Behavior Research Methods (2019) 51:2310–2336
https://doi.org/10.3758/s13428-019-01208-2

CLAD: A corpus-derived Chinese Lexical Association Database

Shu-Yen Lin[1] · Hsueh-Chih Chen[2] · Tao-Hsing Chang[3] · Wei-En Lee[4] · Yao-Ting Sung[5]

Published online: 19 August 2019

Abstract

The application of word associations has become increasingly widespread. However, the association norms produced by traditional free association tests tend not to exceed 10,000 stimulus words, making the number of associated words too small to be representative of the overall language. In this study we used text corpora totaling over 400 million Chinese words, along with a multitude of association measures, to automatically construct a Chinese Lexical Association Database (CLAD) comprising the lexical association of over 80,000 words. Comparison of the CLAD with a database of traditional Chinese word association norms shows that word associations extracted from large text corpora are similar in strength to those elicited from free association tests but contain a much greater number of associative word pairs. Additionally, the relatively small numbers of participants involved in the creation of traditional norms result in relatively coarse scales of association measurement, whereas the differentiation of association strengths is greatly enhanced in the CLAD. The CLAD provides researchers with a great supplement to traditional word association norms. A query website at www.chinesereadability.net/LexicalAssociation/CLAD/ affords access to the database.

Keywords Word association · Lexical association · Association measures · Word co-occurrence · Corpus-derived · Corpus-based · Chinese text corpora

Words are represented in memory as groups in associative structures, bound together through the specification of values along semantic and episodic dimensions (Masson, 1995; McRae & Boisvert, 1998; Seidenberg, Waters, Sanders, & Langer, 1984). Many scholars have demonstrated different associative strengths between words through a variety of experiments, notably word priming. For example, subjects respond to *nurse* faster than they normally would if it follows a highly associated word such as *doctor* (Meyer & Schvaneveldt, 1971).

Word association has become one of the most common methods of exploring cognitive structures (Fazio, 2001; Plaut & Booth, 2000; Preece, 1976). One frequently used method to obtain word association data is to run free association tests in which a series of stimulus words are presented to respondents who must quickly reply with the word that first comes to mind (the response) upon reading or listening to the stimulus. The underlying assumption of this form of word association test is that stimulus–response relations reflect the structure of words and concepts in the long-term memory. Differences between individuals in lexical associations can thus be used to reveal characteristics about people, such as personality, thinking patterns, affective structure, and so forth (Bargh, Chen, & Burrows, 1996; Crossley, Salsbury, & McNamara, 2015; Greenwald, McGhee, & Schwartz, 1998; Merten & Fischer, 1999; Wu & Chen, 2017).

Beyond using word association to examine individual cognitive structures and processes, psychologists have also investigated the normality of word associations in order to capture the shared relational representations in lexical memory. Several researchers have collected word associations from word association tests and constructed word association

✉ Yao-Ting Sung
sungtc@ntnu.edu.tw

[1] Chinese Language and Technology Center, National Taiwan Normal University, Taipei, Taiwan

[2] Department of Educational Psychology and Counseling/Chinese Language and Technology Center/Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei, Taiwan

[3] Department of Computer Science & Information Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[4] Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan

[5] Department of Educational Psychology and Counseling/Chinese Language and Technology Center/Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei, Taiwan

norms (De Deyne & Storms, 2008; Jenkins, 1970; Kiss, Armstrong, Milroy, & Piper, 1973; Nelson, McEvoy, & Schreiber, 1998, 2004; Palermo & Jenkins, 1964).

The importance of word association norms is mostly twofold: First, through word associations, researchers can control or manipulate the associative strength of words and precisely select the vocabulary they wish to study (e.g., Nelson, McKinney, Gee, & Janczura, 1998; Siyanova-Chanturia, Conklin, & Van Heuven, 2011). Second, the calculation of lexical relevance has become an important part of language technology and has been widely used in typo detection and correction, automatic text summarization, word sense disambiguation, topic shift detection, and other language processing tasks (e.g., Matsuo & Ishizuka, 2004; Netzer, Feldman, Goldenberg, & Fresko, 2012; Sung, Chang, Lin, Hsieh, & Chang, 2016; Tseng, Chen, Chang, & Sung, 2019). Currently we are likely unable to fully imagine the scope of applications for word association, and new applications are being explored all the time. For instance, Li, Schloss, and Follmer (2017) showed that word association is a good predictor of political party affiliation. Roininen, Arvola, and Lähteenmäki (2006) discovered that word association is a quick and effective way to collect local consumer food preferences. Word association clearly has commercial and social value, and new applications are likely to continue to be put forward.

The Chen Chinese word association norms

The Chen norms are the largest and most widely used Chinese set of word association norms at present (Hu et al., 2017; Huang, Chen, Huang, & Liu, 2009; Huang, Chen, & Liu, 2012). Hsueh-Chih Chen and his colleagues first selected 900 high-frequency and 900 low-frequency two-character words that are nouns, verbs, adjectives, or adverbs from the *Concise Mandarin Chinese Dictionary* (http://language.moe.gov.tw/index.aspx). Low-frequency words here refer to words with below-average frequencies, whereas high-frequency words refer to the most frequent 13% of words. A preliminary study was conducted to rate the imageability of the words—that is, the extent to which a word evokes a mental image. After one-third of the words with more inconsistent ratings were removed, the imageabilities of the remaining 1,200 words were quite evenly distributed. Then, 1,417 18- to-22-year-old Taiwanese university students enrolled in Introduction to Psychology courses participated in a free association test. Each participant read 200 stimulus words from a paper test book and was tasked with writing down the first word that came to mind. A total of 217 participants whose responses were deemed inappropriate or incomplete were deleted. For each stimulus word, the norms thus collected responses from 200 different participants. We have constructed a query system for the Chen norms, at www.chinesereadability.net/LexicalAssociation/Norm/.

Limitations of conventional word association norms

Unfortunately, conventional word association norms have limitations that need to be overcome for wide and large-scale applications to be possible. First, word association tests have traditionally only examined a few thousand words, leaving the associative structure of many words unexamined. Furthermore, in past norm construction, despite great effort to recruit thousands of participants usually each participant responded to only one or a few hundred stimuli (Kiss et al., 1973; Nelson et al., 1998, 2004). Thus, the average stimulus word typically received only one to two hundred responses, resulting in a relatively small number of associated words and associative strength measurements that could be more precise. Another problem with word association tests is that the associations of the secondary senses of a word can be suppressed by the primary sense in the free association process. For example, in the study of Kiss et al. the stimulus word *table* only invoked responses associated with the sense of a piece of furniture. However, *table* also commonly refers to a method of displaying information. Words like *figure* and *appendix*, which were not elicited by Kiss et al., would be very likely to be conjured by the reader or listener when this sense of *table* was used in context.

Considering the limitations of using associative norms, we think it necessary to construct a comprehensive (including all common words) and representative (gathered from the natural linguistic productions of a substantial number of people) word association database that includes associations with a wide variety of word meanings. Besides eliciting real-time responses, the use of text corpora may be the most convenient method of obtaining large quantities of linguistic and cognitive information. Corpus-based studies have for many years been used in language-related research, including data mining (Aggarwal & Zhai, 2012), pedagogical and specialized lexicography (Kilgarriff & Grefenstette, 2003), machine translation and learning (Liu, Hsaio, Lee, Chang, & Kuo, 2016; Rauf & Schwenk, 2011; Sung et al., 2015), artificial intelligence (Boden, 1998; McNamara, Crossley, & Roscoe, 2013), and numerous other examples (Chen, Liu, Chen, Wang, & Chen, 2016; Lee, Juan, Tseng, Chen, & Tseng, 2015). Because the linkages between words represent the relationships between the concepts embodied in human language (Li & Zhao, 2017), the numerous connections between words in a large corpus when distilled into an association database could be a great supplement to traditional association norms. We aimed to produce exactly such a corpus-derived Chinese Lexical Association Database (CLAD).

Computing lexical association strength using word (co-)occurrence frequency data

Word co-occurrence and lexical association Deese (1966) stated that "almost all the basic propositions of current association theory derive from the sequential nature of events in human experience" (p. 1). This property arises from the fact that experience is not accumulated from random events. Thus, unlike a lottery ticket, where the next winning number cannot be predicted from the last winner, the percent chance that any given word will appear is altered by what other words have already appeared. For instance, because *apple* is essential to the telling of the famous fairy tale *Snow White*, and because *red* is one of the intrinsic properties of most apples, these words will frequently appear in the same text. Although some words are logically or practically related, others have more idiosyncratic connections. When expressing the idea that tea is highly concentrated, English speakers use the phrase *strong tea*. Although *powerful* is very close in meaning to *strong*, native speakers consistently think that the phrase *powerful tea* is odd (Halliday, 1966). Similar examples are numerous in which the intuitive perception of a word's tighter relation with one word than another cannot be accounted for on pure syntactic or semantic grounds. It seems that a very good reason to explain the native sense of many fixed lexical usages is a sufficient exposure to the combination (i.e., co-occurrence) of lexical items in text. To summarize, whether reinforced by life experience, natural properties, or lexical idiosyncrasy, these examples show that the co-occurrence of words in text can be used to index lexical association.

Spence and Owens (1990) tested the relation between lexical association and word co-occurrence. They first drew from Palermo and Jenkins's (1964) word association norms for concrete noun stimuli. For each stimulus, its most common concrete noun response was selected to form an experimental word pair. Then they perused the one-million-word Brown Corpus (Kučera & Francis, 1967) for other concrete nouns that appeared in the corpus with the same frequencies as the response words. These equal-frequency but unrelated words were matched with the stimulus words to generate control pairs. It was found that words of the experimental pairs co-occurred much more frequently in the Brown Corpus than the words of the control pairs. The strong, positive correlation between association rates in the norm and co-occurrence rates in the corpus supports the role of lexical co-occurrence in lexical association. Similarly, Charles and Miller (1989) put forth the hypothesis that "the cue for learning to associate direct antonyms is not their substitutability, but rather their relatively frequent co-occurrence in the same sentence" (p. 357), which was subsequently supported by Justeson and Katz (1991) through their analysis of empirical data.

However, by employing the co-occurrence-oriented approach to retrieve lexical association data, we do not mean to claim that repeated co-occurrence of word forms is a way (or even the only way) to build up lexical association into the structure of our mental lexicon. For the purpose of the present study, we focus on extracting lexical association from text corpora. It is not our goal here to prove whether, or to what extent, association bonds cause co-occurrence or vice versa.

It is nonetheless relevant to ask whether the information retrieved from text corpora based on word co-occurrence is always word associations. The inquiry led us to ponder the definition of word association. McRae and Boisvert (1998) noted that "researchers have typically circumvented the definitional problem [of word association] by operationalizing associative relatedness in terms of word association norms" (p. 569). Given that the current association norms are all restricted in size, the operationalized definition has a fundamental flaw: If a word is not found associated with another word in the norms, one can always argue that the set of norms is simply not large enough. But then again, would corpus-derived word association data differ systematically from traditional norms for which subject recruitment was unlimited? Some recent research has provided a tentative affirmative answer: In four word-priming experiments, Hare, Jones, Thomson, Kelly, and McRae (2009) discovered that certain priming effects can be more adequately explained by event knowledge than by normative word associations. They proposed that nouns should tend to prime other nouns present in related events (e.g., *sale–shopper*, *barn–hay*, *key–door*), and concluded that event-based word relationships are encoded in semantic memory and construed as part of word meaning. Although discussing a more general framework, McRae, Khalkhali, and Hare (2012) commented that "association proper is learning-based; word association [in the operationalized sense] is retrieval or production-based" (p. 45). In line with this argument, we point out that learning may occur implicitly (Frensch & Rünger, 2003), whereas the process of retrieval is mainly explicit. More importantly, what is implicitly learned, including lexical association, may not always be able to be retrieved explicitly, and we would like our corpus-derived lexical association data to fill this gap.

Co-occurrence window Some associated words, such as *strong tea*, are almost always adjacent words, whereas other associative words—for instance, *Snow White* and *apple*—are often separated by some number of words. Technically, whether two words can be said to co-occur depends on the distance between them (i.e., the number of intervening words). Studies of word co-occurrence thus often define a specific segment of consecutive text words, known as the *co-occurrence window* or simply *window* (Sinclair, 1991), within which both words occur simultaneously.

Lexical association measures Lexical association measures are methods for computing association strengths between words

on the basis of their (co-)occurrences in text corpora. Many measures originate in statistical sampling that determines the degree of association by calculating how strongly words co-occur more often than expected by chance (Manning & Schütze, 1999). Some measures compute the entropy of the immediate context of the words by assuming that words occur as units in an information-theoretically noisy environment (Cover & Thomas, 1991; Krenn, 2000). Other measures compute the cosine or dice similarity score of the words based on a vector space model (Frakes & Baeza-Yates, 1992).

Pecina (2010) offered one of the most comprehensive lists of lexical association measures based on the three approaches sketched above. The study also evaluated the performance of the association measures in identifying human-rated associative pairs. Measures of the second and third approaches generally fell into the lower half of performance ranks, while many of the statistically oriented measures performed equally well, topping the rank list. Other relevant studies on word association have also generally relied on statistical measures (Chung & Lee, 2001; Evert & Krenn, 2005; Michelbacher, Evert, & Schütze, 2011; Petrović, Šnajder, & Bašić, 2010). To construct our lexical association database we also used statistical association measures.

The tacit premise that underlies all the statistical measures is the thinking that word association is a hidden parameter that is reflected by the word (co-)occurrence frequencies. We have already pointed out the importance of co-occurrence frequency for word associations. That the frequencies of words occurring individually can also be essential indicators of word association can easily be comprehended. Suppose that word A has identical co-occurrence frequencies with words B and C. However, B's individual frequency is much higher than C's. Consequently, A may just be one of many not-so-relevant words for B, whereas it may be a strong associate for C.

The formulas of the 55 statistical measures given in Pecina (2010) are relisted in the Appendix of this article. The statistical association measures all employ part or all of the frequency data, as displayed in a contingency table (see Table 1).

Directionality of word association measures If one word strongly elicits another while the elicitation in the other direction is relatively weak, the two words are asymmetrically associated. For example, Michelbacher, Evert, and Schütze (2011) showed that in the University of South Florida Association norms (Nelson et al., 1998) the pairs *bird* and *canary* are asymmetrically related: 69% of subjects give *bird* as a response to *canary*, but only 6% give *canary* as a response for *bird*. By contrast, symmetrical association is mutual, that is, it tends to be equally strong in both directions. A good example of symmetric association is *good* and *bad*: The percentages of the subjects give *good* as a response for *bad* and the other way around are 75% and 76%, respectively.

Nearly all corpus-based statistical association measures are nondirectional. In the present study, we have confined ourselves mainly to the symmetrical measures. Among the association measures applied in the CLAD, only two measures—conditional probability and reverse conditional probability—yield asymmetric association strengths for word pairs. The directionality of word association has received increasing attention in recent years. Some recent studies address the importance of investigating asymmetry in word association and/or introduce directional association measures (e.g., Gries, 2013; Hutchison, Heap, Neely, & Thomas, 2014; Michelbacher et al., 2011). Currently, there are only a few asymmetrical association measures, and their abilities to account for behavioral data have been mixed (Gries & Ellis, 2015). In view of the theoretically more precise computation that directional measures can bring about, it is hoped that a sufficient number of asymmetrical measures will emerge that we can test in the near future.

Table 1 Contingency table of observed frequencies of (co-)occurrences of a word pair

$a = f(xy)$	$b = f(x\bar{y})$	$f(x^*)$
$c = f(\bar{x}y)$	$d = f(\bar{x}\,\bar{y})$	$f(\bar{x}^*)$
$f(^*y)$	$f(^*\bar{y})$	N

$f(xy)$ = number of times word x and word y co-occur. $f(x\bar{y})$ = number of times that word x occurs, and word y does not ($\bar{y}$ = any word except y). $f(x^*)$ = sum of $f(xy)$ and $f(x\bar{y})$—that is, the occurrence frequency of x (* = any word). N = size of the corpus.

Approaches to the evaluation of word association measures

The success of extracting word association data from text corpora hinges on the effectiveness of word association measures. Budanitsky and Hirst (2006) noted three kinds of approaches to the evaluation of word association measures. The first kind is a theoretical approach that attempts to optimize preferred mathematical properties. They believed that theoretical evaluations have some uses, but are rather limited in providing evidence as to which measure is superior over another and to what extent. In another approach, association measures are evaluated with respect to their performance in the framework of a particular application with standard references (e.g., typo detection and correction). If some system in natural language processing (NLP), artificial intelligence (AI), or any applied sciences requires measurement of lexical association, different measures can be compared by checking which one makes the system yield the most effective results. Still another approach is based on comparison with human judgments. Setting human judgments as the golden standard gives the assessment of how "good" or "bad" a measure is by its congruence with human performance. These sorts of assessments

are currently the most commonly employed (Cramer, Wandmacher, & Waltinger, 2011; Johns & Jones, 2010; Recchia & Jones, 2009).

For this study we employed the third approach to evaluate the word association measures used for construction of the CLAD. We calculated coverage rates of the Chen norms by the CLAD and tested whether the CLAD can predict primed lexical decision latencies nearly as well as (or better than) the Chen norms.

The following article is organized in five sections. The Construction Pipeline of the CLAD section describes how we constructed the CLAD by using natural language processing techniques, including 55 different measures of calculating word associative strength and a 400+-million-word Chinese-language corpus. Next, the How to Use the CLAD section introduces an online query system for using the CLAD. In the How the CLAD Supplements Traditional Association Norms section, we show how the CLAD can supplement traditional association norms. The Comparing the CLAD and the Chen Norms in Predicting Priming Effects section reveals that the CLAD has a greater predictive value than the Chen norms when accounting for primed lexical decision latency. Finally, in the Discussion and Conclusion section, we discuss in a broader context the potential and limitations of automatically constructed word association databases and traditional norms, and make concluding remarks.

Construction pipeline of the CLAD

The automatic construction of a corpus-derived lexical association database involves applying association measures to the word (co-)occurrence data extracted from text corpora. Figure 1 shows the construction pipeline of the CLAD. The various components of the pipeline are described in detail below.

The text corpora

The various text corpora used for construction of the CLAD are briefly described below. The most substantial portion of the corpora was made up of carefully copyedited texts, to ensure language accuracy. The downside of using published texts is the higher concentration of formal language. To supplement with informal Chinese data, we also used a web-derived corpus. All the texts in each corpus were originally written in Chinese rather than translated to Chinese from another language.

The UDN corpus was collected from three newspapers issued by United Daily News in Taiwan from 2000 to 2012, covering a great variety of theme categories including commentary, culture, education, entertainment, fiction, health, humor, life & style, local news, money & business, people, politics, sports, tech & science, travel, and world news. News on politics, money & business, and sports were the top three contributors. We believe that these theme categories, as perceived by most people, are less dominant in real language use than in the UDN corpus, so we detected these texts through our automatic text classification methods, retained every fourth text, and deleted the others.

The Sinica Corpus was issued by Academia Sinica (http://asbc.iis.sinica.edu.tw/). The collection of texts ranged from 1981 to 2007 covering 14 different text types, including reviews, advertisements or captioned-illustrations, letters, announcements, fiction stories and allegories, prose, biographies and autobiographies, poetry, quotations, manuals, drama scripts, conversations, speeches, and minutes of meetings.

We also utilized a collection of children's and adolescent books as well as novels for adults. The books for young people were published in recent years and covered life education, stories, and natural science. The publishing dates of the novels ranged from the 1940s to the 2010s, and mostly included science fiction and romance novels.

The only web corpus we used was derived from a bulletin board system called PTT, managed and used mostly by college students in Taiwan (https://www.ptt.cc/index.html). The corpus was drawn from all posts from 2000 to 2016 on the Happy, Sad, Angry, and Hate boards. These web texts were cleaned by removing disqualified context, such as signature files and tag clouds.

The total number of words in the corpora surpasses 400 million; the contribution of each corpus is shown in Table 2.

Extracting the keywords

We use the nomenclature of "keyword" and "associate" for the CLAD in the place of "stimulus" and "response," which are reserved for when referring to association norms. Prior to the extraction of keywords, all texts in the corpora were processed by the Chinese word segmentation and tagging tool developed by Academia Sinica (Tsai & Chen, 2004). A total of 84,674 keywords were extracted by excluding low-frequency, extra-high-frequency, and certain nongeneral words that fall into the criteria below.

1. Words with frequencies below 100.
2. Functional words—that is, words of the five categories: conjunctions, exclamations, prepositions, particles, and pronouns.
3. The 100 highest-frequency words. Most of these words are also function words.
4. Proper nouns with frequencies lower than 500.
5. Time words and quantity words in which numerical characters were more frequent than nonnumerical characters. Numerical characters refer to the Chinese equivalents of the

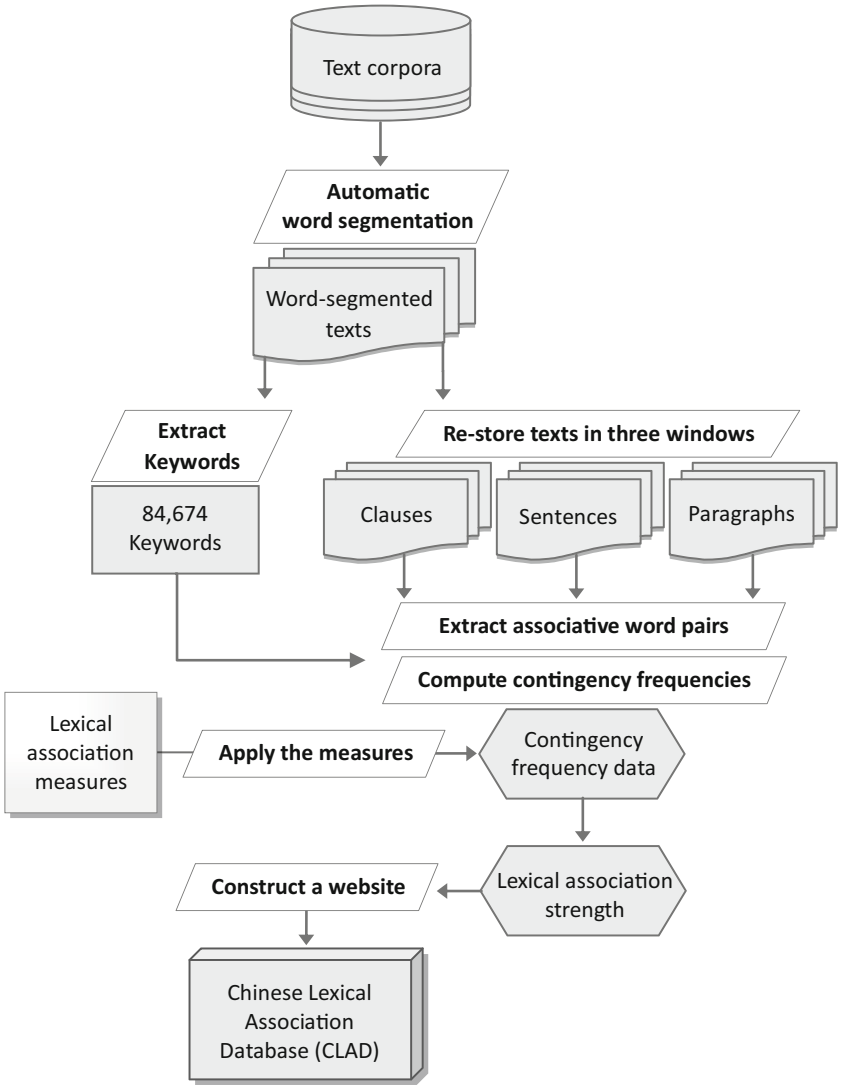


Fig. 1 Construction pipeline of the Chinese Lexical Association Database (CLAD)

ten Arabic numerals. For example, 二十日 "the twentieth day of a month" (this three-character string is treated as a single word by most Chinese word segmenters) includes two numerical characters (二 "two" and 十 "ten") and one nonnumber (日 "day"). Because number characters exceeded nonnumbers, the word was excluded. However, 星期三 "Wednesday" was included, because it contains two nonnumerical characters 星 and 期, which together mean "week," and only a single number 三 "three."

Table 2 Corpora used for the construction of the CLAD and their sizes

Corpus	Number of Word Tokens
United Daily News (UDN) Corpus	253,952,479
Books for children and adolescents	99,340,090
Novels	59,307,704
PTT (a bulletin board system)	10,734,678
Sinica Corpus	9,343,428
Total	432,678,379

Restoring the corpus texts in three co-occurrence windows

In the field of lexicography, whose goal is to pinpoint fixed expressions (also called multiword expressions or collocations) such as *black box* and *roll in*, window size is typically set to five or six words (Church & Hanks, 1990; Smadja, 1993). On the one hand, inquiries of lexical association in the broader, psycholinguistic sense as discussed herein have a greater need to highlight semantic relationships, such as *doctor*/*nurse* and *doctor*/*health*, that often span over larger scales of text words. On the other hand, window size must be limited. Spence and Owens's (1990) study revealed that when window size exceeds 200 words, co-occurrence rate loses its power to explain word association. Yet it would be

wise to also make use of the textual structures pre-defined by language users, namely, the segments delimited by punctuation marks. In the end, we decided to use three co-occurrence windows—clauses, sentences, and paragraphs. By computing lexical association under different windows, we hope that the CLAD can satisfy the needs of a greater variety of users. The three windows are described in detail below:

A paragraph is a string of words delimited by a line break. Like most languages, paragraphs in Chinese indicate how subtopics are organized throughout the text. Its words should be more related among each other than they are related to the words in other paragraphs.

A sentence is separated by any two of the following punctuation marks—periods, question marks, exclamation marks, or semicolons. Unlike the English sentence, which is primarily a grammatical unit, a Chinese sentence is a discursive or rhetoric unit that performs a coherent communicative function. When translated into English, typically a Chinese sentence turns into multiple English sentences.

A clause is framed by two commas or a comma and another punctuation mark. Chinese commas differ greatly from their English equivalents: Though Chinese word strings separated by commas are sometimes similar to clauses in English, they are often more similar to the English sentence unit.

Table 3 gives the total amounts of clauses, sentences, and paragraphs in the corpora. It also shows that the three linguistically defined windows are roughly constrained by size. The mean length of a clause is 6.5 words. On average, a sentence is about four times as long as a clause, and a paragraph is nearly four times the size of a sentence.

Extracting the associative word pairs and their contingency frequency data

Regardless of sequence, any two words that appeared within a window were treated as co-occurring. For instance, a clause consisting of four different words would generate six word pairs, as illustrated below.

Original clause: 計算詞彙聯想強度 “compute word association strength”

Word-segmented clause: 計算 詞彙 聯想 強度 “compute,” “word,” “association,” “strength”
Word pairs: 計算 詞彙 “compute” “word”
計算 聯想 “compute” “association”
計算 強度 “compute” “strength”
詞彙 聯想 “word” “association”
詞彙 強度 “word” “strength”
聯想 強度 “association” “strength”

Table 3 Total number and mean size for each co-occurrence window type

Window type	Total number	Size in Words	
		Mean	*SD*
Clause	68,559,948	6.5	4.9
Sentence	17,007,069	26.3	21.7
Paragraph	4,310,468	96.0	129.6

A total of about 7% of the corpora lack paragraph information and were excluded from the construction of the paragraphs.

We then computed the co-occurring frequencies of those word pairs that were composed of the earlier extracted 84,674 words. Frequencies were defined in terms of how many times a word pair occurred in the clauses, sentences, or paragraphs of the corpora. Say the hypothetical word pair 計算+詞彙 also appeared in 1,000 other clauses, its clause window frequency would be 1,001.

One of the most serious weaknesses of co-occurrence-based lexical association measures is the tendency to obtain unstable results when the co-occurrence frequency is very small (Church & Hanks, 1990). To filter out low-frequency word pairs, we set a relative frequency threshold of *z*-score, followed by a raw frequency limit: The co-occurrence frequencies between a keyword and its co-occurring words were first transformed into *z*-scores. Word pairs with *z*-scores less than 0 were then deleted. Word pairs whose *z*-scores were higher than 0, but raw frequencies were lower than 5, were removed subsequently.

For the remaining word pairs, we computed their contingency frequencies as shown in Table 1. All frequencies were computed using co-occurrence windows as the count units, such that the sizes of the corpora stand for the total numbers of clauses, sentences, or paragraphs in the corpora.

Computing lexical association strengths and clustering the association measures

We then applied the 55 statistical association measures listed in the Appendix to these contingency frequency data to derive the word association strengths of the associative word pairs.

Tan, Kumar, and Srivastava (2004) demonstrated two theoretical scenarios in which many statistical association measures become consistent with each other. Their theoretical argument was borne out by our empirical study, which revealed that the association strengths yielded by many measures were highly correlated. The analysis we carried out was hierarchical divisive clustering, also commonly known as DIANA (Kaufman & Rousseeuw, 1990). The clustering results can be visualized by a dendrogram (see Fig. 2). Association measures conjoined at a lower branching node have stronger correlations with each other, with the distance metric between cluster members (i.e., the height) being smaller. The number of resultant clusters depends on the threshold imposed on the distance metric. We chose a relatively strict threshold, manifested by the blue line, to ensure great within-

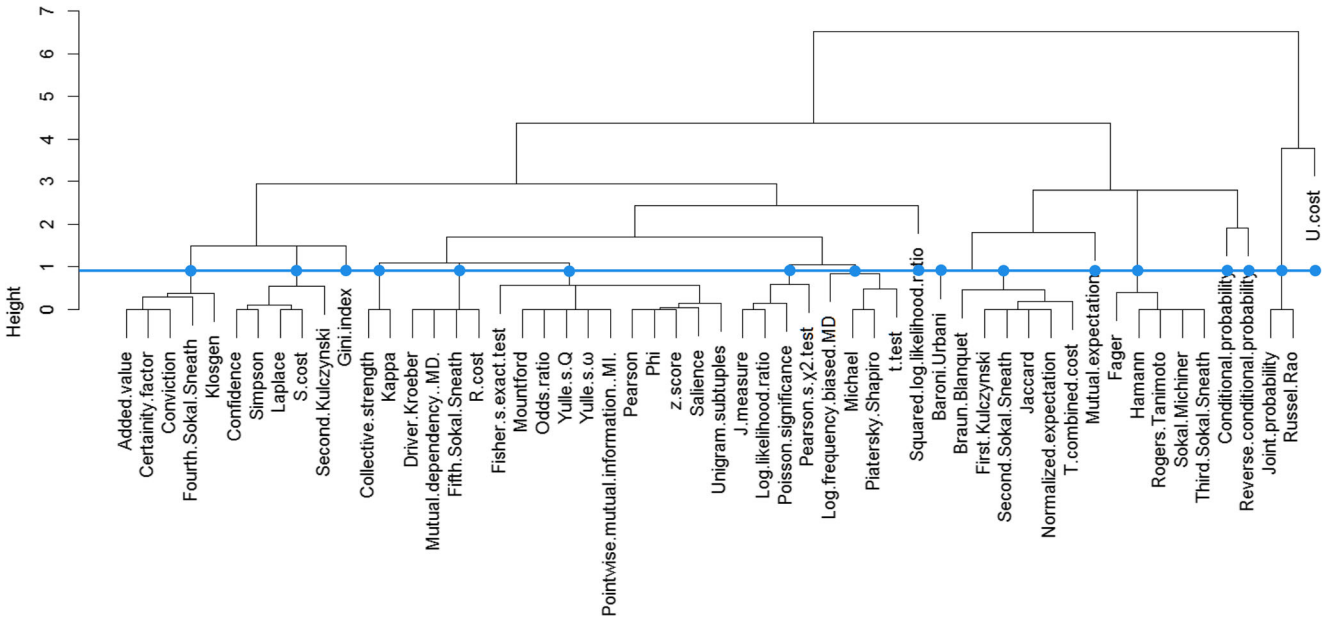


Fig. 2 Dendrogram of the word association measures applied to create the CLAD

cluster similarity. Each vertical line that the blue line crosses represents a group of measures that was identified as a cluster. Accordingly, the measures were divided into 17 clusters.

We calculated the Spearman coefficients between the measures in each cluster. Nearly all the coefficients were higher than .96. To avoid including redundant data, the CLAD presents each cluster by the measure having the highest average coefficient with the other measures in the cluster. The 17 clusters of association measures were represented by the following measures: added value, Baroni-Urbani, conditional probability, Gini index, Jaccard, joint probability, kappa, log likelihood ratio, Michael, mutual expectation, R cost, reverse conditional probability, Simpson, Sokal–Michener, squared log likelihood ratio, U cost, and unigram subtuples.

Constructing a query website

The CLAD contains the occurrence frequency and part-of-speech (POS) data of the associative word pairs and their association strengths. These were computed by the 17 representative association measures on the basis of the 400+-million-word Chinese-language corpus. A website (www.chinesereadability.net/LexicalAssociation/CLAD/) for querying and downloading these data was constructed.

How to use the CLAD

View the whole word list

The "View the whole word list" option allows the user to access all available keywords, 100 per page (see Fig. 3). The most prominent POS (i.e., the most frequently tagged part of speech) and frequency (i.e., number of occurrences in clauses, sentences, or paragraphs in the corpora) of each keyword are also displayed. The keywords are sorted by default from the smallest stroke number to the largest. Clicking on the "Word" header will show the words in the reversed order. If the user clicks on "POS" or "Frequency," the list will be sorted in accordance with those criteria instead.

Keyword search

Upon the user clicking the "Keyword search" radio button, a dialog box opens below it. Users can input one or more character(s) into the box to display all the words that start with the entered character(s). This function is exemplified with the character 美 "beautiful" in Fig. 4.

POS and frequency search

Upon clicking the "POS and frequency search" option, users can search for words of a particular POS and/or within a specific frequency range (see Fig. 5). The default setting for the POS option includes all POSs (i.e., "No restrictions"). To browse the vocabulary of a specific POS, first click on "No restrictions." Next, click on the desired POS in the drop-down menu. Each POS code is followed by a gloss.

To specify the frequency range of keywords, the user must first choose a window in the drop-down menu of "Choose the frequency window." The lowest and highest word frequencies under each window type are given in the captions. The user can then enter in the two dialog boxes underneath "Range of frequency" the upper and lower frequency bounds of the

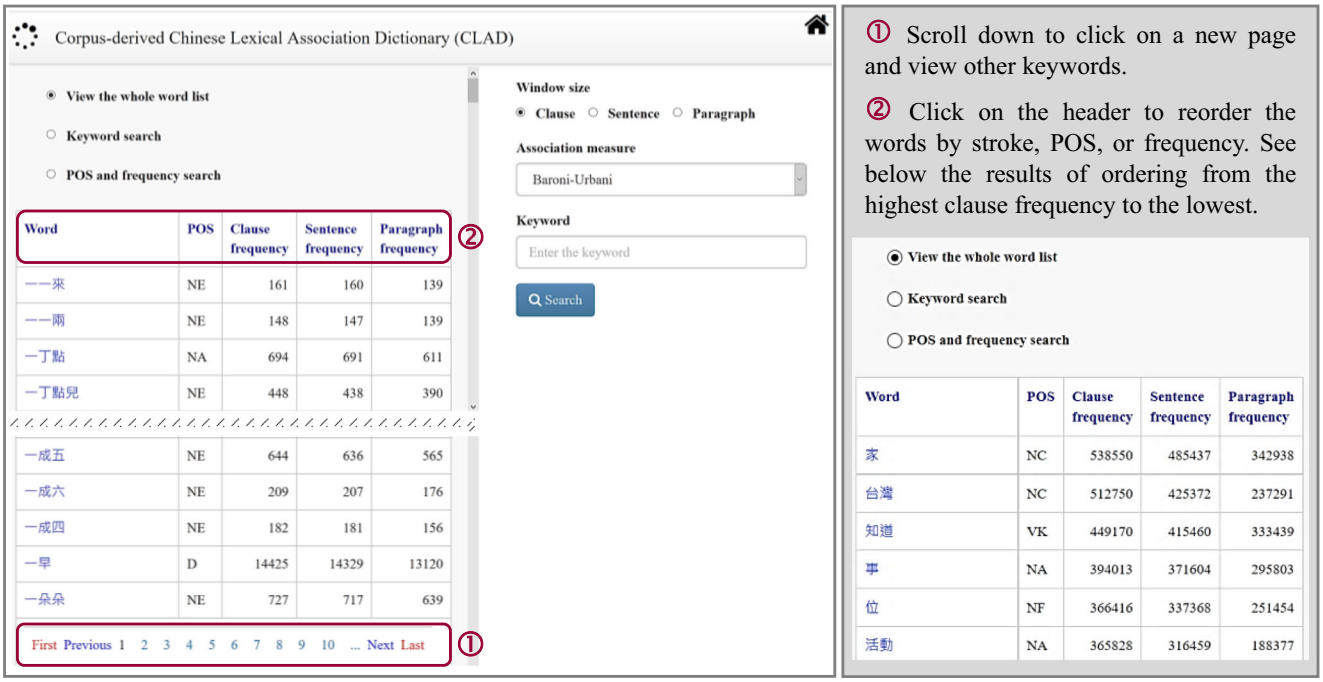


Fig. 3 Viewing the whole word list of the CLAD

keywords to be searched for. Notice that the frequency window only pertains to how often the words occur in a given window size. To specify the co-occurrence window under which association strengths were computed, the user must do another selection, to be described below.

Displaying association data

Clicking on any of the keywords shown on the left column, such as 美人魚 "mermaid," prompts the system to display in the right column its associates, as well as the co-occurrence frequencies and association strengths with the associates (see Fig. 6). In addition to using the left column to select a keyword, the user can directly input the word in the "Keyword" dialog box in the right column.

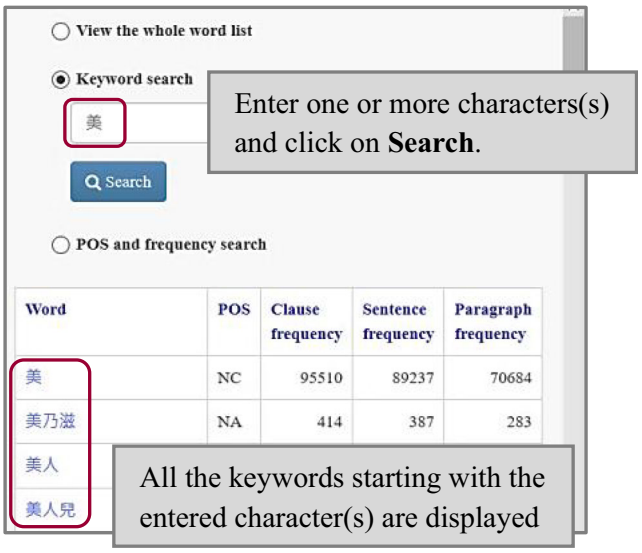


Fig. 4 Using the keyword search function

Selecting co-occurrence window and association measure

Users can choose the desired co-occurrence window of association by clicking the corresponding button for "Clause," "Sentence," or "Paragraph" (see Fig. 6). Click on "Association measure" to select from the 17 measures in the resultant drop-down menu.

Downloading the association data

Click on "Download the .txt file" at the top of the associates table to download the table to the computer (see Fig. 6). The user can drag the icon of the downloaded plain text file to a spreadsheet editor such as Microsoft Excel for optimizing visual display or for further analysis.

How the CLAD supplements traditional association norms

Enlarging the size of lexical association data

The Chen norms described above is the largest set of Chinese word association norms at present. The CLAD provides a greater wealth of lexical association data than the Chen norms

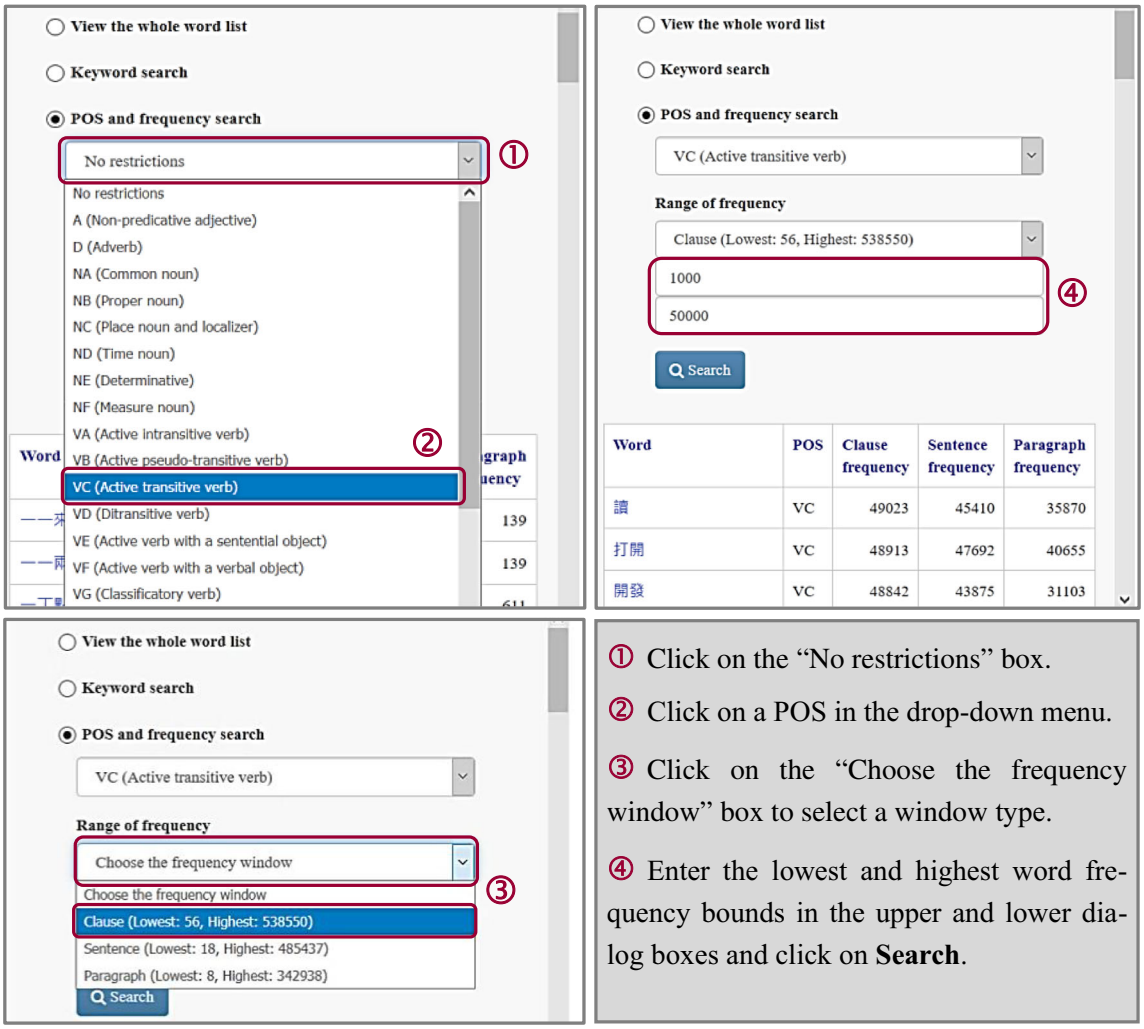


Fig. 5 Using the part-of-speech (POS) and frequency search function

in terms of both the total number of associative word pairs and the average number of associates for each word.

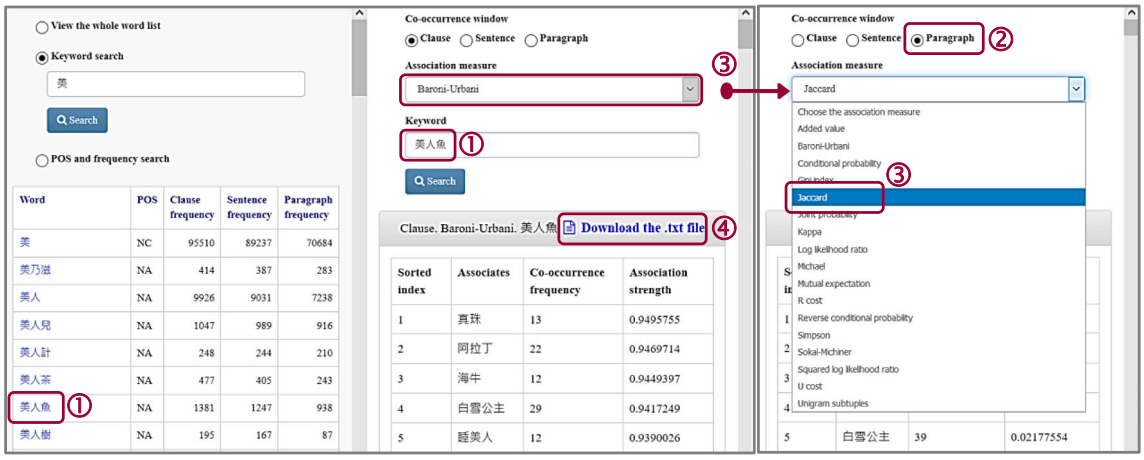
In terms of the overall size, the Chen norms include about 100,000 word pairs, whereas the CLAD consists of nearly 17 million word pairs using the clause window, more than 165 times that of the Chen norms. When using the sentence or the paragraph window, the CLAD is 644 or 2,015 times larger than the Chen norms, respectively.

For each stimulus word, the Chen norms elicited 86 different response words on average. Under the clause window, the CLAD has, on average, 401 different associates for each keyword, more than four times the number of the Chen norms. Using the sentence or paragraph window, the average increases to 1,523 or 4,516 associates for a keyword, more than 17 or 52 times, respectively, the number for the Chen norms (see Table 4).

Increasing the differentiation of association strengths

The relatively small numbers of participants involved in the creation of traditional norms have led to associative strength measurements that could be more precise. For an average stimulus in the Chen norms, 68.8% of the responses had a frequency of 1 (i.e., only one participant responded with that particular word to that stimulus). Normative association strengths are computed by dividing the number of particular responses with the number of respondents. Since the number of respondents is the same for every stimulus word in the Chen norms, the association strengths of these once-occurring stimulus–response pairs are identical. In other words, the normative statistics suggests the counterintuitive notion that the majority of associated words are bound with very similar strengths.

In contrast, words in the CLAD are related to the large majority of their associates with different strengths. An average of 98% of the associates of a word can be differentiated by unique association strengths if applying Baroni-Urbani or 11 other association measures. Unique association strengths make up 94% of the associations of an average word for the measures Sokal–Michener and U cost. In contrast to the highly unique strengths of those measures, the average percentages of unique strengths



① Click on a keyword in the left column, or simply enter the keyword in the right column to display its association data.

② Click to select another co-occurrence window.

③ Click on **Association measure** to choose another measure in the drop-down menu, and display the association data computed with the measure.

④ Click to download the association data.

Fig. 6 Selecting the co-occurrence window and association measure to display and download association data

for the other three measures—Simpson, conditional probability, and joint probability—ranged from 8% to 15%.

Covering the traditional word association norms

For the CLAD to be a good supplement to traditional norms, it should not only provide a greater number of associative data, but also cover the traditional norms (here, the Chen norms). Coverage of the Chen norms can be assessed by discovering how many responses are given with corresponding association strengths in the CLAD. However, the CLAD contains a substantially greater number of associative pairs with considerably higher associative strength differentiation, so it does not seem adequate to assess coverage in a word-for-word fashion such as by computing their correlation coefficients. Considering that responses occurring more frequently in the Chen norms should have relatively, but not exactly, higher association strengths in the CLAD, we utilized the summed relative frequency (SRF) to compute the coverage rate.

Computing coverage rate Taking the stimulus 音樂 "music" as an example, it is described below how to assess whether the responses of an individual stimulus in the Chen norms were covered by the CLAD. Coverage of the norms was then calculated by averaging the coverage rates of the stimuli.

In the Chen norms, the most frequently elicited response to 音樂 "music" was 古典 "classical," totaling ten tokens (i.e., there were ten participants who responded with 古典 "classical" to 音樂 "music"). The response 欣賞 "to enjoy, to appreciate" had the second highest frequency of nine, followed by

Table 4 Sizes of the Chen norms and the CLAD

Database	Number of Keywords/Stimuli	Number of Associative Word Pairs	Number of Associates/Responses per Word	
			Mean	*SD*
CLAD				
Clause window	84,674	16,997,424	401	1,338
Sentence window	84,674	66,313,915	1,523	3,569
Paragraph window	84,674	207,538,494	4,516	6,889
Chen norms	1,200	103,006	86	17

the responses 美妙 "pleasant, wonderful" and 歌手 "singer," both occurring eight times. To compute the SRF of a response, we first summed the frequencies of both the response and the other responses of higher response frequency. The summed frequency was then divided by the total number of response tokens of the stimulus, which was always 200. Accordingly, the SRFs of 古典 "classical" and 欣賞 "to enjoy, to appreciate" were 10/200 = 5% and (9+10)/200 = 9.5%, respectively. 美妙 "pleasant, wonderful" and 歌手 "singer" occurred equally frequently, so they had the same SRF of (8+8+9+10)/200 = 17.5%. The calculation of SRFs continued for the other response words in the same way.

These SRFs were then mapped onto the associates of 音樂 "music" in the CLAD. Coverage of 古典 "classical" was affirmed if it was one of the associates with the highest 5% of association strengths in the CLAD. Suppose 音樂 "music" has 100 associates in the CLAD, then 古典 "classical" must be among the five associates (100×5% = 5) with the highest associative strengths in order to be deemed to be covered. Coverage was negated if 古典 "classical" was not an associate of 音樂 "music," or if it belonged to the 95% of associates with weaker associative strengths. Likewise, for 欣賞 "to enjoy, to appreciate" or 美妙 "pleasant, wonderful" 歌手 "singer" to be covered, they must be among the strongest 9.5% or 17.5% of the associates in the CLAD. The checking continued for the other responses in the same way. The number of covered responses was divided by the total number of different responses (i.e., response types) to return the coverage rate of 音樂 "music."

Coverage rates computed using associates and stimuli in varying strength and frequency ranges For the following two reasons, we hold that coverage rates computed using stimuli with different word frequencies and associates of different association strengths (i.e., instead of only using all the associates and all the stimuli) can be used to help evaluate the CLAD more efficiently. We will then make two predictions for a good corpus-derived word association database regarding how coverage rates should change with the variation of word frequencies and associative strengths.

A terminological distinction should be noted at this juncture: We use the term "word frequency" to denote how often a word occurs in the corpora, whereas "response frequency" refers to how frequently a word was elicited in a word association test. High- and low-frequency words are thus the more common and the relatively rare words in the corpora.

Our first reason is based on the fact that in comparison to low-frequency words, high-frequency words possess larger numbers of co-occurring instances in text corpora. In statistical terms, the larger sample reduces the probability of inaccurate computation by the association measure. Accordingly, corpus-derived lexical association data obtained using high-frequency words should enjoy a higher likelihood of reflecting lexical relationships in the mental lexicon. One can therefore make the prediction that the number of normative responses covered by the CLAD should increase when using stimuli with increasing word frequencies.

The second reason is related to the assumption underlying the entire history of free association study; that is, words with stronger cognitive bonds are easier to retrieve, so that more frequently elicited responses possess stronger association strengths in the human mind. One may then reasonably assert that the Chen norms responses should be found possessing relatively high association strengths in the CLAD. To put in another way, on the whole associates in the CLAD that are also responses in the Chen norms should be more strongly associated with the keywords than those associates not found in the Chen norms. Predictably, when we increase the number of tested associates, but lower their association strengths at the same time, coverage rates should increase at a decreasing rate.

To test the two predictions, coverage rates were computed using four sets of associates and four sets of stimuli in varying strength and frequency ranges. Always starting with the associate with the strongest association strength, we picked the top 10%, the top 40%, the top 70%, and the top 100% (i.e., all) of the associates according to their association strengths in the CLAD. As such, with the number of associates increased to four, seven, and ten times larger (i.e., from 10% to 40%, 70%, and 100%), their overall association strengths gradually decreased. Likewise, always starting with the word of the highest frequency, we picked the top 25%, the top 50%, the top 75%, and the top 100% (i.e., all) of the stimuli according to their word frequencies in the corpora, so that the overall word frequencies of the four stimulus sets gradually decreased.

Table 5 presents the coverage rates computed by using the various sets of associates and stimuli under three window sizes for the 17 association measures. To examine the first prediction, we compared the coverage rates based on the four sets of stimuli. (Each comparison was done using the same set of associates to rule out the effect of associative strengths.) The critical rationale underlying our analysis is that if the word frequency of a stimulus does not play a role in the variation of coverage rate, coverage rates obtained using the four frequency-varying sets of stimuli should be identical.[1] However, if using high-frequency words to construct word association data has a better chance to align with the norms, the highest coverage rate

[1] We should be safe to assume that the different sizes of the stimulus sets can be exempted from being responsible for the variation of coverage rates, because the Chen norms employed 1,200 stimuli, and each set of stimuli contained a relatively large number of words.

should be yielded by using the stimuli at the highest frequency range of 25%. Using the stimuli at the top 50% of word frequency would yield the second highest coverage rate, and so on and so forth.

The prediction was confirmed by the vast majority of the comparisons of coverage rates shown in Table 5. For example, for the log likelihood ratio measure, the coverage rate computed using all the associates against all the stimuli under the paragraph window was .72. The rates increased to .84, .92, and .94 when the word frequencies of the tested stimuli increased to the top 75%, 50%, and finally to the highest frequency range of 25% (see the boldfaced rates in Table 5).

The opposite trend, in which higher word frequencies of the stimuli were associated with lower coverage rates, was found in a single comparison for Simpson under the paragraph window using 10% of the associates based on the stimuli of the highest 50% and 25% of word frequencies (the coverage rates were .41 and .39, respectively), in two comparisons for added value, and in several comparisons for Sokal–Michener and squared log likelihood ratio.

To test the second prediction, the coverage rates obtained using 40%, 70%, and 100% of the associates were divided by the coverage rate obtained using the strongest 10% of associates. There were three possibilities regarding the size of the resultant quotients. First, if the covered responses were evenly distributed among the CLAD associates, the only factor that affects the resultant quotients would be the number of associates used for the computation. Therefore, the quotients would be 4, 7, and 10 (40%÷10% = 4, 70%÷10% = 7, 100%÷10% = 10). The average quotient would be 7 [(4+7+10)/3 = 7].

The second possibility, which is demonstrative of our prediction for a good association database, is that association strengths of the normative word pairs are relatively high in the CLAD, such that more than 10% of the normative responses are covered by the strongest 10% of the CLAD associates. Compared with the first possibility, the divisor (i.e., the coverage rate yielded by using the strongest 10% of associates) would become larger, and the resultant quotients would be smaller than 4, 7, and 10, whose average would be smaller than 7. If the association strengths of the normative responses became even stronger in the CLAD, even more responses would be covered by the associates with high association strengths. Consequently, the resultant average quotients would become even smaller than 7.

The third possibility occurs when the normative responses tend to be weaker associates in the CLAD, resulting in average quotients that are larger than 7. Table 6 gives the actual average quotients computed for the 17 association measures. Only two measures, Sokal–Michener and squared log likelihood ratio, returned average quotients larger than 7. The other, much smaller average quotients show that many of the association measures yielded relatively high association strengths for the stimulus–response pairs in the Chen norms.

The results of testing the two predictions can be used not only to appraise the CLAD on the whole, but also to evaluate the effectiveness of the association measures individually. The first prediction result rendered Simpson, added value, Sokal–Michener, and squared log likelihood ratio less favorable than the other measures. It is probably not coincidental that these four measures also performed less well for the second prediction.

In addition to confirming the CLAD's alignment with the Chen norms, promoting the corpus-based approach to constructing lexical association data requires that we provide more supportive evidence, such as the explanatory power of the CLAD on human behavioral performance, a topic that we now turn to.

Comparing the CLAD and the Chen norms in predicting priming effects

Lexical associations derived from both text corpora and free association tests have been identified as an important variable in predicting word-priming effects (Balota & Paul, 1996; Brunellière, Perre, Tran, & Bonnotte, 2017; Günther, Dudschig, & Kaup, 2016; Hutchison, 2003; Lupker, 1984; Shelton & Martin, 1992). We are interested in discovering whether the CLAD might exhibit as much predictive power in human behavior as traditional association norms do. A word-priming experiment using a lexical decision task (LDT) was conducted for this study. In most LDT priming studies, participants are asked to make the "word" versus "nonword" decision upon reading a letter string (i.e., the target) that is preceded by a priming word. If the prime and the target are related, a priming effect is anticipated such that the reaction time would be shorter than if the two words were unrelated.

Utilizing the Chinese Lexicon Project as a baseline database

There have been recent attempts to compile mega-scale real-time behavioral responses pertaining to lexical items (e.g., the English Lexicon Project; Balota et al., 2007). They provide substantial materials on the basis of which researchers can conduct related studies and test theoretical models. A large database of Chinese lexical decision performance in a neutral (i.e., no priming) condition was published recently for more than 25,000 traditional Chinese two-character words (Chinese Lexicon Project; Tse et al., 2017). In this project, participants were asked to decide whether a two-character string visually presented to them formed a legitimate Chinese word. The reaction time and accuracy data were collected. Past large behavioral databases have been found to be quite robust with respect to relatively large sets of independent variables

Table 5 Coverage rates of the Chen norms in the CLAD, computed using stimuli and associates in varying frequency and strength

	Clause Window																Sentence Window								
	All				75%				50%				25%				All				75%				50%
	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %
Added value	.12	.30	.38	.42	.16	.39	.49	.55	.21	.50	.63	.70	.23	.55	.70	.78	.23	.46	.54	.59	.30	.57	.68	.73	.36
Baroni-Urbani	.17	.31	.38	.42	.22	.40	.49	.54	.31	.53	.63	.69	.40	.63	.73	.78	.29	.46	.54	.58	.37	.58	.67	.72	.50
Conditional probability	.17	.31	.38	.42	.22	.41	.49	.55	.29	.54	.64	.70	.36	.63	.73	.79	.28	.47	.54	.59	.36	.59	.68	.73	.46
Gini index	.17	.31	.38	.43	.23	.41	.49	.55	.31	.53	.63	.70	.37	.60	.71	.78	.32	.48	.55	.59	.41	.60	.68	.73	.52
Jaccard	.19	.33	.40	.43	.25	.43	.51	.55	.34	.57	.66	.70	.41	.65	.74	.79	.33	.50	.56	.59	.42	.63	.70	.73	.54
Joint probability	.17	.31	.38	.42	.22	.41	.49	.55	.29	.54	.64	.70	.36	.63	.73	.79	.28	.47	.54	.59	.36	.59	.68	.73	.46
Kappa	.19	.33	.39	.43	.25	.42	.50	.55	.34	.56	.64	.70	.42	.64	.72	.78	.33	.49	.55	.59	.43	.62	.69	.73	.55
Log likelihood ratio	.19	.32	.39	.43	.25	.42	.50	.55	.34	.54	.63	.70	.40	.61	.71	.78	.34	.49	.55	.59	.44	.61	.68	.73	.54
Michael	.19	.33	.39	.43	.25	.43	.50	.55	.33	.56	.64	.70	.41	.64	.72	.79	.32	.49	.55	.59	.41	.62	.69	.73	.53
Mutual expectation	.19	.33	.40	.43	.25	.43	.51	.55	.33	.56	.65	.70	.40	.65	.74	.79	.32	.50	.56	.59	.41	.63	.70	.73	.51
R cost	.18	.33	.39	.43	.23	.42	.51	.55	.31	.55	.65	.70	.38	.63	.73	.79	.33	.50	.56	.59	.42	.62	.70	.73	.53
Reverse conditional probability	.10	.29	.38	.42	.14	.37	.49	.54	.18	.48	.62	.69	.21	.53	.69	.77	.19	.43	.54	.59	.24	.54	.67	.73	.30
Simpson	.12	.29	.38	.42	.16	.38	.48	.54	.21	.49	.62	.69	.23	.55	.70	.78	.22	.44	.53	.58	.28	.56	.67	.73	.35
Sokal–Michener	.03	.16	.27	.37	.04	.20	.34	.47	.05	.24	.41	.59	.04	.22	.41	.64	.05	.19	.32	.49	.06	.22	.38	.60	.05
Squared log likelihood ratio	.04	.19	.30	.39	.06	.25	.39	.50	.07	.30	.48	.62	.06	.29	.49	.68	.07	.24	.37	.52	.08	.29	.45	.63	.08
U cost	.05	.18	.29	.38	.07	.23	.37	.49	.08	.29	.47	.62	.10	.33	.53	.70	.10	.29	.42	.54	.12	.35	.52	.66	.14
Unigram subtuples	.14	.31	.39	.43	.18	.40	.50	.55	.24	.52	.64	.70	.29	.59	.72	.78	.27	.48	.56	.59	.34	.60	.69	.73	.42

	Sentence Window							Paragraph Window																
	50%			25%				All				75%				50%				25%				
	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All	10 %	40 %	70 %	All
Added value	.68	.80	.86	.36	.70	.83	.90	.30	.55	.66	.71	.37	.66	.78	.84	.41	.74	.86	.92	.38	.73	.86	.93
Baroni-Urbani	.71	.80	.85	.59	.81	.88	.91	.36	.56	.64	.70	.45	.67	.75	.82	.58	.80	.87	.91	.65	.87	.92	.94
Conditional probability	.74	.82	.86	.54	.81	.88	.91	.35	.58	.66	.71	.44	.71	.80	.84	.55	.83	.90	.92	.59	.87	.92	.94
Gini index	.73	.81	.86	.58	.78	.86	.91	.41	.60	.67	.71	.51	.72	.80	.84	.62	.83	.89	.92	.66	.85	.91	.94
Jaccard	.76	.83	.86	.60	.82	.88	.91	.42	.61	.68	.71	.52	.73	.80	.84	.62	.84	.90	.92	.64	.87	.92	.94
Joint probability	.74	.82	.86	.54	.81	.88	.91	.35	.58	.66	.71	.44	.71	.80	.84	.55	.83	.90	.92	.59	.87	.92	.94
Kappa	.75	.81	.86	.62	.81	.86	.91	.43	.61	.68	.71	.53	.73	.80	.84	.64	.84	.89	.92	.66	.87	.92	.94
Log likelihood ratio	.73	.81	.86	.60	.78	.86	.91	.44	.62	.68	**.72**	.54	.74	.80	**.84**	.64	.84	.89	**.92**	.67	.86	.91	**.94**
Michael	.75	.82	.86	.60	.81	.86	.91	.39	.60	.68	.71	.49	.73	.80	.84	.60	.84	.90	.92	.64	.88	.92	.94
Mutual expectation	.75	.83	.86	.57	.82	.88	.91	.42	.62	.68	.72	.50	.74	.81	.84	.59	.84	.90	.92	.61	.87	.92	.94
R cost	.75	.83	.86	.60	.81	.88	.91	.44	.62	.68	.72	.53	.74	.81	.84	.63	.84	.90	.92	.67	.87	.92	.94

Table 5 (continued)

Reverse conditional probability	.64	.78	.85	.32	.67	.83	.90	.24	.52	.64	.71	.29	.61	.75	.83	.32	.68	.83	.91	.33	.69	.85	.93		
Simpson	.68	.80	.86	.36	.71	.84	.90	.28	.54	.65	.70	.35	.66	.77	.83	.41	.74	.86	.92	.39	.74	.87	.93		
Sokal–Michener	.21	.39	.69	.04	.16	.33	.70	.06	.19	.31	.57	.06	.19	.33	.66	.04	.14	.29	.70	.04	.12	.25	.70		
Squared log likelihood ratio	.29	.48	.72	.06	.23	.42	.73	.08	.23	.37	.60	.08	.24	.39	.68	.05	.19	.35	.72	.04	.14	.29	.71		
U cost	.40	.60	.77	.21	.54	.73	.86	.15	.39	.54	.66	.17	.46	.63	.77	.22	.55	.73	.86	.32	.70	.83	.92		
Unigram subtuples	.72	.82	.86	.47	.77	.87	.91	.35	.59	.67	.71	.43	.70	.79	.84	.49	.79	.88	.92	.50	.81	.90	.94		

(Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). The neutral reaction time data utilized in the present priming experiment were derived from the Chinese Lexicon Project.

The priming experiment

Method

Participants A total of 60 native Chinese speakers, 25 males and 35 females, aged between 21 and 43 participated in the experiment. All worked as either research assistants or post-doctoral fellows at National Taiwan Normal University. None of them are on the present research team, and all volunteered to take part in the priming experiment.

Materials In all, 102 related and 102 unrelated prime–target pairs were selected using the following criteria and procedure. First, we extracted all associative word pairs on the basis of both the Chen norms and the CLAD. We used the CLAD data under the paragraph window as it provided a much larger pool than the clause and sentence windows. The words also had to appear in the Chinese Lexicon Project. The qualified word pairs were then grouped according to their association strengths in the Chen norms. In each group, word pairs were sorted by the word frequencies of target words (i.e., the Chen norm responses/CLAD associates). We then selected word pairs at equal intervals of target word frequencies, such that all ranges of word frequencies were well represented. The interval was identical for all the groups (every 200th pair) except for the group of the least strong association (i.e., the associative pairs occurring only once in the Chen norms), whose interval was doubly large (every 400th pair). The percentages of the tested word pairs according to their normative association strengths are given in Table 7.

The unrelated word pairs were created by recombining the primes and targets of the related pairs, with the constraint that the paired words were not associated in either the forward or the backward direction in either the Chen norms or the CLAD. We executed a selection algorithm on a computer over 1,000 iterations, and 102 pairings was the maximum number that could be extracted.

A total of 102 nonwords were generated by recombining the distinct characters of the tested words. Three native speakers made the word-versus-nonword judgment without knowing the purpose of the study beforehand. The characters were recombined several times until uniform agreement of nonword was reached by the raters. The words that were paired with the nonwords were different from the primes and targets.

Two lists of testing items were constructed and counterbalanced so that each word occurred only once in a list. Each participant was tested on one of the two lists. As a result, each testing item received responses from 30

Table 6 Tendencies of the association measures to align with the Chen norms

Measure	Window Type		
	Clause	Sentence	Paragraph
Added value	2.98	2.25	2.11
Baroni-Urbani	2.04	1.65	1.57
Conditional probability	2.13	1.78	1.70
Gini index	2.04	1.59	1.49
Jaccard	1.91	1.57	1.50
Joint probability	2.13	1.78	1.70
Kappa	1.88	1.53	1.46
Log likelihood ratio	1.88	1.51	1.42
Michael	1.93	1.58	1.56
Mutual expectation	1.94	1.62	1.55
R cost	2.05	1.57	1.45
Reverse conditional probability	3.35	2.60	2.53
Simpson	2.96	2.29	2.14
Sokal–Michener	8.24	7.74	7.77
Squared log likelihood ratio	6.96	6.25	7.28
U cost	5.48	4.07	3.23
Unigram subtuples	2.63	1.94	1.80

Smaller values indicate a stronger tendency to align.

participants. Each list was divided into two blocks, with each block consisting of 102 trials (25 or 26 related pairs, 26 or 25 unrelated pairs, and 51 word–nonword pairs). The order of the blocks was also counterbalanced across participants. The order of the word pairs in each block were randomized for each participant.

Procedure The experiment was run on a PC with 19-in. flat-screen display via the E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). The instructions emphasized both speed and accuracy. Prior to the experimental session, participants took a practice session consisting of 20 randomized trials (five related pairs, five unrelated pairs, and ten word–nonword pairs). The displayed words were presented in font size 36. Participants read each two-character word from left to right. Prime and target words were displayed in the font types 新細明體 (PMingLiU) and 標楷體 (DFKai-SB), respectively. Each trial began with a fixation mark (+) appearing in the center of the screen for 500 ms. The prime appeared for 200 ms, followed by a blank screen for 100 ms (stimulus onset asynchrony (SOA) = 300 ms). After the blank, the target appeared, which remained on the screen until the participant pressed the key labeled either 是 (the L key) or 否 (the A key), indicating word and nonword status, respectively. Once they pressed a key, participants received a 200-ms message that informed them whether they gave the right or the wrong answer. The next trial began after a 1,500-ms blank-screen intertrial interval. There was a short break between the two blocks of the experimental session. Participants were asked to determine the length of the break on their own while remaining in their seats.

Results

Researchers have demonstrated that *z*-score transformation of raw reaction times in a priming experiment reduces variance across participants (Bush, Hess, & Wolford, 1993; Faust, Balota, Spieler, & Ferraro, 1999), and consequently increases reliability of item-based analysis (Hutchison, Balota, Cortese, & Watson, 2008). The results of the present experiment thus underwent a procedure of standardization that transformed each raw reaction time (henceforth, raw RT) into a *z*-score (*z*RT) based on a participant's overall RTs. Then we trimmed the results by discarding reactions greater than two standard deviations from the mean. The trimming removed 2.65% of the overall reactions. Applying an intersubject Cronbach's alpha reliability test to both the raw RTs and the *z*-scores attested the variance-reducing effect of *z*-score transformation: For one testing list, the Cronbach's alpha coefficients based on the raw RTs and *z*RTs were .62 and .71, respectively. For the other testing list, the coefficients were .72 and .83, respectively.

For a one-way analysis of variance test of the priming effect (i.e., a significant reaction time difference between the related and unrelated word pairs), the *F* value also increased

Table 7 Percentages of word pairs in the Chen norms and the priming experiment, according to their normative association strengths

Association Strength in the Chen Norms	Percentage in the Chen Norms	Percentage in the Priming Experiment
≧.03	7.2	15.7
≧.015 & < .03	10.5	26.5
= .01	13.2	21.6
= .005	69.2	36.3

Association strength is calculated by dividing the number of particular responses by the number of respondents, which is always 200 for a stimulus in the Chen norms.

as a result of the *z*-score transformation. For the raw-RT-based priming effect, the mean was 25.51 ms (SD = 67.41 ms), $F(1, 202) = 7.22$, $p < .01$. When *z*-scores were used instead, the statistics were $M = 0.11$, $SD = 0.23$, $F(1, 202) = 9.32$, $p < .01$.

Multiple regression analyses

The influence of word-related features in semantic priming has been investigated in several studies. In terms of the analytics adopted, these studies fall into two groups (Mandera, Keuleers, & Brysbaert, 2017). The traditional group of factorial design analysis has been shown to contain several potential confounds (Balota et al., 2004; Forster, 2000). For example, there are few, if any, words whose characteristics differ in only one dimension. As a result, the item sets can hardly be matched in the other dimensions across conditions. Researchers' implicit knowledge may bias their selection of test items, which would then work to achieve their desired effects. Researchers also tend to utilize items that have extreme values across a certain characteristic such that list contexts often vary across experiments. Most importantly, because many of the semantic or psychological features of words

Table 8 Means, standard deviations, and ranges for the predictor variables used in the regression analyses

Predictor Variables		Mean	*SD*	Range
Prime	Stroke	21.61	6.07	(7, 43)
(related and unrelated)	Word frequency	7.81	1.47	(4.81, 10.65)
	Left orthographic neighbors	59.36	54.10	(0, 298)
	Right orthographic neighbors	102.07	121.31	(1, 654)
	zRT_{CLP}	– 0.13	0.43	(– 0.87, 1.05)
Target	Stroke	21.48	6.87	(5, 40)
	Word frequency	8.92	1.29	(5.87, 11.87)
	Left orthographic neighbors	80.5	74.61	(2, 289)
	Right orthographic neighbors	102.16	131.35	(2, 654)
	zRT_{CLP}	– 0.32	0.33	(– 0.95, 1.17)
Associative	Chen Norms			
	Forward	.016324	.016619	(.005, .105)
	Backward	.225938	.280689	(.004, 1)
	CLAD			
	Added value	0.048222	0.076665	(– 9.4e-05, 0.515925)
	Baroni-Urbani	0.551819	0.202924	(0.058325, 0.903283)
	Conditional probability	0.038788	0.057208	(0.000319, 0.449438)
	Gini index	6e-06	1.8e-05	(4.63e-10, 8.8e-05)
	Jaccard	0.00828	0.012235	(9.5e-05, 0.071429)
	Joint probability	2.5e-05	4.6e-05	(1e-06, 0.000274)
	Kappa	0.015398	0.023186	(– 0.000159615, 0.133028)
	Log likelihood ratio	513.248121	1,134.09	(0.180993, 6,932.112495)
	Michael	9.30e-05	0.000177	(2e-06, 0.001083)
	Mutual expectation	1.00e-06	4.00e-06	(4.42e-10, 0.000029)
	R cost	0.002334	0.007047	(3e-06, 0.058224)
	Reverse conditional probability	0.023922	0.060867	(9.6e-05, 0.518809)
	Simpson	0.051452	0.077473	(0.002340824, 0.518809)
	Sokal–Michener	0.996286	0.004341	(0.975672, 0.999796)
	Squared log likelihood ratio	16.441521	30.86232	(0.174048, 201.581215)
	U cost	0.351233	0.263277	(0.002478, 0.979513)
	Unigram subtuples	2.079332	1.534075	(– 1.313466, 6.388045)

Stroke = number of strokes of the two characters of a word. Word frequency = logarithmic transformation (base *e*) of word frequency in the text corpora described in Table 3. Left orthographic neighbors = number of words sharing the same first character. Right orthographic neighbors = number of words sharing the same second character. zRT_{CLP} = standardized RT according to the Chinese Lexicon Project. Associative = Chen and CLAD association strengths computed using the various measures. Forward = proportion of the target in the overall responses to the prime in the Chen norms. Backward = proportion of the prime as a stimulus in the overall associations in the Chen norms when the target was a response.

Table 9 R^2 of the multiple regression models and beta weights for the lexical and behavioral variables predicting the priming effect

	Model R^2	Un Prime Stroke	Un Prime L Ortho	Un Prime R Ortho	Un Prime Freq	Un Prime zRT_{CLP}	Rel Prime Stroke	Rel Prime L Ortho	Rel Prime R Ortho	Rel Prime Freq	Rel Prime zRT_{CLP}	Target Stroke	Target L Ortho	Target R Ortho	Target Freq	Target zRT_{CLP}
Added value	.24	.08	.15	– .01	.06	.14	.01	– .04	– .05	– .15	– .10	.06	– .01	.15	– .33†	.12
Baroni-Urbani	.29*	.08	.10	– .01	.13	.19	.00	– .05	– .06	– .17	– .11	.07	.02	.16	.02	.23
Conditional probability	.24	.08	.16	.00	.06	.15	.01	– .05	– .06	– .13	– .12	.05	– .02	.15	– .35†	.12
Gini index	.26†	.10	.17	– .03	.09	.16	.00	– .07	– .05	– .23	– .14	.05	– .02	.14	– .31†	.14
Jaccard	.29*	.11	.13	.00	.18	.25†	.02	– .10	– .09	– .27†	– .16	.05	– .01	.17	– .21	.17
Joint probability	.28*	.11	.16	– .01	.14	.19	.03	– .10	– .06	– .32*	– .15	.06	– .01	.15	– .34†	.14
Kappa	.29*	.11	.13	.00	.18	.25†	.02	– .10	– .09	– .27†	– .16	.05	– .01	.17	– .21	.17
Log likelihood ratio	.29*	.11	.16	– .03	.16	.21	.02	– .10	– .06	– .30*	– .16	.04	– .01	.15	– .29	.14
Michael	.28*	.11	.16	– .02	.15	.19	.03	– .10	– .06	– .32*	– .15	.06	– .01	.15	– .33†	.14
Mutual expectation	.29*	.10	.18	– .02	.16	.22†	.04	– .11	– .08	– .27†	– .15	.03	– .01	.15	– .29	.13
R cost	.26†	.09	.16	– .01	.10	.19	.02	– .06	– .07	– .20	– .14	.05	– .02	.15	– .29	.14
Reverse cond. prob.	.24	.08	.14	– .02	.05	.14	.00	– .04	– .04	– .16	– .08	.08	.00	.15	– .30	.12
Simpson	.24	.08	.15	– .01	.05	.14	.01	– .04	– .05	– .15	– .10	.06	– .01	.15	– .33†	.12
Sokal–Michener	.25	.05	.13	– .03	.08	.14	.02	– .05	– .06	– .09	– .09	.06	– .02	.14	– .12	.14
Squared log likelihood	.24	.08	.14	– .01	.04	.14	.01	– .05	– .06	– .08	– .08	.06	– .01	.14	– .27	.12
U cost	.27*	.08	.11	.01	.02	.15	.01	– .04	– .07	– .20	– .07	.06	– .03	.12	– .16	.20
Unigram subtuples	.26†	.09	.12	– .01	.11	.17	– .01	– .05	– .06	– .13	– .10	.06	.01	.17	– .22	.15

Un = unrelated; Rel = related; Freq = frequency; L = left; R = right; Ortho = orthographic neighbors; zRT_{CLP} = *z*-score standardized reaction time from the Chinese Lexicon Project (Tse et al., 2017). * $p < .05$; † $p < .10$.

Table 10 Beta weights for the associative variables predicting the priming effect

CLAD		Chen Norms	
		Forward	Backward
Added value	.09	.12	– .01
Baroni-Urbani	.37*	– .01	– .01
Conditional probability	.12	.09	– .01
Gini index	.22†	.07	– .04
Jaccard	.34*	– .02	– .03
Joint probability	.30*	.05	– .02
Kappa	.34*	– .02	– .03
Log likelihood ratio	.32*	.02	– .02
Michael	.31*	.05	– .02
Mutual expectation	.31*	.03	– .01
R cost	.21	.04	– .03
Reverse cond. prob.	.10	.14	– .01
Simpson	.08	.12	– .01
Sokal–Michener	.19	.13	.03
Squared log likelihood	.08	.14	.02
U cost	.27*	.16	.03
Unigram subtuples	.21	.06	– .02

* $p < .05$, † $p < .10$.

are continuous variables, an analysis of the priming effect should indicate its extent, not just its presence or absence (McRae, De Sa, & Seidenberg, 1997).

An alternative methodological approach that could minimize the problems of factorial designs is to model the semantic priming at the item level through regression-based analysis. For example, Hutchison, Balota, Cortese, and Watson (2008) examined 15 predictor variables that modulate the size of the semantic-priming effect through a multiple regression procedure. Using a similar analytical method, we conducted simultaneous multiple regression analyses at item-based level to ascertain the value of various possible predictors in accounting for the variance of the observed priming effect.

Predictor variables The predictor variables entered into the regression model included the lexical and behavioral characteristics of the related primes, the unrelated primes, and the targets, as well as the associative strengths in the Chen norms and the CLAD. Description and summary statistics of the variables are given in Table 8.

Model R^2 and regression coefficients Because the CLAD variables based on the association measures were entered into the regression models individually, we ran 17 multiple regression analyses in total. Table 9 gives the R^2 values of the models and the standardized regression coefficients, or beta weights, for the variables, based on the related primes, the unrelated primes, and the targets. The row names in Table 9 have the sole function of specifying the regression models into which the referred associative variables were entered. The beta weights for the associative variables themselves are given in Table 10.

The regression models based on the following association measures—Baroni-Urbani, Jaccard, joint probability, kappa, log likelihood ratio, Michael, and mutual expectation—were able to account for a larger proportion of the priming variance (R^2 = .28 or .29, all ps < .05) than the other models. These association measures also turned out to be the strongest predictors among the variables, with the beta weights reaching between .30 and .37 (all ps < .05). The Chen norms did not exhibit significant predictive power in the priming results, with the largest beta weights being only .16 and .03, for forward and backward associations, respectively.

The results of the priming experiment indicate the superiority of the CLAD over the Chen norms in accounting for human performance. The multiple regression analysis also provides users a guide to which measures (i.e., those with larger beta weights) are probably more useful for applications of the CLAD. Moreover, as most association measures are derived from statistical or probabilistic models, the theoretically oriented models (or measures) need to be verified by empirical observations, for which the findings of our study can provide a useful aid. Specifically, the priming results give information on the varying degrees of validity of the association measures in extracting lexical association from text corpora. However, given the limited scope of the experiment, the generalizability of our findings needs to be investigated by further research.

The other three variables that predicted priming reliably were the word frequencies of related primes and targets, as well as the zRTs of unrelated primes (with largest beta weights of – .32, – .35, and .25, respectively). Priming was increased following words that occurred less frequently in the corpora. Greater priming was also evident when the targets were low-frequency words. Previous research has also shown that priming effects are stronger for low-frequency words than for high-frequency words (Becker, 1979; Yap, Tse, & Balota, 2009).

Other than their RTs, no unrelated-prime-oriented predictors were significant or accounted for a large amount of variance. The magnitude of influence of unrelated primes has seldom been analyzed in the past (although see Hutchison et al., 2008, who also found significant regression coefficients for the RTs of unrelated primes). More research is needed to explore their effect, and a regression analysis seems to suit this purpose better than a factorial design analysis.

As a variable equivalent to word length in alphabetical languages, the stroke number of neither the primes nor the

targets showed any significant or strong impact on priming (beta weights ranging between – .01 and .11). Such results may support the whole-word access theory that Chinese words or, more generally, Chinese characters tend to be processed as holistic perceptual units as opposed to combinations of individual elements (Chialant & Caramazza, 2013; Myers, Huang, & Wang, 2006).

The other variables based on the related primes and targets seemed to exert influences on priming in similar degrees, but in opposite directions. Priming increased with longer target $z\mathrm{RT}_{\mathrm{CLP}}$ but shorter prime $z\mathrm{RT}_{\mathrm{CLP}}$ (the largest beta weights of .23 and – .16, respectively). Targets with more right orthographic neighbors and primes with fewer right orthographic neighbors enhanced priming (the largest beta weights of .17 and – .09, respectively). Priming was greater following primes with fewer left orthographic neighbors, but this tendency was much reduced for targets (the largest beta weights of – .11 and – .03, respectively). Nonetheless, because the beta weights for these variables were not significant, the statistics gave an indication that need to be further investigated.

Among the findings of the present experiment, perhaps the most important, the one that underscores all the others, is what is shared by the information presented in Table 10 and Table 6—there is quite a close correspondence between an association measure's ability to predict word priming and the degree that it covers the Chen norms. Table 10 shows that the largest beta weights from regression analyses (i.e., at least .31) were derived from the association measures Baroni-Urbani, Jaccard, kappa, log likelihood ratio, Michael, and mutual expectation. Table 6 reveals that the same association measures displayed some of the strongest tendencies (i.e., quotients of 1.57 or below under the paragraph window) to align with the Chen norms. The other two association measures that aligned most strongly with the Chen norms were Gini index and R cost (quotients of 1.49 and 1.45). Although their beta weights from the regression analyses were not as large as .31, they output near-significant coefficients of .22 and .21, which were still greater than the other nonsignificant measures.

Although the explanatory power of the Chen norms for the current priming variances was not very impressive, the fact that the association measures with better predictive ability aligned better with the Chen norms sheds important light on how we could evaluate the norms. It is possible that the association strengths expressed by the norms are generally on the right track, but that a higher level of granularity of the association strengths is required in order to account for the variance in priming. Recall that in comparison to the mostly unique values of the association strengths in the CLAD, normative association strengths tend not to be so rigorously distinguished from each other, due to the relatively small numbers of participants involved in their creation. Because multiple regression analyses are very sensitive to nuance in the magnitude of the numerical variables, we speculate that although the Chen norms and the better-performing association measures are similar in strength, the greater differentiation of strengths yielded by the association measures resulted in their greater ability to explain priming variance.

Discussion and conclusion

To expand the size or scope of a lexical association database is a goal not unique to this study, but one shared by many of those who constructed traditional association norms. When constructing new norms, researchers often add new stimuli to existing norms. For example, the Palermo and Jenkins's (1964) association norms are an extension of the Minnesota norms (Jenkins, 1970), and the Edinburgh Word Association Thesaurus (Kiss et al., 1973) contains the stimuli used by Palermo and Jenkins. In addition to incorporating more stimuli, another way to increase norm size is to multiply the set size (i.e., the number of different responses of a stimulus) by using a continuous instead of a discrete association task. In a continuous task, participants are asked to generate more than one response in a sequence, whereas only one response is allowed for discrete associations.

The collection of associative responses for the Chen norms and several frequently cited English association norms were all accomplished through discrete association tasks. By contrast, De Deyne and Storms (2008) constructed word association norms for 1,424 Dutch words using continuous tasks. For each stimulus word, three association responses were gathered per participant, and then they compared the set sizes of the first, second, and third responses. The second and third responses led to a substantial increase of response types: The total amount of response types almost tripled that when calculating only the first response. Furthermore, the first responses were more uniform, whereas the variability increased in the second and third responses. From these results, we can see that continuous association tasks can elicit weak associates, and consequently, it is an effective method for expanding the breadth of associative norms.

Compared to discrete association, continuous association gives subjects much more time to make weaker associations. Doing so not only yields more response types, but also increases the differentiation of association strengths. However, if we were to attempt to explore all the relationships within the mental lexicon with continuous association, we might encounter difficulty removing sources of interference. As De Deyne and Storms (2008) pointed out, continuous association educes chaining and retrieval inhibition (McEvoy & Nelson, 1982). It is inadvisable to allow subjects an unlimited number of responses or to impose no time limit, for otherwise subjects will start responding to their own responses rather than to the target stimulus.

Despite the practical inability of associative norms to cover all words and their lexical relationships, we do not think that corpus-derived lexical association references such as the CLAD can completely replace norms. For example, Joyce (2005) pointed out that word association norms could enhance the ability of bilingual or learner dictionaries to assist in the recollection of terms that would otherwise remain on the tip of the tongue (Brown & McNeill, 1966). When speakers encounter the tip-of-the-tongue phenomenon, they are proactively searching for relevant vocabulary in a way similar to the free association process. Therefore, traditional association norms are likely more appropriate for constructing systems meant to assist users with recollecting words.

Although a full-blown evaluation of the association measures is outside of the scope of this article, it is pertinent to underline the need for the assessment of association measures. We would remind the readers that aside from the challenge of establishing truly objective evaluative criteria, applying association measures itself is a strenuous task. The researchers must first gather an immense amount of linguistic data, preprocess the corpus, compute all the information necessary to calculate association strengths, and only then is it possible to actually apply the measures. In view of these hurdles, we believe the CLAD can provide a wealth of research material for the evaluation of association measures and save future researchers the time and effort normally spent on prep-work.

In this study, we applied 55 statistically oriented association measures. For future research, we are excited to see that new computational algorithms for learning features embedded in language are burgeoning, such as HAL (Lund & Burgess, 1996), LSA (Landauer & Dumais, 1997), BEAGLE (Jones & Mewhort, 2007), Contextual_SOM (Zhao, Li, & Kohonen, 2011), and most recently, Word2Vec (Hsu, Lee, Chang, & Sung, 2018; Mikolov, Chen, Corrado, & Dean, 2013). We look forward to more studies in the future on the design of new methods of measuring the strength of word associations using corpus data, or on how to transform techniques in related fields, such as NLP or AI, into word association measures.

In view of the obstacles to creating more comprehensive association norms, we hope that associations distilled from large text corpora can effectively supplement traditional association norms. Furthermore, for behavior-related systems that require lexical association data, wide applicability cannot be accomplished without the supporting word association database that covers a sufficient amount of association information, and the CLAD or similar large corpus-based association databases could fulfill this role. With the CLAD, we hope to have provided researchers a convenient and comprehensive database that inspires future innovative research.

Acknowledgements This research is supported by the Chinese Language and Technology Center of National Taiwan Normal University from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan. The authors are also grateful for the support of the Higher Education Sprout Project Office and the Ministry of Science and Technology, Taiwan, under Grant MOST 104-2511-S-003-017-MY3, 104-2511-S-003-019-MY3, 104-2511-S-151-001-MY3, and 107-2511-H-003-022-MY3.

Appendix

The inventory of word association measures used in this study to compute word association strength

#	Measure	Formula	Reference
1.	Joint probability	$P(xy)$	(Giuliano, 1964)
2.	Conditional probability	$P(y\|x)$	(Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999)
3.	Reverse conditional probability	$P(x\|y)$	(Gregory et al., 1999)
4.	Pointwise mutual information	$\log\frac{P(xy)}{P(x*)P(*y)}$	(Church & Hanks, 1990)
5.	Mutual dependency (MD)	$\log\frac{P(xy)^2}{P(x*)P(*y)}$	(Thanopoulos, Fakotakis, & Kokkinakis, 2002)
6.	Log frequency biased MD	$\log\frac{P(xy)^2}{P(x*)P(*y)}+\log P(xy)$	(Thanopoulos et al., 2002)
7.	Normalized expectation	$\frac{2f(xy)}{f(x*)+f(*y)}$	(Smadja & McKeown, 1990)
8.	Mutual expectation	$\frac{2f(xy)}{f(x*)+f(*y)}\times P(xy)$	(Dias, Guilloré, Bassano, & Lopes, 2000)
9.	Salience	$\log\frac{P(xy)}{P(x*)P(*y)}\cdot\log f(xy)$	(Kilgarriff & Tugwell, 2001)
10.	Pearson's χ^2 test	$\sum_{i,j}\frac{\left(f_{ij}-\hat{f}_{ij}\right)^2}{\hat{f}_{ij}}$	(Manning & Schütze, 1999)
11.	Fisher's exact test	$\frac{f(x*)!f(\bar{x}*)!f(*y)!f(*\bar{y})!}{N!f(xy)!f(x\bar{y})!f(\bar{x}y)!f(\bar{x}\bar{y})!}$	(Pedersen, 1996)
12.	*t* test	$\frac{f(xy)-\hat{f}(xy)}{\sqrt{f(xy)(1-(f(xy)/N))}}$	(Church & Hanks, 1990)
13.	*z* score		(Berry-Rogghe, 1973)

(continued)

#	Measure	Formula	Reference
		$\frac{f(xy)-\hat{f}(xy)}{\sqrt{\hat{f}(xy)\left(1-\left(\hat{f}(xy)/N\right)\right)}}$	
14.	Poisson significance	$\frac{\hat{f}(xy)-f(xy)\log\hat{f}(xy)+\log f(xy)!}{\log N}$	(Quasthoff & Wolff, 2002)
15.	Log likelihood ratio	$-2\sum_{i,j} f_{ij}\log\frac{f_{ij}}{\hat{f}_{ij}}$	(Dunning, 1993)
16.	Squared log likelihood ratio	$-2\sum_{i,j}\frac{\log f_{ij}^2}{\hat{f}_{ij}}$	(Inkpen & Hirst, 2002)
17.	Russel–Rao	$\frac{a}{a+b+c+d}$	(Russel & Rao, 1940)
18.	Sokal–Michener	$\frac{a+d}{a+b+c+d}$	(Sokal & Michener, 1958)
19.	Rogers–Tanimoto	$\frac{a+d}{a+2b+2c+d}$	(Rogers & Tanimoto, 1960)
20.	Hamann	$\frac{(a+d)-(b+c)}{a+b+c+d}$	(Hamann, 1961)
21.	Third Sokal–Sneath	$\frac{b+c}{a+d}$	(Sokal & Sneath, 1963)
22.	Jaccard	$\frac{a}{a+b+c}$	(Jaccard, 1912)
23.	First Kulczynski	$\frac{a}{b+c}$	(Kulczynski, 1927)
24.	Second Sokal–Sneath	$\frac{a}{a+2(b+c)}$	(Sokal & Sneath, 1963)
25.	Second Kulczynski	$\frac{1}{2}\left(\frac{a}{a+b}\right)+\left(\frac{a}{a+c}\right)$	(Kulczynski, 1927)
26.	Fourth Sokal–Sneath	$\frac{1}{4}\left(\frac{a}{a+b}+\frac{a}{a+c}+\frac{d}{d+b}+\frac{d}{d+c}\right)$	(Kulczynski, 1927)
27.	Odds ratio	$\frac{ad}{bc}$	(Tan, Kumar, & Srivastava, 2004)
28.	Yulle's ω	$\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$	(Tan et al., 2004)
29.	Yulle's Q	$\frac{ad-bc}{ad+bc}$	(Tan et al., 2004)
30.	Driver–Kroeber	$\frac{a}{\sqrt{(a+b)(a+c)}}$	(Driver & Kroeber, 1932)
31.	Fifth Sokal–Sneath	$\frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	(Sokal & Sneath, 1963)
32.	Pearson	$\frac{ad-bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	(Pearson, 1950)
33.	Baroni-Urbani	$\frac{a+\sqrt{ad}}{a+b+c+\sqrt{ad}}$	(Baroni-Urbani & Buser, 1976)
34.	Braun–Blanquet	$\frac{a}{\max(a+b,a+c)}$	(Braun-Blanquet, 1932)
35.	Simpson	$\frac{a}{\min(a+b,a+c)}$	(Simpson, 1943)
36.	Michael	$\frac{4(ad-bc)}{(a+d)^2+(b+c)^2}$	(Michael, 1920)
37.	Mountford	$\frac{2a}{2bc+ab+ac}$	(Kaufman & Rousseeuw, 1990)
38.	Fager	$\frac{a}{\sqrt{(a+b)(a+c)}}-\frac{1}{2}\max(b,c)$	(Kaufman & Rousseeuw, 1990)
39.	Unigram subtuples	$\log\frac{ad}{bc}-3.29\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$	(Blaheta & Johnson, 2001)
40.	U cost	$\log\left(1+\frac{\min(b,c)+a}{\max(b,c)+a}\right)$	(Tulloss, 1997)
41.	S cost	$\log\left(2+\frac{\min(b,c)}{a+1}\right)^{-\frac{1}{2}}$	(Tulloss, 1997)
42.	R cost	$\log\left(1+\frac{a}{a+b}\right).\log\left(1+\frac{a}{a+c}\right)$	(Tulloss, 1997)
43.	T combined cost	$\sqrt{U\times S\times R}$	(Tulloss, 1997)
44.	Phi	$\frac{P(xy)-P(x*)P(*y)}{\sqrt{P(x*)P(*y)(1-P(x*))(1-P(*y))}}$	(Tan et al., 2004)
45.	Kappa	$\frac{P(xy)+P(\bar{x}\,\bar{y})-P(x*)P(*y)-P(\bar{x}*)P(*\bar{y})}{1-P(x*)P(*y)-P(\bar{x}*)P(*\bar{y})}$	(Tan et al., 2004)
46.	*J* measure	$\max[P(xy)\log\frac{P(y\|x)}{P(*y)}+P(x\bar{y})\log\frac{P(\bar{y}\|x)}{P(*\bar{y})},$ $P(xy)\log\frac{P(x\|y)}{P(x*)}+P(\bar{x}y)\log\frac{P(\bar{x}\|y)}{P(\bar{x}*)}]$	(Tan et al., 2004)
47.	Gini index	$\max[P(x*)\left(P(y\|x)^2+P(\bar{y}\|x)^2\right)-P(*y)^2$ $+P(\bar{x}*)\left(P(y\|\bar{x})^2+P(\bar{y}\,\|\bar{x})^2\right)-P(*\bar{y})^2,$ $P(*y)\left(P(x\|y)^2+P(\bar{x}\|y)^2\right)-P(x*)^2$ $+P(*\bar{y})\left(P(x\|\bar{y})^2+P(\bar{x}\,\|\bar{y})^2\right)-P(\bar{x}*)^2\,]$	(Tan et al., 2004)
48.	Confidence	$\max[P(y\|x),P(x\|y)]$	(Clark & Boswell, 1991)
49.	Laplace		(Clark & Boswell, 1991)

(continued)

#	Measure	Formula	Reference
		$\max\left[\frac{NP(xy)+1}{NP(x*)+2}, \frac{NP(xy)+1}{NP(*y)+2}\right]$	
50.	Conviction	$\max\left[\frac{P(x*)P(*\overline{y})}{P(x\overline{y})}, \frac{P(\overline{x}*)P(*y)}{P(\overline{x}y)}\right]$	(Brin, Motwani, Ullman, & Tsur, 1997)
51.	Piatetsky-Shapiro	$P(xy) - P(x*)P(*y)$	(Piatetsky-Shapiro, 1991)
52.	Certainty factor	$\max\left[\frac{P(y\|x)-P(*y)}{1-P(*y)}, \frac{P(x\|y)-P(x*)}{1-P(x*)}\right]$	(Shortliffe & Buchanan, 1975)
53.	Added value	$\max[P(y\|x) - P(*y), P(x\|y) - P(x*)]$	(Sahar & Mansour, 1999)
54.	Collective strength	$\frac{P(xy)+P(\overline{x}\,\overline{y})}{P(x*)P(*y) + P(\overline{x}*)P(*\overline{y})} \cdot \frac{1-P(x*)P(*y)-P(\overline{x}*)P(*\overline{y})}{1-P(xy)-P(\overline{x}\,\overline{y})}$	(Aggarwal & Yu, 1998)
55.	Klösgen	$\sqrt{P(xy)}$.Added value	(Klösgen, 1992)

Measures that work under the assumption of statistical independence employ both observed frequencies (as shown in Table 1) and expected frequencies $\hat{f}(xy) = f(x*)f(*y)/N$. The estimated probabilities of (co-)occurrence are expressed by *P*. *P*(*x*|*y*) is interpreted as the probability of word *x* given word *y*.

References

Aggarwal, C. C., & Yu, P. S. (1998). A new framework for itemset generation. In *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems* (pp. 18–24). New York, NY: Association for Computing Machinery. doi:https://doi.org/10.1145/275487.275490

Aggarwal, C. C., & Zhai, C. (Eds.) (2012). Mining text data. New York, NY: Springer. doi:https://doi.org/10.1007/978-1-4614-3223-4

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316. doi:https://doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 827–845. doi:https://doi.org/10.1037/0278-7393.22.4.827

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. doi:https://doi.org/10.3758/BF03193014

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230−244. doi:https://doi.org/10.1037/0022-3514.71.2.230

Baroni-Urbani, C., & Buser, M. W. (1976). Similarity of binary data. *Systematic Zoology*, *25*, 251–259.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 252–259. doi:https://doi.org/10.1037/0096-1523.5.2.252

Berry-Rogghe, G. L. M. (1973). The computation of collocations and their relevance to lexical studies. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 103–112). Edinburgh, UK: University of Edinburgh, Institute for Advanced Studies in the Humanities.

Blaheta, D., & Johnson, M. (2001). Unsupervised learning of multi-word verbs. In *Proceedings of the ACL/EACL 2001 Workshop on the Computational Extraction, Analysis and Exploitation of Collocations* (pp. 54–60). Stroudsburg, PA: Association for Computational Linguistics.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, *103*, 347–356. doi:https://doi.org/10.1016/S0004-3702(98)00055-1

Braun-Blanquet, J. (1932). Plant sociology: The study of plant communities (authorized English trans. of Pflanzensoziologie). New York, NY: McGraw-Hill.

Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic item-set counting and implication rules for market basket data. In *Proceedings of the 1997 ACM-SIGMOD International Conference on Management of Data* (pp. 255–264). New York, NY: Association for Computing Machinery. doi:https://doi.org/10.1145/253260.253325

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325–337. doi:https://doi.org/10.1016/S0022-5371(66)80040-3

Brunellière, A., Perre, L., Tran, T., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*, *70*, 1922–1934. doi:https://doi.org/10.1080/17470218.2016.1215479

Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics*, *32*, 13–47. doi:https://doi.org/10.1162/coli.2006.32.1.13

Bush, L. K., Hess, U., & Wolford, G. (1993). Transformations for within-subject designs: A Monte Carlo investigation. *Psychological Bulletin*, *113*, 566–579. doi:https://doi.org/10.1037/0033-2909.113.3.566

Charles, W. G., & Miller, G. A. (1989). Contexts of antonymous adjectives. *Applied Psycholinguistics*, *10*, 357−375. doi:https://doi.org/10.1017/S0142716400008675

Chen, K.-Y., Liu, S.-H., Chen, B., Wang, H.-M., & Chen, H.-H. (2016). Exploring the use of unsupervised query modeling techniques for speech recognition and summarization. *Speech Communication*, *80*, 49–59. doi:https://doi.org/10.1016/j.specom.2016.03.006

Chialant, D., & Caramazza, A. (2013). Where is morphology and how is it processed? The case of written word recognition. In L. B. Feldman (Ed.), Morphological aspects of language processing (pp. 55–78). Hillsdale, NJ: Erlbaum.

Chung, Y. M., & Lee, J. Y. (2001). A corpus-based approach to comparative evaluation of statistical term association measures. *Journal of the American Society for Information Science and Technology*, *52*, 283–296. doi:https://doi.org/10.1002/1532-2890(2000)9999:9999<::AID-ASI1073>3.0.CO;2-5

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*, 22–29.

Clark, P., & Boswell, R. (1991). Rule induction with CN2: Some recent improvements. In Y. Kodratoff (Ed.), Machine learning — EWSL-91 (Lecture Notes in Computer Science), Vol. 482, pp. 151–163). Berlin, Germany: Springer. doi:https://doi.org/10.1007/BFb0017011

Cover, T. M., & Thomas, J. A. (1991). Elements of information theory. New York, NY: Wiley.

Cramer, I., Wandmacher, T., & Waltinger, U. (2011). Exploring resources for lexical chaining: A comparison of automated semantic relatedness measures and human judgments. In A. Mehler, K.-U. Kühnberger, H. Lobin, H. Lüngen, A. Storrer, & A. Witt (Eds.), Modeling, learning, and processing of text-technological data structures (Studies in Computational Intelligence), Vol. 370, pp. 377–396). Berlin, Germany: Springer. doi:https://doi.org/10.1007/978-3-642-22613-7_18

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*, 570–590. doi:https://doi.org/10.1093/applin/amt056

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*, 198–205. doi:https://doi.org/10.3758/BRM.40.1.198

Deese, J. (1966). The structure of associations in language and thought. Baltimore, MD: Johns Hopkins University Press.

Dias, G., Guilloré, S., Bassano, J.-C., & Lopes, J. G. P. (2000). Combining linguistics with statistics for multiword term extraction: A fruitful association? In *Proceedings of Recherche d'informations Assistee par Ordinateur, Vol. 2* (pp. 1473–1491). Paris, France: Le Centre de Hautes Etudes Internationales *d'*informatique Documentaire.

Driver, H. E., & Kroeber, A. L. (1932). Quantitative expression of cultural relationship. *University of California Publications in American Archaeology and Ethnology*, *31*, 211–256.

Dunning, T. E. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*, 61–74.

Evert, S., & Krenn, B. (2005). Using small random samples for the manual evaluation of statistical association measures. *Computer Speech & Language*, *19*, 450–466. doi:https://doi.org/10.1016/j.csl.2005.02.005

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*, 777–799. doi:https://doi.org/10.1037/0033-2909.125.6.777

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*, 115–141. doi:https://doi.org/10.1080/02699930125908

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, *28*, 1109–1115. doi:https://doi.org/10.3758/BF03211812

Frakes, W. B., & Baeza-Yates, R. A. (Eds.). (1992). Information retrieval: Data structures and algorithms. Upper Saddle River, NJ: Prentice-Hall.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, *12*, 13–18. doi:https://doi.org/10.1111/1467-8721.01213

Giuliano, V. E. (1964). The interpretation of word associations. In M. E. Stevens, V. E. Giuliano & L. B. Heilprin (Eds.), Statistical association methods for mechanized documentation: National Bureau of Standards Miscellaneous Publication, *Vol. 269* (pp. 25–32). Washington, DC: United States Department of Commerce.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi:https://doi.org/10.1037/0022-3514.74.6.1464

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Chicago Linguistics Society, Vol. 35* (pp. 151–166). Chicago, IL: University of Chicago.

Gries, S. T. (2013). 50-something years of work on collocations. *International Journal of Corpus Linguistics*, *18*, 137–165. doi:https://doi.org/10.1075/ijcl.18.1.09gri

Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, *65*, 228–255.

Günther, F., Dudschig, C., & Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology*, *69*, 626–653. doi:https://doi.org/10.1080/17470218.2015.1038280

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), In memory of J. R. Firth (pp. 148–162). London, UK: Longman.

Hamann, U. (1961). Merkmalsbestand und Verwandtschaftsbeziehungen der Farinose: Ein Betrag zum System der Monokotyledonen. *Willdenowia*, *2*, 639–768.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*, 151–167. doi:https://doi.org/10.1016/j.cognition.2009.01.009

Hsu, F.-Y., Lee, H.-M., Chang, T.-H., & Sung, Y.-T. (2018). Automated estimation of item difficulty for multiple-choice tests: An application of word embedding techniques. *Information Processing and Management*, *54*, 969–984. doi:https://doi.org/10.1016/j.ipm.2018.06.007

Hu, J.-F., Chen, Y.-C., Zhuo, S.-L., Chen, H.-C., Chang, Y.-L., & Sung, Y.-T. (2017). Word association norms and associated responses: Reference index for 1,200 two-character Chinese words. *Bulletin of Educational Psychology*, *49*, 137–160. doi:https://doi.org/10.6251/BEP.20161111

Huang, P.-S., Chen, H.-C., Huang, H.-C., & Liu, C.-H. (2009). The development of divergent thinking test of word associative strategy (DTTWAS). *Psychological Testing*, *56*, 153–177. doi:https://doi.org/10.7108/PT.200906.0153

Huang, P.-S., Chen, H.-C., & Liu, C.-H. (2012). The development of Chinese word remote associates test for college students. *Psychological Testing*, *59*, 581–607. doi:https://doi.org/10.7108/PT.201212.0581

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A *micro*analytic review. *Psychonomic Bulletin & Review*, *10*, 785–813. doi:https://doi.org/10.3758/BF03196544

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *Quarterly Journal of Experimental Psychology*, *61*, 1036–1066. doi:https://doi.org/10.1080/17470210701438111

Hutchison, K. A., Heap, S. J., Neely, J. H., & Thomas, M. A. (2014). Attentional control and asymmetric associative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 844–856. doi:https://doi.org/10.1037/a0035781

Inkpen, D. Z., & Hirst, G. (2002). Acquiring collocations for lexical choice between near synonyms. In *Proceedings of the ACL-02 Workshop on Unsupervised Lexical Acquisition, Vol. 9* (pp. 67–

76). Stroudsburg, PA: Association for Computational Linguistics. doi:https://doi.org/10.3115/1118627.1118636

Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, *11*, 37–50.

Jenkins, J. J. (1970). The 1952 Minnesota word association norms. In L. Postman & G. Keppel (Eds.), Norms of word association (pp. 1–38). New York, NY: Academic Press. doi:https://doi.org/10.1016/B978-0-12-563050-4.50004-2

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*, 662−672. doi:https://doi.org/10.3758/PBR.17.5.662

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1−37. doi:https://doi.org/10.1037/0033-295X.114.1.1

Joyce, T. (2005). Constructing a large-scale database of Japanese word associations. *Glottometrics*, *10*, 82–98.

Justeson, J. S., & Katz, S. M. (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics*, *17*, 1−19.

Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. Hoboken, NJ: Wiley.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics*, *29*, 333–347. doi:https://doi.org/10.1162/089120103322711569

Kilgarriff, A., & Tugwell, D. (2001). WORD SKETCH: Extraction and display of significant collocations for lexicography. In *Proceedings of the ACL 2001 Collocations Workshop* (pp. 32−38). Stroudsburg, PA: Association for Computational Linguistics.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey & N. Hamilton-Smith (Eds.), The computer and literary studies (pp. 153–165). Edinburgh, Scotland: Edinburgh University Press.

Klösgen, W. (1992). Problems for knowledge discovery in databases and their treatment in the statistics interpreter explora. *International Journal of Intelligent Systems*, *7*, 649–673. doi:https://doi.org/10.1002/int.4550070707

Krenn, B. (2000). *The usual suspects: Data-oriented models for identification and representation of lexical collocations* (PhD thesis). Saarland University, Germany.

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

Kulczynski, S. (1927). Die Pflanzenassociationen der Pienenen. *Bulletin International de L'Académie Polonaise des Sciences et des Lettres, Classe des Sciences Mathématiques et Naturelles, Série B, Supplement II*, *2*, 57–203.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240. doi:https://doi.org/10.1037/0033-295X.104.2.211

Lee, L.-H., Juan, Y.-C., Tseng, W.-L., Chen, H.-H., & Tseng, Y.-H. (2015). Mining browsing behaviors for objectionable content filtering. *Journal of the Association for Information Science and Technology*, *66*, 930–942. doi:https://doi.org/10.1002/asi.23217

Li, P., Schloss, B., & Follmer, D. J. (2017). Speaking two "Languages" in America: A semantic space analysis of how presidential candidates and their supporters represent abstract political concepts differently. *Behavior Research Methods*, *49*, 1668–1685. doi:https://doi.org/10.3758/s13428-017-0931-5

Li, P., & Zhao, X. (2017). Computational modeling. In A. M. B. de Groot & P. Hagoort (Eds.), Research methods in psycholinguistics and the neurobiology of language: A practical guide (pp. 208–229). Malden, MA: John Wiley & Sons.

Liu, C.-L., Hsaio, W.-H., Lee, C.-H., Chang, T.-H., & Kuo, T.-H. (2016) Semi-supervised text classification with universum learning. *IEEE Transactions on Cybernetics*, *46*, 462–473. doi:https://doi.org/10.1109/TCYB.2015.2403573

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*, 203–208. doi:https://doi.org/10.3758/BF03204766

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior*, *23*, 709–733. doi:https://doi.org/10.1016/S0022-5371(84)90434-1

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. doi:https://doi.org/10.1016/j.jml.2016.04.001

Manning, C., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. Cambridge, MA: The MIT Press.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 3−23. doi:https://doi.org/10.1037/0278-7393.21.1.3

Matsuo, Y., & Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools*, *13*, 157–169. doi:https://doi.org/10.1142/S0218213004001466

McEvoy, C. L., & Nelson, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology*, *95*, 581−634. doi:https://doi.org/10.2307/1422189

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, *45*, 499–515. doi:https://doi.org/10.3758/s13428-012-0258-1

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 558−572. doi:https://doi.org/10.1037/0278-7393.24.3.558

McRae, K., De Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130. doi:https://doi.org/10.1037/0096-3445.126.2.99

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations in adolescents and young adults: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), The adolescent brain: Learning, reasoning, and decision making (pp. 39−66). Washington, DC: American Psychological Association. doi:https://doi.org/10.1037/13493-002

Merten, T., & Fischer, I. (1999). Creativity, personality and word association responses: Associative behaviour in forty supposedly creative persons. *Personality and Individual Differences*, *27*, 933−942. doi:https://doi.org/10.1016/S0191-8869(99)00042-2

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227−234. doi:https://doi.org/10.1037/h0031564

Michael, E. L. (1920). Marine ecology and the coefficient of association. *Journal of Animal Ecology*, *8*, 54−59.

Michelbacher, L., Evert, S., & Schütze, H. (2011). Asymmetry in corpus-derived and human word associations. *Corpus Linguistics and Linguistic Theory*, *7*, 245–276. doi:https://doi.org/10.1515/cllt.2011.012

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of Workshop at International Conference on Learning Representations*. Scottsdale, AZ.

Myers, J., Huang, Y.-C., & Wang, W. (2006). Frequency effects in the processing of Chinese inflection. *Journal of Memory and Language*, *54*, 300–323. doi:https://doi.org/10.1016/j.jml.2005.11.005

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. < http://web.usf.edu/FreeAssociation/ > .

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402–407. doi:https://doi.org/10.3758/BF03195588

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*, 299−324. doi:https://doi.org/10.1037/0033-295X.105.2.299

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*, 521–543. doi:https://doi.org/10.1287/mksc.1120.0713

Palermo, D. S., & Jenkins, J. J. (1964). Word association norms: Grade school through college. Minneapolis, MN: University of Minnesota Press.

Pecina, P. (2010). Lexical association measures and collocation extraction. *Language Resources and Evaluation*, *44*, 137–158. doi:https://doi.org/10.1007/s10579-009-9101-4

Pedersen, T. (1996). Fishing for exactness. In *Proceedings of the South Central SAS Users Group Conference* (pp. 188–200). Austin, TX.

Petrović, S., Šnajder, J., & Bašić, B. D. (2010). Extending lexical association measures for collocation extraction. *Computer Speech & Language*, *24*, 383–394. doi:https://doi.org/10.1016/j.csl.2009.06.001

Piatetsky-Shapiro, G. (1991). Discovery, analysis, and presentation of strong rules. In G. Piatetsky-Shapiro & W. Frawley (Eds.), Knowledge discovery in databases (pp. 229–248). Cambridge, MA: MIT Press.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, *107*, 786−823. doi:https://doi.org/10.1037/0033-295X.107.4.786

Preece, P. F. W. (1976). Mapping cognitive structure: A comparison of methods. *Journal of Educational Psychology*, *68*, 1–8. doi:https://doi.org/10.1037/0022-0663.68.1.1

Quasthoff, U., & Wolff, C. (2002). The Poisson collocation measure and its applications. In *Proceedings of 2nd International Workshop on Computational Approaches to Collocations* (pp. 22–23). Wien, Austria.

Rauf, S. A., & Schwenk, H. (2011). Parallel sentence generation from comparable corpora for improved SMT. *Machine Translation*, *25*, 341–375. doi:https://doi.org/10.1007/s10590-011-9114-9

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, *41*, 647−656. doi:https://doi.org/10.3758/BRM.41.3.647

Rogers, D. J., & Tanimoto, T. T. (1960). A computer program for classifying plants. *Science*, *132*, 1115–1118.

Roininen, K., Arvola, A., & Lähteenmäki, L. (2006). Exploring consumers' perceptions of local food with two different qualitative techniques: Laddering and word association. *Food Quality and Preference*, *17*, 20–30. doi:https://doi.org/10.1016/j.foodqual.2005.04.012

Russel, P. F., & Rao, T. R. (1940). On habitat and association of species of anopheline larvae in southeastern madras. *Journal of Malaria Institute India*, *3*, 153–178.

Sahar, S., & Mansour, Y. (1999). Empirical evaluation of interest-level criteria. In *SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology* (pp. 63–74). Orlando, FL. doi:https://doi.org/10.1117/12.339991

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime user's guide. Pittsburgh, PA: Psychology Software Tools.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, *12*, 315−328. doi:https://doi.org/10.3758/BF03198291

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1191–1210. doi:https://doi.org/10.1037/0278-7393.18.6.1191

Shortliffe, E. D., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*, *23*, 351−379.

Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*, *241*, 1−31.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford, UK: Oxford University Press.

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 776−784. doi:https://doi.org/10.1037/a0022531

Smadja, F. (1993). Retrieving collocations from text: Xtract. *Computational Linguistics*, *19*, 143–177.

Smadja, F. A., & McKeown, K. R. (1990). Automatically extracting and representing collocations for language generation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics* (pp. 252–259). Stroudsburg, PA: Association for Computational Linguistics. doi:https://doi.org/10.3115/981823.981855

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *38*, 1409−1438.

Sokal, R. R., & Sneath, P. H. (1963). Principles of numerical taxonomy. San Francisco, CA: W. H. Freeman and Company.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, *19*, 317–330. doi:https://doi.org/10.1007/BF01074363

Sung, Y.-T., Chang, T.-H., Lin, W.-C., Hsieh, K.-S., & Chang, K.-E. (2016). CRIE: An automated analyzer for Chinese texts. *Behavior Research Methods*, *48*, 1238–1251. doi:https://doi.org/10.3758/s13428-015-0649-1

Sung, Y.-T., Chen, J.-L., Cha, J.-H., Tseng, H.-C., Chang, T.-H., & Chang, K.-E. (2015). Constructing and validating readability models: The method of integrating multilevel linguistic features with machine learning. *Behavior Research Methods*, 47, 340–354. doi:https://doi.org/10.3758/s13428-014-0459-x

Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, *29*, 293–313. doi:https://doi.org/10.1016/S0306-4379(03)00072-3

Thanopoulos, A., Fakotakis, N., & Kokkinakis, G. (2002). Comparative evaluation of collocation extraction metrics. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC), Vol. 2* (pp. 620–625). Stroudsburg, PA: Association for Computational Linguistics.

Tsai, Y.-F., & Chen, K.-J. (2004). Reliable and cost-effective pos-tagging. *International Journal of Computational Linguistics & Chinese Language Processing*, *9*, 83–96.

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

Tseng, H.-C., Chen, B., Chang, T.-H., & Sung, Y.-T. (2019). Integrating LSA-based hierarchical conceptual space and machine learning methods for leveling the readability of domain-specific texts. *Natural Language Engineering, 25*, 331–361. doi:https://doi.org/10.1017/S1351324919000093

Tulloss, R. E. (1997). Assessment of similarity indices for undesirable properties and new tripartite similarity index based on cost functions. In M. E. Palm & I. H. Chapela (Eds.), Mycology in sustainable

development: Expanding concepts, vanishing borders. (pp. 122–143). Boone, NC: Parkway.

Wu, C.-L., & Chen, H.-C. (2017). Normative data for Chinese compound remote associate problems. *Behavior Research Methods*, *49*, 2163–2172. doi:https://doi.org/10.3758/s13428-016-0849-3

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory and Language*, *61*, 303–325. doi:https://doi.org/10.1016/j.jml.2009.07.001

Zhao, X., Li, P., & Kohonen, T. (2011). Contextual self-organizing map: Software for constructing semantic representations. *Behavior Research Methods*, *43*, 77–88. doi:https://doi.org/10.3758/s13428-010-0042-z