with Bonin et al. (2015), 2C-words that more easily arouse

### Table 4. Principal Components (PCs) of the 13 Variables for 3,783 2C-Words.

| Variable   | PC1   | PC2   | PC3   | PC4   |
|------------|-------|-------|-------|-------|
| sF         | 0.82  | -0.00 | 0.07  | 0.04  |
| subTF      | 0.74  | -0.09 | 0.17  | -0.05 |
| Familiarity| 0.69  | 0.05  | -0.08 | 0.07  |
| AOA        | -0.60 | -0.40 | 0.12  | 0.15  |
| EER        | 0.50  | 0.20  | 0.19  | 0.39  |
| Concreteness| -0.03 | 0.80  | -0.24 | -0.06 |
| SER        | 0.19  | 0.72  | -0.01 | 0.32  |
| ST         | 0.28  | 0.57  | -0.38 | -0.03 |
| Imageability| -0.11 | 0.43  | 0.28  | 0.01  |
| NOM        | 0.10  | 0.02  | 0.73  | -0.11 |
| sNOM       | 0.06  | -0.18 | 0.69  | -0.05 |
| NOS        | -0.09 | -0.07 | -0.08 | 0.73  |
| Compositionality| 0.08 | 0.12  | -0.10 | 0.70  |
| Variance (%)| 18.98 | 14.77 | 10.71 | 10.17 |

Note. The boldface of the scores indicated the clustering of the variables that contributed to the corresponding components.

### Table 5. Regression Analysis Results.

| Variable   | 3,783 2C-words | 2,510 2C-words |
|------------|----------------|----------------|
|            | $B$            | $\beta$        | $t$  | Tolerance | $B$            | $\beta$        | $t$  | Tolerance |
| (Constant) | 2.974          | 13.519***      |      |           | 2.207          | 7.57***        |      |           |
| subTF      | -0.524         | -0.336          | 20.566*** | 0.672 | 0.000       | -1.66          | -8.2***      | 0.709       |
| sF         | -0.168         | -0.163          | -9.158*** | 0.568 | -0.222       | -2.18          | -9.579***    | 0.562       |
| NOS        | 0.025          | 0.115           | 8.349***  | 0.939 | 0.027        | 1.127          | 7.254***    | 0.951       |
| ST         | -0.125         | -0.076           | -4.748*** | 0.706 | -0.112       | -0.066         | -3.26***    | 0.718       |
| Familiarity| -0.111         | -0.069          | -4.498*** | 0.752 | -0.191       | -1.17          | -5.968***   | 0.753       |
| Concreteness| -0.065       | -0.068          | -4.355*** | 0.729 | -0.042       | -0.044         | -2.18*      | 0.716       |
| NOM        | -0.257         | -0.052          | -3.726*** | 0.933 | -0.073       | -0.067         | -3.77***    | 0.932       |
| Compositionality| 0.055      | 0.049           | 3.549***  | 0.921 | 0.079        | 0.071          | 4.005***    | 0.933       |
| EER        | -0.062         | -0.048         | -3.224**  | 0.817 | -0.064       | -0.051         | -2.72***    | 0.817       |
| AOA        | 0.076          | 0.049           | 3.036***  | 0.694 | 0.124        | 0.079          | 3.809***    | 0.683       |
| Imageability| —            | 0.025           | 1.859      | 0.989 | —            | 0.020           | 1.191      | 0.987       |
| SER        | —             | 0.005           | 0.275      | 0.633 | —            | 0.023           | 1.063      | 0.632       |
| sNOM       | —             | 0.004           | 0.254      | 0.870 | —            | -0.004          | -2.36**    | 0.874       |

Note. 3,783 2C-words were the targets in the present study, and 2,510 2C-words were the common words in the present study and in Tsang et al. (2018). $^*$p < .1. $^{**}$p < .01. $^{***}$p < .001.
sensory experiences in the users are typically acquired earlier in life. Furthermore, in agreement with Juhasz et al. (2011) and Bonin et al. (2015), 2C-words that more easily evoke mental images are also more likely to evoke sensory experiences. The 2C-words that more easily arouse sensory experiences in the users are usually more concrete in meaning, are more semantically transparent and are more likely to evoke emotional experiences. The more semantically transparent a 2C-word is, the earlier it is likely to be acquired and the fewer meanings it tends to have in the users’ daily lives.

**Principal components.** According to Table 4, the most frequently used 2C-words in daily life, which are familiar to the users and are acquired early by them, also arouse emotional experiences in the users more easily in word recognition. Similar to this case is the way in which bilinguals recognize emotional words in the second language (L2). For example, those who began to learn L2 at a late age in classroom settings are likely to fail to process the affective information of emotional words in L2 during cognitive tasks (Li et al., 2015). After all, one of the two important ways of enlarging one’s vocabulary is to enrich his or her life experiences (Andrews et al., 2009).

It is understandable that the second and third principal component represented the dimensions of semantics and meaningfulness, respectively. However, it is interesting that compositionality clustered with NOS, which will be explained in the next sub-section.

**Features significantly predictive of LD RTs.** In agreement with previous findings, subTF (Cai & Brysbaert, 2010; Tsang et al., 2018; Tse et al., 2017), NOM (Peng et al., 2003), EER (Newcombe et al., 2012), and familiarity and AOA (Chen et al., 2004) predicted changes in participants’ RTs to the 2C-words. That is, consistent with Liu et al.’s (2007) finding on Chinese characters, 2C-words are similar to 1C-words when measured in concreteness. Similar to the case of English compounds (Juhasz et al., 2015), the more semantically transparent a 2C-word is, the more easily it is to be recognized.

The simple correlation was not significant between compositionality and z-scored RT but the partial correlation was. That is, the effect of compositionality must have been affected by other variables such as NOS and ST, which were strongly correlated with compositionality. Compositionality might have a weak power in predicting changes in the LD RTs. However, the beta value was positive for compositionality but negative for ST, which will be explained in the next sub-section.

The cases of sF and NOS also tended to be somewhat controversial. SF has been shown to be a reliable determinant of participant performance in cognitive tasks concerning monosyllabic words in English (Balota et al., 2001, 2004). Similarly, it was significantly predictive of changes to participants’ RTs in the present study. However, Brysbaert and Cortese (2011) held the view that the weight of sF predominantly relied on the frequency quality. The influence of sF could be omitted when frequency values were reliably measured. Furthermore, sF did not appear to be predictive of participants’ RTs in picture naming in Liu et al. (2011). Thus, as indicated in Table 5, sF might have been weaker than subTF in predicting participants’ RTs in the present study.

Similarly, the finding that NOS appeared to predict changes in RTs in participants is compatible with the findings of previous studies on Chinese expressions (Liu et al., 2007; Tsang et al., 2018; Tse et al., 2017) and Japanese kanjis (Tamaoka & Kiyama, 2013). However, Chang et al. (2016) reported that NOS had a null effect on participants’ naming of traditional Chinese characters. In other words, sF or NOS might be slightly more constrained than subTF, concreteness, familiarity, ST, NOM, EER, or AOA in terms of their influences on 2C-word recognition by skilled readers.

**Features not significantly predictive of LD RTs.** Similar to concreteness, imageability involves words’ referents of external objects. Referring to the extent to which a word arouses mental pictures, sounds (Paivio et al., 1968), senses (Westbury et al., 2013), or sensory experiences (Paivio et al., 1968), however, imageability is also regarded as a composite measure (Dellantonio et al., 2014). It is possibly related to sensory experience ratings (Juhasz et al., 2011) or body interaction ratings (Siakaluk et al., 2008). It is similar to SER in indicating the internal bodily-related sensory experiences (Juhasz et al., 2011; Siakaluk et al., 2008).

SER is associated with words’ perceptual information which is grounded on sensory experiences, and can be construed as indications of simulation (Zdrazilova & Pexman, 2013). When seeing the word “gloves,” for example, the reader will be likely to unconsciously simulate the action of how to put gloves on and/or retrieve memories for the experiences of feeling warm and comfortable with gloves on in cold winter. Due to its similarity and significant correlation with SER (Juhasz et al., 2011), imageability may also reflect words’ simulation information to some extent. However, activation of simulation information is rather slow (Barsalou, 2008). For example, Zdrazilova and Pexman (2013) conducted an LD task on abstract-noun recognition and revealed the context availability effect, but failed to do the influences of SER and valence. LD tends to involve a lexical pattern match (Compton & Bradshaw, 1975) and does not seem to be as capable as semantic categorization of tapping the sensory and emotional aspects of words’ meanings (Zdrazilova & Pexman, 2013). In other words, changes in SER or imageability may simply not be sensitive enough to be reflected in LD RTs for a 2C-word in general. In fact, neither SER nor imageability was included in 2C-word normative studies by Sun et al. (2018), Tsang et al. (2018), Tse and Yap (2018), Tse et al. (2017), or Wang et al. (2008). Yao et al. (2017) did seem to reveal a significant imageability effect in LDs on 2C-words, but they did so on a sample of emotional words.
Words with high scores in valence or arousal are likely to have high scores in imageability (Altarriba & Bauer, 2004; Citron et al., 2014; Westbury et al., 2013).

SNOM was most strongly correlated with NOM. We conducted a regression analysis including all the variables except NOM in order to understand the result that sNOM was not significantly predictive of 2C-word LD RTs. Unfortunately, however, sNOM did not contribute significantly to changes in the dependent variable ($\beta = -0.008, t = -0.537, p = .592$), but all the variables except NOM that are listed in Table 5 did ($F(9, 3773) = 199.244, p < .001, R^2 = .322$). More discussion is provided in the next section.

**Findings Particular of 2C-Words**

The significant correlations between AOA and subTF and SER and that between SER and imageability for 2C-words are similar to the cases with words in European languages. That is, consistent with the conclusion that words are the basic units in the mental lexicon (Dronjic, 2011) and in sentence reading (Bai et al., 2008) by skilled readers, the examined 2C-words were taken as whole entities by participants in the present study. The significant correlation between concreteness and imageability of 2C-words is similar to the case of Chinese characters (Liu et al., 2007), confirming the conclusion that 2C-words are similar to 1C-words in semantic representations (Ma et al., 2016).

In comparison with these two points, the present study seemed to have yielded four more interesting results: (1) The correlation was relatively weak between concreteness and imageability. (2) Compositionality clustered with NOS in the PCA. (3) The beta values for compositionality and ST were positive and negative, respectively, in the regression equation. (4) SNOM diverged from NOM in predicting the changes in LD RTs.

On the one hand, the most important thing for a Chinese native speaker to do in literacy learning is to master a few thousand Chinese characters that are most often in use (Li et al., 2019). He or she learns how to read and write a certain number of basic Chinese characters and how to use them in a few cases. Indeed, one meets the minimum requirement for literacy if she/he masters the 3,500 basic Chinese characters in simplified Chinese (State Language Affairs Commission, 1991). Even grade-three primary school students are able to be clearly aware that a 2C-word consists of two Chinese characters (Jiang, 2011). On the other hand, it is repeatedly confirmed that the constituent characters are inevitably perceivable in the recognition of a visually presented 2C-word by skilled reader (Du et al., 2013; Liu et al., 2010; Peng et al., 1999; Tong et al., 2014; Wang et al., 2017; Yang, 2013). Therefore, it should have been participants’ processing of the 2C-words both at the word level and at the constituent-character level that led to the above-mentioned four interesting results.

**Concreteness and imageability.** The relatively weaker correlation between concreteness and imageability can be understood from two aspects. The recognition of a 2C-word is similar to that of a 1C-word in retrieving semantic representations, but 1C-words are more likely than 2C-words to be taken as wholes in cognitive tasks (Ma et al., 2016). Indeed, one has to assign cognition resources to the processing of its individual characters in the early stage of 2C-word recognition (Peng et al., 1999).

It is indicated that the more concrete a word is in meaning, the more imageable it is likely to be (Bonin et al., 2003; Guasch et al., 2016; Yao et al., 2017). Scholars even tend to take concreteness and imageability interchangeably (Westbury et al., 2016; Xia et al., 2016). However, the similarity between these two semantic features appeared to be of a lower degree for 2C-words than for Chinese characters. Of course, more studies are needed to elaborate this point of view.

**Compositionality and NOS.** The factor that compositionality clustered with NOS seemed to suggest participants’ perception of the 2C-words’ visual information. Firstly, the processing of configurational information of both 1C-words and 2C-words helps trigger activations of the semantic representations. For example, 两 (two) and 拳击 (boxing) are similar to 雨 (rain) and 泰山 (Taishan Mount), respectively, only in configuration, and perceptions of 两 and拳击 resulted in activations of semantic representations for雨和泰山, respectively (Ma et al., 2016).

Secondly, the degree to which the constituent characters can be seen as consisting of components plays a role in 2C-word recognition by skilled readers. For example, Wang et al. (2019) conducted a lexical decision task on 2C-words, the first constituent characters of which were composed of two components with one vertically above the other. The researchers also manipulated whether or not the first constituent characters were presented upside-down (presentation mode). If the first constituent characters had a high score in being rated as compositions of two components, participants tended to recognize the 2C-words according to the constituent characters and the N400 amplitude was significantly influenced by presentation mode; if the first constituent characters had a low score in being rated as compositions of two components, however, participants tended to recognize the 2C-words as wholes and no significant change of N400 amplitude was observed under the influence of presentation mode. N400 is an event-related potential component that is supposed to indicate semantic processing.

Thirdly, the constituent characters’ relative positions are also regarded as important in the representation of visual features for 2C-words (Tan & Perfetti, 1999). There is a probability of .97 to obtain a non-word if the constituent characters’ relative positions are exchanged in a 2C-word (Gu et al., 2015). For example, 学校 (school) is a word, but 校学 is not. Indeed, new learners of Chinese are found to be quite
confusing with the constituents’ relative positions (Hao & Li, 2015). It seems very important for the learner to take carefully the constituent characters’ relative positions in a word of two or more Chinese characters.

Skilled readers should be able to be unconsciously aware of the configurational features of the components, constituent characters and words and aware of the spatial positions of the components and constituent characters in the perception of 2C-words according to Taft and Zhu (1997). Similar to the event that compositionality of 2C-words’ first constituent characters affected participants’ LD performance (Wang et al., 2019), the higher compositionality score a 2C-word had, the longer its LD RT was in the present study. In other words, the compositionality scores might have been loaded with a large amount of the raters’ awareness of 2C-words’ visual characteristics. NOS indicates the visual complexity of a 2C-word at the stroke level (Liu et al., 2007; Wang et al., 2020), and compositionality seems to suggest the visual characteristics of configuration and intra-structure at the level of components, constituent characters and words. This might be the reason why compositionality clustered with NOS as shown in Table 4.

Compositionality and ST. Both the words as wholes and the constituent characters can be inevitably perceived in 2C-word recognition (Peng et al., 1999), and this perception may be a function of the relationship diversity between the two constituent characters in 2C-words (Jia et al., 2013). There is a copulative (e.g., 争夺-seize and take by force), determinative (e.g., 鸭毛-duck’s feather), numeral (e.g., 四季-the four seasons), descriptive (e.g., 大雪-heavy snow), adverbial (e.g., 难看-ugly), or reciprocal (e.g., 白痴-idiot) relationship between the constituent characters of a 2C-word. Those with high scores in ST but low scores in compositionality (e.g., 争夺) might be more likely to be taken as whole entities than those with low scores in ST but high scores in compositionality (e.g., 难看). That is, participants were more likely to assign resources of cognition to the individual morphemes and consume cognitive resources in integrating their comprehension of the constituent characters in LDs on a 2C-word of high than of low compositionality. However, ST might function to indicate the semantic features of 2C-words as wholes. Thus, RTs were shorter to the 2C-words that had high scores than to those that had low scores in ST. As a consequence, the beta value was positive for compositionality but negative for ST as indicated in Table 5. In other words, compositionality is different from ST in indicating the semantic relationship between a 2C-word and its constituent characters.

SNOM and NOM. Words in English usually have more than one meaning (Khanna & Cortese, 2011), but it might not be that case for Chinese 2C-words (Xiao, 1991). Indeed, most 2C-words have one meaning according to the dictionary, but participants thought that most words had one and half meanings in the present study. The reason why sNOM diverged from NOM in variance probably lied in the semantically complex relationships between 2C-words and their constituent characters and in how participants behaved in 2C-word recognition.

On the one hand, the number of frequently used words is 2,000 to 5,000 in English (Schmitt, 2007, p. 828). However, a Chinese middle-school student needs to master over 3,000 of the most commonly used Chinese characters. We randomly selected 25 of the 3,500 most commonly used Chinese characters, and found that their mean NOMs was 3.44 (SD=1.73). A 2C-word usually has fewer meanings than a Chinese character (Peng et al., 2003). On the other hand, skilled readers have developed representations for commonly used 2C-words as wholes and also for the constituent characters (Wang & Peng, 1999). Both commonly used 2C-words as whole entities and their constituent characters are perceivable in the early stage of word recognition (Liu et al., 2014; Peng et al., 1994, 1999; Zhang & Peng, 1992). Therefore, participants’ sNOM rating of 2C-words might have been incorporated with their perceptions of several other features of the words. However, empirical evidence is still needed for further research.

Of course, there seem to be at least two limitations to the generalization of the present findings. First, a direct investigation into the semantic relationships between the 2C-words and their constituent characters could have been valuable in better understanding the data. Second, there might have been more factors than the 13 features that affected the data, since the four principal components accounted less than 60% of the variance. Undoubtedly, however, the present study has a potentially significant reference value for further research into Chinese lexical processing. Furthermore, the original data for the 3,783 2C-words are available as Supplemental Material.

Conclusion

Using objective data obtained from dictionaries regarding subTF, NOS, and NOM, as well as subjective data obtained by means of questionnaires concerning familiarity, concreteness, imageability, AOA, sF, sNOM, compositionality, EER, SER, and ST, a normative description was provided for 3,783 2C-words in simplified Chinese. Correlative patterns were identified between the 13 variables, all of which, with the exception of imageability, SER, and sNOM, were significantly predictive of the changes in participants’ RTs in LDs on the target words. In conclusion, skilled readers’ awareness of 2C-words’ features such as concreteness, imageability, compositionality, sNOM, and ST is closely associated with their semantic perception of the constituent characters.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Social Science Foundation of China under Grant 17BYY016.

ORCID iD
Degao Li https://orcid.org/0000-0002-9511-3109

Supplemental Material
Supplemental material for this article is available online.

References
Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. The American Journal of Psychology, 117, 389–410.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. Psychological Review, 116(3), 463–498.
Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. Journal of Experimental Psychology Human Perception & Performance, 34(5), 1277–1287.
Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. Journal of Experimental Psychology General, 133(2), 283–316.
Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. Memory & Cognition, 29(4), 639–647.
Barca, L., Burani, C., & Ardino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. Behavior Research Methods Instruments & Computers, 34, 424–434.
Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychology, 59, 617–645.
Blomberg, F., & Öberg, C. (2015). Swedish and English word ratings of imageability, familiarity and age of acquisition are highly correlated. Nordic Journal of Linguistics, 38(3), 351–364.
Bonin, P., Méot, A., Aubert, L. F., Malardier, N., Niedenthal, P. M., & Capelle-Toczek, M. C. (2003). Concreteness, imageability,
subjective frequency and emotionality ratings for 866 words. *L’année psychologique*, 103(4), 655–694.

Bonin, P., Méot, A., Ferrand, L., & Bugańska, A. (2015). Sensory experience ratings (SERs) for 1,659 French words: Relationships with other psycholinguistic variables and visual word recognition. *Behavior Research Methods*, 47(3), 813–825.

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, 64(3), 545–559.

Cai, Q., & Brysbaert, M. (2010). Subtlex-ch: Chinese word and character frequencies based on film subtitles. *PLoS One*, 5(6), 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(1), 112–122.

Chen, B. G., Wang, L. X., Wang, L. L., & Peng, D. L. (2004). The effect of age of word acquisition and frequency on the identification of Chinese double-character words. *Psychological Science*, 27(5), 1060–1064.

Chen, X., & Dong, Y. (2019). Evaluating objective and subjective frequency measures in L2 lexical processing. *Lingua*, 230, 102738.

Chinese Proficiency Oral Test Center of Beijing Language and Culture University. (2009). *HSK Chinese proficiency test vocabulary guideline (Dictionary of Chinese usage: 8000 words)*. Beijing Language and Culture University Press.

Cieslicka, A. B. (2013). Do nonnative language speakers chew the fat and spill the beans with different brain hemispheres? Investigating idiom decomposability with the divided visual field paradigm. *Journal of Psycholinguistic Research*, 42(6), 475–503.

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014). How are affective word ratings related to lexico-semantic properties? Evidence from the sussex affective word list. *Applied Psycholinguistics*, 35, 313–331.

Compton, A., & Bradshaw, J. L. (1975). Differential hemispheric mediation of nonverbal visual stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, 1(3), 246–252.

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*, 40, 791–794.

Dasgupta, T., Sinha, M., & Basu, A. (2016). Computational models of the representation of bangla compound words in the mental lexicon. *Journal of Psycholinguistic Research*, 45(4), 833–855.

Dellantonio, S., Mulatti, C., Pastore, L., & Job, R. (2014). Measuring inconsistencies can lead you forward: Imageability and the x-ception theory. *Frontiers in Psychology*, 5, 708.

Desrochers, A., Liceras, J. M., Fernández-Fuertes, R., & Thompson, G. L. (2010). Subjective frequency norms for 330 Spanish simple and compound words. *Behavior Research Methods*, 42(1), 109–117.

Dronjic, V. (2011). Mandarin Chinese compounds, their representation, and processing in the visual modality. *Writing Systems Research*, 3(1), 5–21.

Du, Y., Hu, W., Fang, Z., & Zhang, J. X. (2013). Electrophysiological correlates of morphological processing in Chinese compound word recognition. *Frontiers in Human Neuroscience*, 7, 601.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124.

Ge, B. (2018). *Modern Chinese lexicology*. Shandong People’s Publishing House.

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4), 154–159.

Gonthier, I., Desrochers, A., Thompson, G., & Landry, D. (2009). Imagery norms and subjective frequency of 1,760 monosyllabic words in the French language. *Revue Canadienne de Psychologie Expérimentale*, 63(2), 139–149.

Grigoriev, A., & OshepkoKov, I. (2013). Objective age of acquisition norms for a set of 286 words in Russian: Relationships with other psycholinguistic variables. *Behavior Research Methods*, 45(4), 1208–1217.

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, 48(4), 1358–1369.

Gu, J., Li, X., & Liversedeg, S. P. (2015). Character order processing in Chinese reading. *Journal of Experimental Psychology: Human Perception & Performance*, 41(1), 127–137.

Hao, M. L., & Li, L. (2015). Factors affecting cognitive processing of Chinese compounds: Evidence from non-proficient L2 learner. *Language Teaching and Linguistic Studies*, 2, 8–18.

Huang, H. W., & Lee, C. Y. (2018). Number of meanings and number of senses: An ERP study of sublexical ambiguities in reading Chinese disyllabic compounds. *Frontiers in Psychology*, 9, 324.

Hubers, F., Cucchiari, C., Strik, H., & Dijkstra, T. (2019). Normative data of Dutch idiomatic expressions: Subjective judgments you can bank on. *Frontiers in Psychology*, 10, 1075.

Institute of Linguistics, Chinese Academy of Social Science. (1988). *Modern Chinese dictionary*. The Commercial Press.

Jiang, Z. (2011). How to look at children-make words. *Contemporary Preschool Education*, 3, 38–40.

Jia, X., Wang, S., Zhang, B., & Zhang, J. X. (2013). Electrophysiological evidence for relation information activation in Chinese compound word comprehension. *Neuropsychologia*, 51(7), 1296–1301.

Juhasz, B. J., Lai, Y. H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods*, 47(4), 1004–1019.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*, 64(9), 1683–1691.

Khanna, M. M., & Cortese, M. J. (2011). Age of acquisition estimates for 1,208 ambiguous and polysemous words. *Behavior Research Methods*, 43(1), 89–96.

Language Teaching and Research Institute of Beijing Language and Culture University. (1986). *Modern Chinese frequency dictionary*. Beijing Language and Culture University Press.
Li, D., Liu, H., & Ma, B. (2015). Association of affect with vertical position in L1 but not in L2 in unbalanced bilinguals. *Frontiers in Psychology, 6*, 693.

Li, D., Xu, J., & Zhang, J. (2019). Semantic primitives and “heart. *Journal of China University of Petroleum (Edition of Social Sciences), 35*(1), 81–85.

Li, D., Zhang, Y., & Wang, X. (2016). Descriptive norms for 350 Chinese idioms with seven syntactic structures. *Behavior Research Methods, 48*(4), 1678–1693.

Liu, D., Chung, K. K. H., Zhang, Y., & Lu, Z. (2014). Sensitivity to the positional information of morphemes inside Chinese compound words and its relationship with word reading. *Reading and Writing, 27*(3), 431–450.

Liu, P. D., Chung, K. K. H., McBride-Chang, C., & Tong, X. (2010). Holistic versus analytic processing: Evidence for a different approach to processing of Chinese at the word and character levels in Chinese children. *Journal of Experimental Child Psychology, 107*(4), 466–478.

Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed picture naming norms for Mandarin Chinese. *PLoS One, 6*(1), e16505.

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

Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition, 22*, 441–457.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology General, 143*(2), 895–913.

Li, X. J. (2014). *Modern Chinese standard dictionary* (3rd ed.). Foreign Language Teaching and Research Press.

Lv, S. (1979). *Chinese grammar analysis*. The Commercial Press.

Ma, B., Wang, X., & Li, D. (2016). The processing of visual and phonological configurations of Chinese one- and two-character words in a priming task of semantic categorization. *Frontiers in Psychology, 6*, 1918.

Miwa, K., Libben, G., Dijkstra, T., & Baayen, H. (2014). The time-course of lexical activation in Japanese morphographic word recognition: Evidence for a character-driven processing model. *Quarterly Journal of Experimental Psychology, 67*(1), 79–113.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A, 50*, 528–559.

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in Human Neuroscience, 6*, 275.

Paivo, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(Suppl. 1), 1–25.

Peng, D. L., Deng, Y., & Chen, B. G. (2003). The polymy effect in Chinese 1C-word identification. *Acta Psychologica Sinica, 35*(5), 569–575.

Peng, D. L., Ding, G. S., Wang, C. M., Taft, M., & Zhu, X. P. (1999). The processing of Chinese reversible words: The role of morphemes in lexical access. *Acta Psychologica Sinica, 31*(1), 36–44.

Peng, D. L., Li, Y. P., & Liu, Z. Z. (1994). Identification of the Chinese two-character word under repetition priming condition. *Acta Psychologica Sinica, 26*(4), 393–400.

Schmitt, N. (2007). Current perspectives on vocabulary teaching and learning. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 827–841). Springer.

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*, 374–379.

Schulte Im Walde, S., & Borgwaldt, S. R. (2015). Association norms for German noun compounds and their constituents. *Behavior Research Methods, 47*(4), 1199–1221.

Siakaluk, P. D., Pexman, P. M., Aguiler, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect. *Cognition, 106*, 433–443.

State Language Affairs Commission. (1991). “The common Chinese characters in modern Chinese” in manual for standard Chinese literacy. Chinese Press.

State Language Affairs Commission. (2008). *Lexicon of common words in contemporary China*. The commercial Press.

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*(6), 2606–2629.

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*(8), 1541–1570.

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology Learning Memory and Cognition, 23*(3), 761–775.

Tamaoka, K., & Kiyama, S. (2013). The effects of visual complexity for Japanese kanji processing with high and low frequencies. *Reading and Writing, 26*(2), 205–223.

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*(2), 382–393.

Tong, X., Chung, K. K., & McBride, C. (2014). Two-character Chinese compound word processing in Chinese children with and without dyslexia: ERP evidence. *Developmental Neuropsychology, 39*(4), 285–301.

Tsang, Y. K., Huang, J., Lui, M., Xue, M., Chan, Y. F., Wang, S., & Chen, H. C. (2018). Meld-sch: A megastudy of lexical variables in the visual recognition of Chinese characters: A megastudy analysis. *Quarterly Journal of Experimental Psychology, 71*(9), 2022–2038.

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*(4), 1503–1519.
Wang, C. M., & Peng, D. L. (1999). The roles of surface frequencies, cumulative morpheme frequencies, and semantic transparencies in the processing of compound words. *Acta Psychologica Sinica, 31*(3), 266–273.

Wang, R., Huang, S., Zhou, Y., & Cai, Z. G. (2020). Chinese character handwriting: A large-scale behavioral study and a database. *Behavior Research Methods, 52*(1), 82–96.

Wang, T., Zhuang, B., & Li, D. (2019). The recognition of Chinese two-character words the first characters of which consist of two components with one above the other. *Chinese Social Sciences Today, 1648*, 7.

Wang, X., Ma, B., & Li, D. (2017). Similarities between Chinese 2R- and 2C-words in the processing of positional information of the words’ components. *Journal of Psychological Science, 40*(5), 1075–1083.

Wang, Y. N., Zhou, L. M., & Luo, Y. J. (2008). The pilot establishment and evaluation of Chinese affective word system. *Chinese Mental Health Journal, 22*(8), 39–43.

Westbury, C. F., Cribben, I., & Cummine, J. (2016). Imaging imageability: Behavioral effects and neural correlates of its interaction with affect and context. *Frontiers in Human Neuroscience, 10*, 346.

Westbury, C. F., Shaoul, C., Hollis, G., Smithson, L., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2013). Now you see it, now you don’t: On emotion, context, and the algorithmic prediction of human imageability judgments. *Frontiers in Psychology, 4*, 991.

Xiao, T. F. (1991). On the polysemy of Chinese words. *Journal of Shanxi University (Philosophy and Social Science Edition), 1*, 78–81.

Xia, Q., Wang, L., & Peng, G. (2016). Nouns and verbs in Chinese are processed differently: Evidence from an ERP study on monosyllabic and disyllabic word processing. *Journal of Neurolinguistics, 40*, 66–78.

Yael, W., Tami, K., & Tali, B. (2015). The effects of orthographic transparency and familiarity on reading Hebrew words in adults with and without dyslexia. *Annals of Dyslexia, 65*(2), 84–102.

Yang, J. (2013). Preview effects of plausibility and character order in reading Chinese transposed words: Evidence from eye movements. *Journal of Research in Reading, 36*(S1), S18–S34.

Yan, G. L., Du, C., Bian, Q., & Bai, X. J. (2014). Eye movement study on the effect of noun word concreteness. *Psychological Exploration, 34*(3), 206–212.

Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2017). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods, 49*(4), 1374–1385.

Yap, M. J., Liow, S. J., Jail, S. B., & Faizal, S. S. (2010). The Malay lexicon project: A database of lexical statistics for 9,592 words. *Behavior Research Methods, 42*(4), 992–1003.

Yee, L. T. S. (2017). Valence, arousal, familiarity, concreteness, and imageability ratings for 292 two-character Chinese nouns in Cantonese speakers in Hong Kong. *PLos One, 12*(3), e0174569.

Yuan, C. F., & Huang, C. N. (1998b). A database study on Chinese morphemes. *Applied Linguistics, 3*, 83–88.

Yuan, H. (2001). *Modern Chinese polysemous words dictionary*. Shuhai Press.

Zdrazilova, L., & Pexman, P. M. (2013). Grasping the invisible: Semantic processing of abstract words. *Psychonomic Bulletin & Review, 20*(6), 1312–1318.

Zhang, B. Y., & Peng, D. L. (1992). Decomposed storage in the Chinese lexicon. In H.-C. Chen & Q. T. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 131–149). Amsterdam, North-Holland.

Zhang, X., Han, Z., & Bi, Y. (2013). Are abstract and concrete concepts organized differently? Evidence from the blocked translation paradigm. *Applied Psycholinguistics, 34*(5), 1059–1092.

Zhou, X., & Marslen-Wilson, W. D. (2009). Pseudo-homophone effects in processing Chinese compound words. *Language and Cognitive Processes, 24*(7–8), 1009–1038.