

Published in "Physica A: Statistical Mechanics and its Applications 491(): 898–911, 2018"
which should be cited to refer to this work.

Quantitative learning strategies based on word networks

Yue-Tian-Yi Zhao ^a, Zi-Yang Jia ^b, Yong Tang ^{c,d,*}, Jason Jie Xiong ^e,
Yi-Cheng Zhang ^d

^a School of Foreign Languages, University of Electronic Science and Technology of China, Chengdu, 610054, China

^b Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA

^c School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 610054, China

^d Department of Physics, University of Fribourg, Chemin du Musée 3, CH-1700 Fribourg, Switzerland

^e Department of Computer Information Systems and Supply Chain Management, Walker College of Business, Appalachian State University, Boone, NC 28608, USA

H I G H L I G H T S

- Quantitative English words learning strategies based on network and usage are proposed.
 - English words network combining etymology and usage information is constructed and analyzed.
 - The results show the strategies has significant improvement in learning efficiency and effectiveness.
-

Learning English requires a considerable effort, but the way that vocabulary is introduced in textbooks is not optimized for learning efficiency. With the increasing population of English learners, learning process optimization will have significant impact and improvement towards English learning and teaching. The recent developments of big data analysis and complex network science provide additional opportunities to design and further investigate the strategies in English learning. In this paper, quantitative English learning strategies based on word network and word usage information are proposed. The strategies integrate the words frequency with topological structural information. By analyzing the influence of connected learned words, the learning weights for the unlearned words and dynamically updating of the network are studied and analyzed. The results suggest that quantitative strategies significantly improve learning efficiency while maintaining effectiveness. Especially, the optimized-weight-first strategy and segmented strategies outperform other strategies. The results provide opportunities for researchers and practitioners to reconsider the way of English teaching and designing vocabularies quantitatively by balancing the efficiency and learning costs based on the word network.

Keywords:

Word network
Network analysis
Quantitative linguistics
Vocabulary building
Corpus-based linguistic analysis

1. Introduction

English is playing an increasingly important role in bringing social, cultural, and economic influences in non-English speaking countries. English has become omnipresent and dominating in multidimensional aspects of international communications [1]. As the *de facto* global language, English is used by a large body of population [2] and it is learned globally by

* Corresponding author at: School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 610054, China.

E-mail address: tangyong@uestc.edu.cn (Y. Tang).

more than one billion people according to British Council [1]. Required as a compulsory course in China, English is taught at different levels, ranging from elementary school level to graduate school level and is part of National Higher Education Entrance Examination (NCEE). English exam performance is key to education and career opportunities in the future. Given the growing importance, the growing population of English learners, and the tremendous efforts spent on teaching and learning English as a Foreign Language (EFL), it is essential yet challenging to improve English teaching methodologies to keep up with the needs of fast changing circumstances. From the global perspective, improvement and optimization in English learning techniques have significant impacts and benefits. However, improvement of English learning requires mastering a considerable number of English words. In this study, we focus on strategy to enhance English learning process based on the word networks.

English words are the building blocks of English language. There are startling 171 000 English words in the second edition of Oxford English Dictionary [3], and it is still growing as new words are introduced and accepted every day. It is neither possible nor necessary to learn every and each word. Rather, mastering a certain number of English words is needed for daily communication and larger vocabulary is required in complicated contexts. It is believed that about 2000 words are necessary for daily communication. Developing a strong vocabulary leads to better reading comprehension [4] and listening comprehension [5]. It has been pointed out that vocabulary development and instruction are important to master the language [6]. Different exams require different levels of vocabularies. For a typical college student from China, 4500 words is needed to pass the College English Test Level 4 (CET4), 5500 for CET6 [7,8], 3900 for the Test of English as a Foreign Language (TOEFL) [9], and 4500 for the International English Language Testing System (IELTS) [10,11]. Due to the lack of English environment and mainly driven by exam pressures [12], students spend a significant amount of time in memorizing words. Although modern English teaching emphasizes on the actual usage, the situation where students need to grasp a large number of English words in a short time of period hardly changes. While proper vocabularies impact the efficiency and outcome, however, the traditional vocabularies required in different levels of textbooks and exams are composed of words in textbooks. These vocabularies are normally in alphabetical order without considering the inter-relationships between the words. It will be helpful to utilize the optimized vocabulary rather than the traditional ones. Thus, maximizing efficiency and effectiveness while minimizing learning efforts can be achieved in English learning. Being motivated to achieve this specific goal, in this study, we propose a novel learning strategy in English vocabulary building utilizing the usage information and relationships embedded in word networks.

The paper is structured as follows. Firstly, the traditional methods of English vocabulary teaching are introduced in Section 2. Section 3 presents the data used in this study. In Section 4, the research methodologies are provided. In Section 5, the learning strategies are discussed. Finally, Section 6 concludes the study with discussions.

2. Literature review

Though there is a large body of literature offering insights of English vocabulary building from words teaching, there is a lack of focus on the design of particular strategies in English learning, *i.e.* the learning order of words. Some researchers investigated methods for learners to acquire English words using lexical relations of synonymy, antonym, and gradation [13]. However, this study limits to adjective words only. As most previous studies limited to pedagogical methods, this paper focuses on the quantitative learning strategies. The vocabularies designed for early stage beginners normally focus on words for daily communication, usually fewer than 1000 words. For intermediate level students, words are built for advanced tasks like reading and writing where larger vocabulary is needed. The literature on vocabulary teaching mainly discusses the latter phrase of building a vocabulary, which is about 5000 words as required by most of the exams mentioned above. Some literature discusses vocabulary learning from the cognitive aspects of English teaching methodologies [14,15], including contextual guessing, dictionary use, note-taking, contextual encoding, and rote rehearsal. For example, some researchers identify there is a positive relationship between the learning with the outcome of English tests [16] and some discuss a design of pedagogically-sound vocabulary notebooks for classroom use based on principles of language memory and research [17]. Another report showed that the metacognitive strategies are the most frequently used vocabulary learning strategies for learners [18], which is compatible with two other studies [19,20]. Furthermore, it was found that vocabulary size is associated with learning strategies adoptions [19]. Some reports revealed that learners who are more strategic do faster than the less ones in gaining proficiency [20]. The choice and use of strategies are based on the task, learners themselves, and the learning context [21]. Some studies investigated the influential factors of vocabulary acquisition and discuss temporal, spatial, and meaning cues in classroom teaching [22]. A discussion on language learning strategy is given in [23]. Incidental exposures to words in reading are also important for learners to vocabulary acquisition [15,24].

For other languages, two studies propose to learn Spanish words through etymology and mnemonics [25,26]. More relevantly, two prior works focus on the learning strategies for Chinese characters based on the network of characters [27,28]. Chinese characters are normally composed by parts of semantic components (meaning) and phonetic components (pronunciation). Two or more simple and primitive components form the complex characters. These structural relations between simple characters and complex characters can be used to form Chinese character networks where vertices are characters and edges have combination relations of characters. Another interesting research identified that small-world and non-Poisson distributions can be found from the properties of character networks [29]. By utilizing topology information of character networks, another research set weights for vertices according to the usage frequencies and introduce learning costs for characters based on the number of sub-characters and the number of unlearned sub-characters to measure the difficulties in

learning the characters [28]. Generally speaking, a character with more sub-characters and more unlearned sub-characters is more difficult to learn than those with fewer sub-characters and fewer unlearned sub-characters. Thus the learning cost will be higher. The centralities combining frequencies and structure positions derived from the network structure are generalized for all characters. A strategy of learning characters can then be conducted by ranking the characters from the highest to the lowest. The total achieved accumulated usage frequency for a given learning cost is used to evaluate the efficiency of different strategies, *i.e.* learning orders. In this way, a strategy outperforms another if the former achieves a higher accumulated usage frequency than the latter given the same learning cost. In a similar framework, another interesting research further defines the centrality of a character as the ratio of usage frequency to the learning cost [27]. Based on the centralities, all characters are ranked in descending order before performing a topological sorting in which primitive characters are rearranged before the corresponding compound characters. By conducting the rank, an improved strategy for Chinese character learning strategy is achieved. While two studies focused on the simplified Chinese characters [27,28], another research used the information of combining components and character usage frequency to propose a learning mechanism for traditional Chinese characters [30].

These works are based on the character formation structures of Chinese characters. However, belonging to two different language families, alphabetic English words are formed in a different way than the Chinese characters. English words are composed of prefixes, suffixes, and roots. English words can thus form word networks by same components or meaning relationships, like antonyms and synonyms. For example, *previous* and *preview* share the same prefix *pre-*; *enjoy* has an antonym *dislike* and a synonym *appreciate*. From the network analysis perspective, words can be connected with each other to form word networks semantically, orthography, or phonology according to their meanings, syntactic, spelling, or pronunciation [31]. With the development of complex network theory [32,33], there are some study on the network structures of human language [34–39], it was shown that Chinese phrases networks exhibit small world patterns [40]. Some studies observed that the language network built from the co-occurrences of words show the small-world effect and the degree display scale-free distribution [41,42]. Using syllables information, a research built network for Portuguese [43] and another built for Chinese [44], respectively. There are also researches that focus on Romance languages. For example, a research compared the networks of English and Spanish words based on phonological similarities [31]. It is revealed that the English words of islands in the networks are only phonologically similar to each other, while the Spanish words are similar both phonologically and semantically. Some researchers focused on the growth of early semantic networks for children of 30 months [45]. Their findings challenge the preferential attachment process. However, this is usually for adult semantic networks and is not likely to be the model of the early development processes for children. Thus, the learning strategy should not solely rely on the connectivity but also take the semantic and phonological information into considerations. In a similar quantitative linguistics research on the language development for children, they apply network analysis and find that smaller network size leads to large connectivity [46]. A report discussed the semantic networks and the growth [47]. Small-world and scale-free phenomena are revealed in empirical studies on semantic networks. Their results also hint for learning history variables like age and usage frequency. Some researchers showed words of similar concepts form a small-world network with scale-free feature [48]. Unfortunately, the information extracted from the language networks are not used for designing a better words learning strategy. By utilizing the information, an intelligent information system can be built. A research investigated the effect of using etymology in English vocabulary building [49]. They find that experimental group with the treatment of etymology strategy outperform the control group in TOEFL test. An article demonstrated the knowledge of word associations and suffix are related to the vocabulary size and TOEFL test performance [50]. There are some English etymology dictionaries [51,52] and Spanish dictionaries [26]. However, these dictionaries are only organized based on the etymology without considering more information like usage frequency or relations between words and do not indicate any particular word learning orders. In this work, we build English word network. In a wider perspective of the information system, the methods of constructing and studying the lexical network of words are also important in the quantitative analysis of text mining, sentiment analysis, and corpus-based linguistic tools for information systems modeling, and applications of natural language processing [53–55].

3. Data

In this research, we propose an English words learning strategy based on a word–word relationship network N and usage frequency. In the word–word relationship network, a word w_i is connected with its synonyms, antonyms, related words in close usages or contexts, and words that share the same roots. The usage frequency f_i of a word w_i shows how frequently it is used in English. To indicate the degree of difficulty to learn a word, the learning difficulty d_i is adopted from the *difficulty index* of Dictionary.com [56]. The difficulty index is a comprehensive index for each word. General speaking, the larger the value is, the more difficult the word is. However, in a more detailed approach, one can take into considerations of the length, difficulties of unlearned sub-components, the portion of unlearned connected words, and its frequency, etc. To keep it simple, in this research, we adopt the difficulty index of Dictionary.com [56] and leave more detailed approaches in future studies. The learning cost c_i is introduced to describe the dynamic difficulty to learn word w_i by combining the difficulty d_i , the learning situation of connected words. The quantitative definition of learning cost is given later. By combining all these information, a learning order for English words is achieved to minimize total learning cost meanwhile maximizing the total accumulated usage frequencies.

A set of 49 527 English words is analyzed. For each word w_i , the following information is obtained, including synonyms Syn_i , antonyms Ant_i , related words Rel_i , and similar words Sim_i from online dictionaries of Oxforddictionaries.com [3],

Table 1

Basic information of different types of networks. In this table, we summary the number of vertices N_v , number of edges N_e , average degree $\langle d \rangle$, and average clustering coefficient $\langle C \rangle$ for different networks: N_{sim} , N_{ass} , N_{ant} , N_{syn} , N_{rel} , and a random network N_{random} . Among them, the N_{ass} is merged from N_{ant} , N_{syn} , and N_{rel} by keeping only one edge between vertices, i.e. the weights of all edges is set to 1. Every network has 49 527 words, and the numbers of edges in the two main networks N_{sim} and N_{ass} networks are 4 855 275 and 1 874 432, respectively. Compared to the results of words to the random network with the same number of vertices and edges, N_{random} has a very smaller value of $\langle C \rangle$. This indicates that our word networks are relatively denser than random counterpart.

| Network | N_v | N_e | $\langle d \rangle$ | $\langle C \rangle$ |
|--------------|--------|-----------|---------------------|---------------------|
| N_{sim} | 49 527 | 4 855 275 | 196.066 | 0.905 |
| N_{ass} | 49 527 | 1 874 432 | 75.693 | 0.233 |
| N_{ant} | 49 527 | 222 041 | 8.966 | 0.036 |
| N_{syn} | 49 527 | 761 138 | 30.736 | 0.333 |
| N_{rel} | 49 527 | 1 172 254 | 47.338 | 0.228 |
| N_{random} | 49 527 | 4 853 744 | 196.004 | 0.004 |

Dictionary.com [56], Thesaurus.com [57], Merriam–Webster.com [58], and Webster-dictionary.org [59]. The usage frequency f_i is taken from the 1-grams word list of Google’s Web 1T 5-gram Version 1 project [60] in which the observed frequency and ranking of each word are calculated from publicly accessible web pages with over one trillion n-grams words. For a pair of words w_i and w_j , if they share the same component, then w_i and w_j are structurally similar to each other. We then obtain the similar words Sim_i of w_i from the component information on Dictionary.com [56]. To focus on the research question, we treat the affixes such as *-ion*, *-ly*, *-ing*, etc. as word components like other Latin and Greek roots such as *ab-*, *dic-*, *ex-*, etc. Furthermore, for the data availability, we treat singular (without *-s*) and plural form (with *-s*) as different words. For example, *turn* and *turns* are treated as two different words with different usage frequency. In Dictionary.com [56], we can get the component information. However, the words in the form of third-person singular are treated the same to the original form, so there is no component of *-s*. Table 1 shows the basic information of the networks: similar words network N_{sim} , associated words network N_{ass} , antonyms network N_{ant} , synonyms network N_{syn} , and related words network N_{rel} . As it is shown, there are millions of edges in these networks, indicating rich information embedded among the English words. The average degree $\langle d \rangle$ provides a quick impression how the words are connected. As illustrated in Fig. 1, we plotted the topological structures of similar words network N_{sim} , associated words network N_{ass} and the subsets of N_{ass} : antonyms network N_{ant} , synonyms network N_{syn} , and related words network N_{rel} . The vertex size is proportional to the degree and the color depth is proportional to the frequency. In each network, we also label some selected common words with highest degree and frequency. As we can see, those words are the most used daily words.

On average, each word has dozens of directly connected words. Thus the learning of a word can influence a large number of neighbor words. The average clustering coefficient $\langle C \rangle$ is an indicator of local connectivity. For the similar word network N_{sim} , $\langle C \rangle = 0.905$ is significantly higher than the rest of networks and values of all networks are not small. This also suggests that the word networks are not poorly connected, in other words, the connectivities contain rich information of relationships between English words.

It is also interesting to see the frequency distribution of the English words. In Fig. 2, the rank and frequency in log–log scale for words used in this study is plotted. The words follow Zipf’s law, which is widely observed in language studies [61], with an exponent of $\alpha = -1.5691$. This indicates that the sample word dataset is a typical set of words without changing the distribution.

4. Conceptions and properties

4.1. Associated and similar words

There are two kinds of words connected to a given English word w_i , either being associated with meanings or structurally similar. The antonym words Ant_i are those with opposite meaning to w_i , the synonym words Syn_i are those with exactly or nearly the same meaning as w_i , the related words Rel_i are those with connected meanings to w_i in usages. All three kinds of words form the joint set of w_i , the associated words Ass_i . Those words share certain components, like *root*, *prefix*, or *suffix* with w_i are structurally similar words, denoted as Sim_i . A vocabulary of 49 527 words from online dictionaries [3,56–58] was collected and the associated words Ass_i and structurally similar words Sim_i for each word w_i were obtained.

4.2. Word network

Based on the analysis, the synonyms, antonyms, and related words are connected in meanings. They may look significantly different, for example, *perfect* is a synonym word to *great*, but the two words ‘look’ differently. However, even those structurally similar words share same components and ‘look’ like each other, they might not necessarily be connected in meanings. For example, *station* and *observation* share the same component ‘*-ation*’. However, they have different meanings. To construct the word–word relationship network N , we first look up all synonyms, antonyms, related words, and words

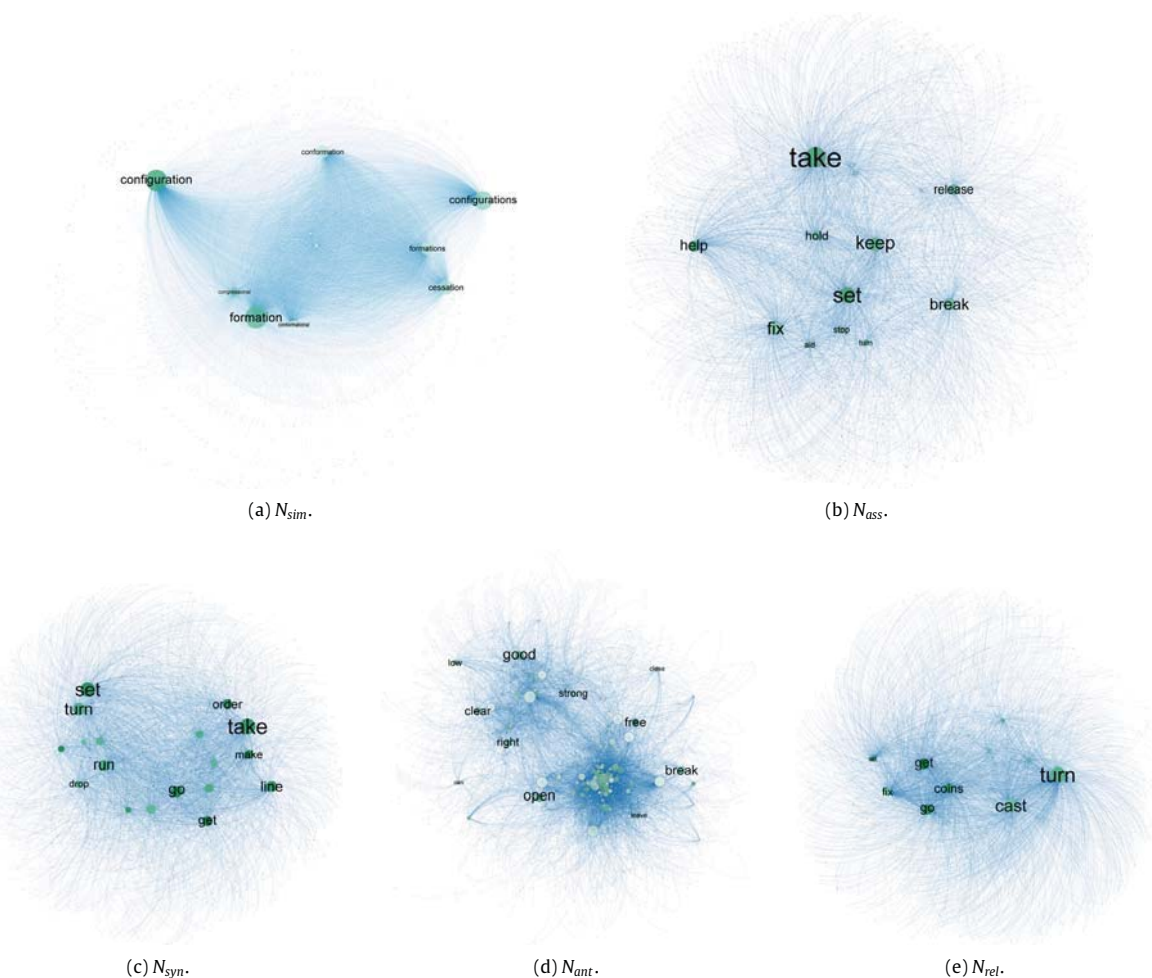


Fig. 1. Topological structures of all five English word networks. In this figure, we virtualize the word networks of 49 527 in our English word dataset. For simpler and better visualization, only top 5% words with highest degrees and their edges are drawn in (a) Similar network N_{sim} , (b) Associated network N_{ass} , (c) Synonym network N_{syn} , (d) Antonym network N_{ant} , and (e) Related network N_{rel} .

sharing same components for each word from online dictionaries [3,56–58]. In Fig. 3, we show the combined network of both Ass_i and Sim_i for all 49 527 words. For better visualization, the top 5% words are shown in the figure.

Fig. 4 provides a detailed example of how words are connected. Take the word *advocate* as example. *admit* and *advise* share the component *ad-* with *advocate*, *provoke*, *vocal* and *invoke* share the transformed component *voc*, *vok* of *voice* with *advocate*, and *rotate* and *operate* share the component *-ate* with *advocate*. These words form the structurally similar words $Sim_{advocate}$ of *advocate*. The associated words consist of synonyms, antonyms and related words. In this case, *favor* and *promote* are synonyms of *advocate*, *prevent* is the antonym, *help* and *guard* are the related words. The related words Rel_i are collected from Dictionary.com [56] for those words that are not synonyms or antonyms, but extensively connected in language contexts or domains. The relationships of structural similar and meanings associated are summarized in Table 2. The basic information of the words in the example is also listed in Table 3.

4.3. Frequency and degree

The importance of a word depends on two aspects: first is static, including the use frequency and the degree of the word on the network of itself. The other is dynamic learning cost which changes when its connected neighbor words are learned. The frequency f_i of word w_i is normally a count of appearances in a corpus, it directly indicates the importance of the word in the use of a language. The higher the frequency f_i of the word w_i has, the more important the word w_i is. If we only consider the words in an isolated manner, a learning order of words can be straightly obtained based on the frequencies. The frequency of a word is an important attribute. However, words are mutually connected among each other rather than isolated individuals. A higher frequency does not necessarily indicate that the word should be learned first. For example,

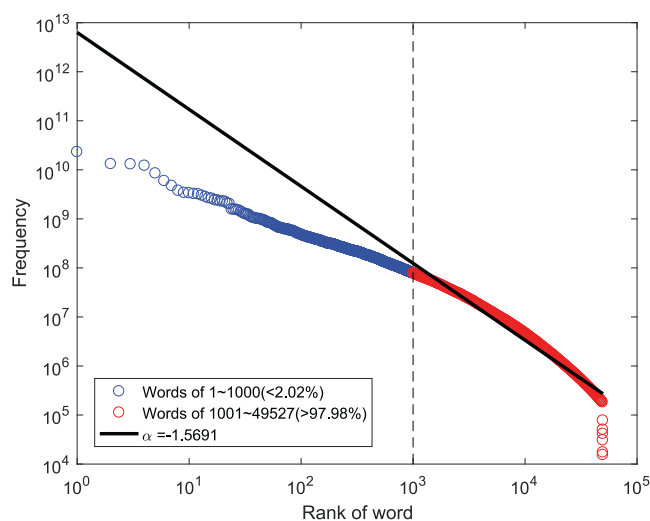


Fig. 2. Log-log scaled frequency and rank. We calculate the logged frequency and rank of the 45,927 words in our dataset. It follows a Zipf's law with an exponent of $\alpha = -1.5691$. We had shown all of the 49 527 words as circles in the figure. Since the x-axis is increasing by exponential growth, the words on right side (red ones), which means their ranks are between 1,000 and 49 527, take more than 97.98% of all the words. The left part of the plot only has less than 2.02% words. So the slope of the regression line is largely determined by the words on the right side, and the regression line looks like fitting to the right part of the plot.

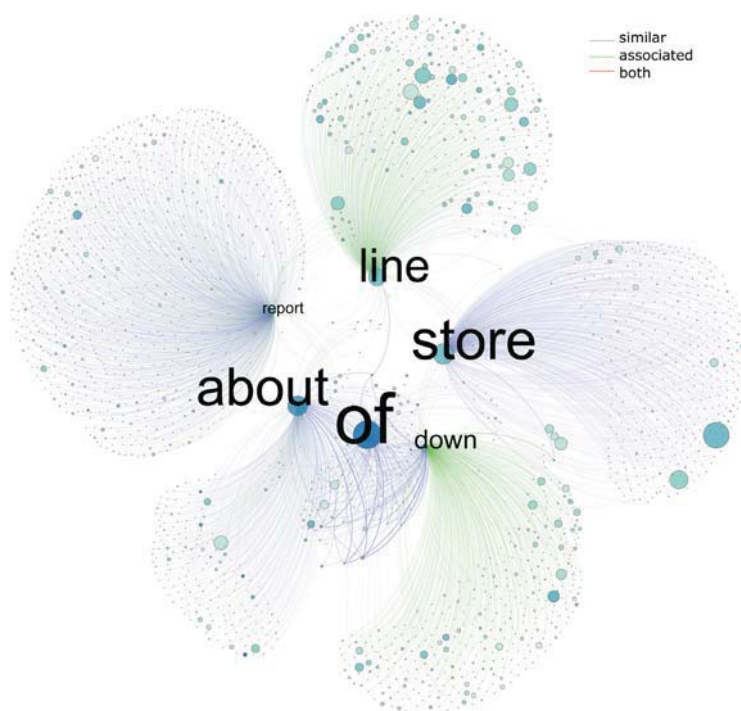


Fig. 3. Topological structure of the total combined network. We combine the similar word network N_{sim} with the associated word network N_{ass} into a total network $N_{total} = N_{sim} \cup N_{ass}$. Thus, the total network N_{total} carry inter-relationship information of both structural similarity from N_{sim} and the meaning associations from N_{ass} . Since N_{total} has millions of edges, the top 5% words with highest degrees were selected. The vertex size is proportional to the degree, while the color depth is proportional to the frequency. The network is drawn using Gephi package [62] and the layout is manually adjusted for better visualization.

given two words w_i and w_j , if w_i has a slightly higher frequency than w_j , i.e. $f_i > f_j$ but w_i has far fewer connected neighbors, i.e. $|Sim_i| + |Ass_i| < |Sim_j| + |Ass_j|$, then it will be better to learn w_j first, since w_j has larger influence over the network than w_i . By learning w_j , it benefits other words belonging to Sim_j and Ass_j . Considering the influence of learning some words in its structurally similar words and meaning associated words might bring upon the word. This topological information

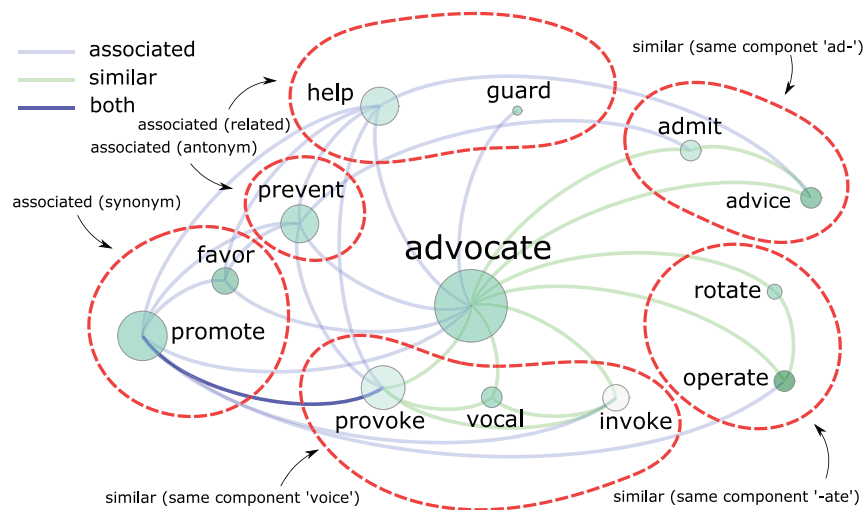


Fig. 4. Sample network of word *advocate*. In this sample network, how the total network N_{total} is formed using synonyms, antonyms, related words, and similar words is investigated. As it is shown, the word *advocate* has 7 similar words, sharing same components 'ad-', '-ate', and 'voice'; 5 associated words including 2 synonyms, 1 antonym, and 2 related words. Together, these 7 similar words and 5 associated words are directly connected with *advocate*. The vertex size is proportional to the degree and the color depth is proportional to the frequency. The edges are colored for associated in blue, similar in green, and especially, if two words are both associated and similar, the edge is colored in gray, e.g. *promote* and *provoke* share component 'pro-' and are also related words in meaning.

Table 2

Relations of sample network of word *advocate* in Fig. 4. Similar words are words that share one or more components and these structurally similar words are marked as S. For those words associated with words in meanings are marked as A. It is denoted $A[s]$ for synonyms, $A[a]$ for antonyms, and $A[r]$ for related words. For example, *promote* and *provoke* are both structurally similar and sharing the same component 'pro-'. Also, they are associated with meanings as synonyms and related words, marked as $S A[sr]$. As the example shown in Fig. 4, considering all pairs in our English words, both structural similarity and meaning associated information are combined to build the network.

| | Advocate | Admit | Advice | Provoke | Vocal | Invoke | Operate | Rotate | Promote | Favor | Prevent | Help | Guard |
|----------|----------|-------|--------|----------|-------|--------|---------|--------|----------|--------|----------|----------|-------|
| Advocate | | S | S | S | S | S | S | S | A[s r] | A[s r] | A[a] | A[r s] | A[r] |
| Admit | S | | S | | | | | | | | A[a] | | |
| Advice | S | S | | | | | | | | | | | |
| Provoke | S | | | | S | S | | | S A[s r] | | A[a] | A[s r] | |
| Vocal | S | | | S | | S | | | | | | A[a] | |
| Invoke | S | | | S | S | | | | A[r] | | | | |
| Operate | S | | | | | | | S | A[s r] | | | | |
| Rotate | S | | | | | | S | | | | | | |
| Promote | A[s r] | | | S A[s r] | | A[r] | A[s r] | | | A[s r] | A[a] | A[s r] | |
| Favor | A[s r] | | | | | | | | | | A[a] | A[s r] | |
| Prevent | A[a] | A[a] | | A[a] | | | | | A[s r] | A[a] | | A[s a r] | |
| Help | A[r s] | | A[s r] | A[a] | | | | | A[s r] | A[s r] | A[s a r] | | |
| Guard | A[r] | | | | | | | | | | | | |

should be utilized with the frequency information. There is a point of balance and trade-off needed between frequency and topological importance.

4.4. Learning cost

To measure how difficult it is to learn a word, the difficulty index is adopted from Dictionary.com [56] as the initial learning cost c_i for word w_i . A higher learning cost indicates more efforts are required to learn the word. The learning cost c_i is constantly changing and updated in the process of vocabulary building process with more and more words are learned. The learning cost c_i of word w_i will decline when more and more similar words and associated words are learned by the learner, as the learner can benefit from previous learning of those connected words. The learning cost c_i is defined in Eq. (1) as

$$c_i = d_i \left(1 - \frac{\Delta f_i^{Ass}}{f_i^{Ass} + f_i} \right) \left(1 - \frac{\Delta f_i^{Sim}}{f_i^{Sim} + f_i} \right) \quad (1)$$

where c_i refers to the learning cost of w_i , d_i is the normalized difficulty index [56] as the initial learning difficulty, f_i^{Ass} and f_i^{Sim} are the total frequency of the associated words Ass_i and similar words Sim_i , respectively. Δf_i^{Ass} and Δf_i^{Sim} are the

Table 3

Basic information of words in the sample network of word *advocate* in Fig. 4. In this table, the frequency f_i , the degree d_i , the number of similar words $|Sim_i| + 1$, the number of associated words $|Ass_i| + 1$, and the weight W_i for each word is demonstrated.

| w_i | f_i | d_i | $ Sim_i + 1$ | $ Ass_i + 1$ | W_i |
|----------|--------|--------|---------------|---------------|--------|
| Advocate | 0.0127 | 0.4286 | 8 | 6 | 0.0095 |
| Admit | 0.0144 | 0.6286 | 3 | 2 | 0.0014 |
| Advice | 0.1166 | 0.3143 | 3 | 2 | 0.0109 |
| Provoke | 0.0016 | 0.8000 | 5 | 4 | 0.0005 |
| Vocal | 0.0149 | 0.3714 | 4 | 1 | 0.0009 |
| Invoke | 0.0038 | 1.0000 | 4 | 2 | 0.0005 |
| Operate | 0.0299 | 0.2857 | 3 | 2 | 0.0028 |
| Rotate | 0.0047 | 0.2000 | 3 | 1 | 0.0002 |
| Promote | 0.0542 | 0.4286 | 2 | 8 | 0.0135 |
| Favor | 0.0194 | 0.1143 | 1 | 5 | 0.0015 |
| Prevent | 0.0610 | 0.4571 | 1 | 7 | 0.0067 |
| Help | 1.0000 | 0.0857 | 1 | 7 | 0.1094 |
| Guard | 0.0346 | 0.1714 | 1 | 2 | 0.0011 |

accumulated learned frequency of associated words and similar words, respectively. From this definition, the learning cost c_i is proportional to the initial learning difficulty d_i as well as to the accumulated learned frequencies Δf_i^{Ass} and Δf_i^{Sim} . It is inversely proportional to the frequency f_i itself as well as to the total frequencies f_i^{Ass} and f_i^{Sim} .

4.5. Topological weight

As mentioned previously, the frequency is not the only standard to measure the importance of a word. The topological weight or importance should be considered as well. The word w_i is connected with its similar words Sim_i and associated words Ass_i on the network. The number of words in Sim_i and Ass_i shows the degree w_i has. From the perspective of network, the higher the degree is, the more importance the vertex has. In this way, the topological weight W_i for word w_i is defined as:

$$W_i = \frac{(|Sim_i| + 1)(|Ass_i| + 1)f_i}{(|Sim_i|_{max} + 1)(|Ass_i|_{max} + 1)f_{max}} \quad (2)$$

where $|Sim_i|$ and $|Ass_i|$ are the numbers of similar and associated words, respectively. $|Sim_i|_{max}$ and $|Ass_i|_{max}$ stands for the maximum numbers among all words. This equation considers the normalized contribution of both degree and frequency. Given that Sim_i or Ass_i might be 0 for isolated words, without loss of generality, $|Sim_i| + 1$ and $|Ass_i| + 1$ is utilized instead.

5. Learning strategies and results

5.1. Weighted strategies

Once the learning cost c_i and weight W_i is defined, the word learning strategies are presented. A strategy is an order of words for a learner to learn words one by one. For example, for the n words $\{w_1, w_2, w_3, \dots, w_n\}$, a naive strategy is just the order without changes from w_1, w_2 , etc., till w_n . The challenge is to enrich accumulated frequency as fast as possible while keeping the learning costs as low as possible. Different learning strategies have different performances. It is suggested that frequency plays a significant role in learning words. By ranking all words in descending order of frequency, a simple strategy without using any other topological information is conducted. It is defined as naive strategy *maximum frequency first strategy* denoted as $S^{frequency}$. Considering the learning cost as defined in Eq. (1), a new strategy combined with both frequency and topological information can be concluded. It is defined as *minimum learning cost first strategy* denoted as S^{cost} . Unfortunately, many existing vocabularies do not have specific strategies, those words appear in randomness only according the used texts or articles. Besides, 'random' is an extreme situation and the random strategy is included to represent a strategy where no strategies applied at all. So, we introduce a *random strategy* denoted as S^{random} . Considering many vocabularies only list words in an alphabetical order from A to Z. From what we observed in schools from high schools to universities and even our own personal experiences, at least in China, it is not rare but surprisingly common for students to spend great time in memorizing mechanically words list in alphabetical, since almost all available word lists are simply assembled in alphabetical order from A to Z. Accordingly, we generate a *dictionary strategy* by listing words in dictionary order, we denote this strategy as $S^{dictionary}$. If we simply rank the words by weight W_i defined in Eq. (2), we get a *weight first strategy* denoted as S^{weight} . By combining the weight and learning cost together, we can revise the weight in Eq. (2) as

$$W_i^* = \frac{W_i}{c_i} \quad (3)$$

This indicates that a word with lower learning cost and larger weight is more likely to be learned firstly than those with higher cost and smaller weight. We call the strategy based on W_i^* *optimized weight first strategy* denoted as S^{weight^*} . An

Table 4

Learning procedure of sample words in Fig. 4. According to the optimized weight first strategy S^{weight^*} , the learning order for these sample words is *help, advice, advocate, promote, favor, prevent, provoke, admit, operate, invoke, rotate, guard, and vocal*.

| Round | Advocate | Admit | Advice | Provoke | Vocal | Invoke | Operate | Rotate | Promote | Favor | Prevent | Help | Guard |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Round 1 c_i | 0.4286 | 0.6286 | 0.3143 | 0.8000 | 0.3714 | 1.0000 | 0.2857 | 0.2000 | 0.4286 | 0.1143 | 0.4571 | 0.0857 | 0.1714 |
| W_i^* | 0.0222 | 0.0022 | 0.0348 | 0.0006 | 0.0025 | 0.0005 | 0.0098 | 0.0011 | 0.0316 | 0.0133 | 0.0146 | 1.2762 | 0.0063 |
| Round 2 c_i | 0.0659 | 0.6286 | 0.0328 | 0.0836 | 0.3714 | 1.0000 | 0.2857 | 0.2000 | 0.0662 | 0.0147 | 0.0641 | - | 0.1714 |
| W_i^* | 0.1445 | 0.0022 | 0.3330 | 0.0059 | 0.0025 | 0.0005 | 0.0098 | 0.0011 | 0.2049 | 0.1036 | 0.1040 | - | 0.0063 |
| Round 3 c_i | 0.0272 | 0.1184 | - | 0.0836 | 0.3714 | 1.0000 | 0.2857 | 0.2000 | 0.0662 | 0.0147 | 0.0641 | - | 0.1714 |
| W_i^* | 0.3504 | 0.0114 | - | 0.0059 | 0.0025 | 0.0005 | 0.0098 | 0.0011 | 0.2049 | 0.1036 | 0.1040 | - | 0.0063 |
| Round 4 c_i | - | 0.0629 | - | 0.0714 | 0.2281 | 0.6140 | 0.2089 | 0.1462 | 0.0615 | 0.0134 | 0.0592 | - | 0.1254 |
| W_i^* | - | 0.0215 | - | 0.0069 | 0.0041 | 0.0008 | 0.0134 | 0.0015 | 0.2202 | 0.1134 | 0.1128 | - | 0.0086 |
| Round 5 c_i | - | 0.0629 | - | 0.0104 | 0.2281 | 0.0401 | 0.0742 | 0.1462 | - | 0.0080 | 0.0379 | - | 0.1254 |
| W_i^* | - | 0.0215 | - | 0.0471 | 0.0041 | 0.0117 | 0.0377 | 0.0015 | - | 0.1898 | 0.1762 | - | 0.0086 |
| Round 6 c_i | - | 0.0629 | - | 0.0104 | 0.2281 | 0.0401 | 0.0742 | 0.1462 | - | - | 0.0302 | - | 0.1254 |
| W_i^* | - | 0.0215 | - | 0.0471 | 0.0041 | 0.0117 | 0.0377 | 0.0015 | - | - | 0.2207 | - | 0.0086 |
| Round 7 c_i | - | 0.0120 | - | 0.0003 | 0.2281 | 0.0401 | 0.0742 | 0.1462 | - | - | - | - | 0.1254 |
| W_i^* | - | 0.1126 | - | 1.8771 | 0.0041 | 0.0117 | 0.0377 | 0.0015 | - | - | - | - | 0.0086 |
| Round 8 c_i | - | 0.0120 | - | - | 0.2104 | 0.0370 | 0.0742 | 0.1462 | - | - | - | - | 0.1254 |
| W_i^* | - | 0.1126 | - | - | 0.0044 | 0.0127 | 0.0377 | 0.0015 | - | - | - | - | 0.0086 |
| Round 9 c_i | - | - | - | - | 0.2104 | 0.0370 | 0.0742 | 0.1462 | - | - | - | - | 0.1254 |
| W_i^* | - | - | - | - | 0.0044 | 0.0127 | 0.0377 | 0.0015 | - | - | - | - | 0.0086 |
| Round 10 c_i | - | - | - | - | 0.2104 | 0.0370 | - | 0.0198 | - | - | - | - | 0.1254 |
| W_i^* | - | - | - | - | 0.0044 | 0.0127 | - | 0.0111 | - | - | - | - | 0.0086 |
| Round 11 c_i | - | - | - | - | 0.1676 | - | - | 0.0198 | - | - | - | - | 0.1254 |
| W_i^* | - | - | - | - | 0.0056 | - | - | 0.0111 | - | - | - | - | 0.0086 |
| Round 12 c_i | - | - | - | - | 0.1676 | - | - | - | - | - | - | - | 0.1254 |
| W_i^* | - | - | - | - | 0.0056 | - | - | - | - | - | - | - | 0.0086 |
| Round 13 c_i | - | - | - | - | 0.1676 | - | - | - | - | - | - | - | - |
| W_i^* | - | - | - | - | 0.0056 | - | - | - | - | - | - | - | - |

intelligent information system is thus constructed based on the strategy. By utilizing the intelligent information system, it is expected that the learning effectiveness and efficiency will increase significantly in the future.

To demonstrate how the learning order of considered words is determined in the learning process for the learning strategy S^{weight^*} , we take the *advocate* and its connected words as sample. As shown in Fig. 4 and Tables 2 and 3, the sample network contains of 13 words. Each step is presented in Table 4. In the first round, we update the learning costs and optimized weights using Eqs. (1) and (3), respectively. We choose the word *help* as the first to learn, which has the largest weight. In the second round, we update the learning costs of *advocate, advice, promote, favor* and *prevent*, which have associated relationships with *help*. Since no words are structurally similar to *help*, no learning costs of similar words need to be updated. Then, we update weights of *advocate, advice, promote, favor*, and *prevent* according to Eq. (3). From Table 3, we can see that *invoke* has the largest difficulty index d_i as the initial learning cost c_i . In the process of learning on the whole word network, the learning cost is reduced when *advocate* and *provoke* are learned as similar words and *promote* is learned as associated word. By utilizing the proposed learning strategy, after these words are learned, the c_i of *invoke* is reduced and thus makes its W_i^* increased. Similarly, we repeat this updating process to find the word for each round till the last word *vocal*. In conclusion, using the proposed strategy, the word with largest optimized weight value in each round and dynamically updated the weight values of unlearned words.

To compare the efficiency of different strategies, the growths of accumulated frequency $\sum f_i$, accumulated weight $\sum W_i$, and accumulated number of learned words $|w_{learned}|$ against the accumulated learning cost $\sum c_i$ are analyzed. In the process of learning, based on the order defined in a learning strategy, when new words are learned with some sacrifice of overcoming the learning costs, they also contribute in bringing more learned frequency, weight, and number of words. A good strategy is thought to gain high accumulated frequency $\sum f_i$, accumulated weight $\sum W_i$, and accumulated number of learned words $|w_{learned}|$ with least accumulated learning cost $\sum c_i$. In other words, for a given $\sum c_i$, if a strategy S_a outperforms another strategy S_b in the manner of the three learning outcomes, then S_a is more efficient compared to S_b . In Fig. 5, for the six strategies, we plot the accumulated learning cost $\sum c_i$ against the accumulated frequency $\sum f_i$ in Fig. 5(a), accumulated weight $\sum W_i$ in Fig. 5(b), and number of learned words $|w_{learned}|$ in Fig. 5(c), respectively. As shown in the figures, it is suggested that the *dictionary* strategy $S^{dictionary}$ and the *random* strategy S^{random} have the worst performance in all three cases. This demonstrates how inefficient and costly for learners to memorize words in rote using the widely used word vocabularies from A to Z in textbooks or just randomly. All other strategies demonstrate significantly improved performances compared to $S^{dictionary}$ and S^{random} . Specifically, in Fig. 5(a), we represent the accumulated Learning Cost $\sum c_i$ - Accumulated Frequency $\sum f_i$. Point A in Fig. 5(a): $\sum c_i$ reaches at 226.69 out of 3533.18 and the $\sum f_i$ reaches at 0.84, S^{weight} and S^{weight^*} covers 4534 and 22 534 words, respectively. Fig. 5(b) shows the Accumulated Learning Cost $\sum c_i$ against Accumulated Weight $\sum W_i$. Point B in Fig. 5(b): when accumulated weight of S^{weight^*} reaches 99%, the learning cost is only 180.72 out of 3533.18. In Fig. 5(c), it represents the relationship between Accumulated Learning Cost $\sum c_i$ against Accumulated Number of Learned Words $|w_{learned}|$. Point C in Fig. 5(c): when the accumulated word number $|w_{learned}|$ reaches 42 000 out of total 49 527 words, the learning cost of S^{weight^*} is only 505.94 out of 3533.18.

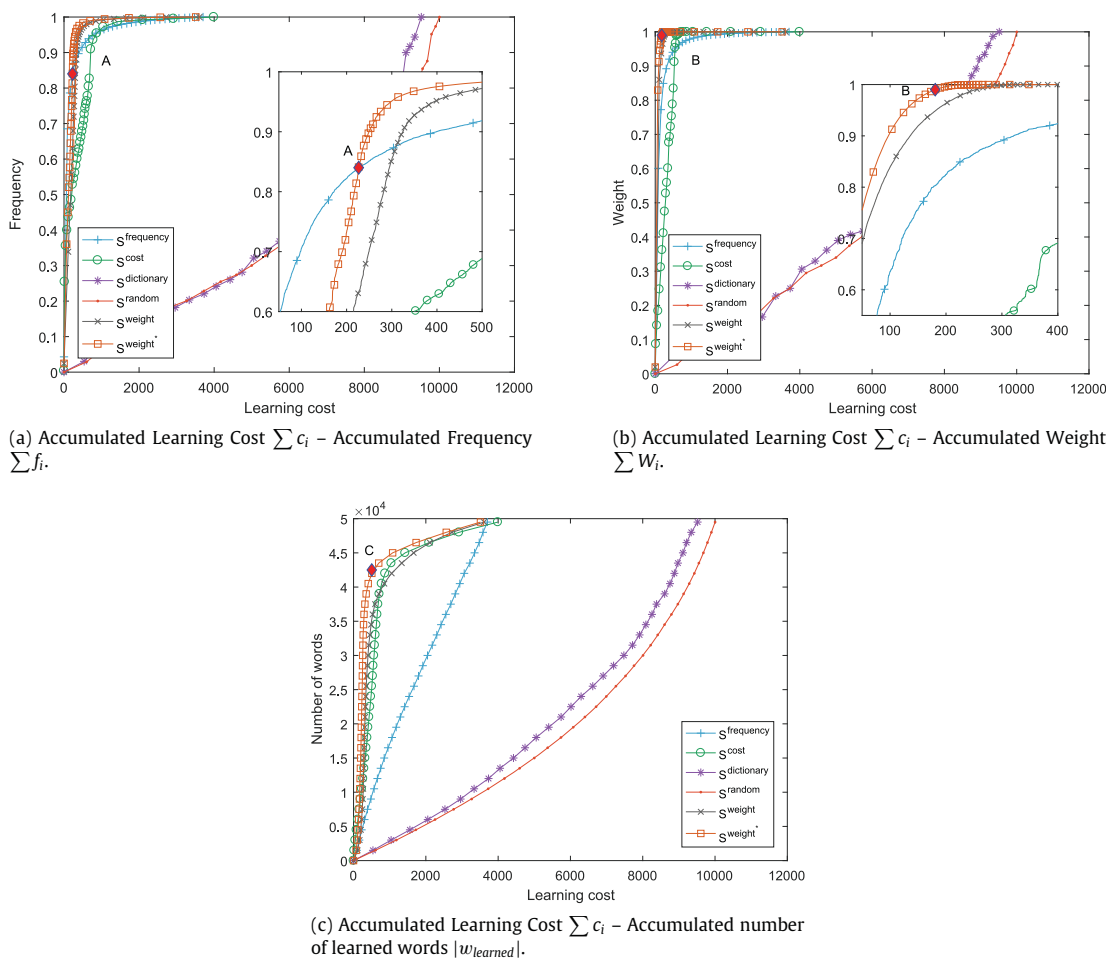


Fig. 5. Comparison of weighted strategies.

The frequency-based strategy $S^{frequency}$ performs poorly compared to the other three strategies of S^{cost} , S^{weight} , and S^{weight^*} in accumulated number of learned words $|w_{learned}|$ as shown in Fig. 5(c). This suggests that the naive strategy, which is purely based on word frequency might not be a good option. For a given accumulated learning cost $\sum c_i$, a much smaller number of words are learned in $S^{frequency}$ which is very inefficient. For the three strategies of S^{cost} , S^{weight} , and S^{weight^*} , the curves show steep slopes and climb very fast to reach the maximum value. This shows these strategies are very efficient in reaching high accumulated frequency $\sum f_i$ and accumulated weight $\sum W_i$ as shown in Figs. 5(a) and (b), respectively. In Fig. 5(a), the frequency strategy has the best performance in initial stage but eventually loses to weight first strategy S^{weight} and optimized weight first strategy S^{weight^*} . Among all strategies, the optimized weight strategy S^{weight^*} has the best performance in all three results, as shown in Fig. 5(a)–(c). Based on details of how strategies perform in the learning processes, we present the results in Tables 6–8 for the accumulated learning costs for different percentages of $\sum f_i$, $\sum W_i$, and $|w_{learned}|$, respectively. The optimized weight strategy S^{weight^*} significantly outperforms the rest of the strategies. As a glimpse of the result, In Table 5, top 20 words to be learned first out of the total 49 527 words is listed.

5.2. Segmented strategies

Since it is more likely to learn the most frequent words first in real life, we have designed several strategies and made comparisons to gain an idea on the efficiency of them.

We added two kinds of new strategies: (1) Learn in Intervals: Segmented S^{weight^*} . We first rank all words in order of frequency. Then we divide the words into five intervals: 1–10, 11–100, 101–1000, 1001–10 000, and the rest words. Intervals are learned one by one. In each interval, we adopt the S^{weight^*} strategy. For those words in the same interval, words with the larger values of optimized weights will be learned first. (2) Learn top words according to frequency first. In this approach, three strategies are considered: (a) S^{weight^*} from 2.5%: top 2.5% (1238 words) are learned according to frequency first, then

Table 5

Top 20 words according to optimized weight strategy $S^{weight*}$. We list the learning rank r_i , the frequency f_i , the number of similar words $|Sim_i| + 1$, the number of associated words $|Ass_i| + 1$, the initial learning difficulty index d_i , the learning cost when it is learned c_i , the weight W_i and, the optimized weight W_i^* for these top 20 words. We see that some words have smaller values of learning costs than the initial difficulty, this indicates their costs are getting smaller in the process of learning.

| r_i | w_i | f_i^c | $ Sim_i $ | $ Ass_i $ | d_i^c | c_i^c | W_i^b | W_i^{*a} |
|-------|---------|---------|-----------|-----------|---------|---------|---------|------------|
| 1 | of | 56.9 | 266 | 81 | 3.41 | 3.41 | 10.0 | 29.3 |
| 2 | about | 5.30 | 369 | 142 | 2.27 | 0.42 | 2.25 | 54.3 |
| 3 | down | 0.97 | 288 | 804 | 2.27 | 0.37 | 1.82 | 49.0 |
| 4 | line | 1.21 | 119 | 1267 | 1.14 | 1.14 | 1.48 | 13.0 |
| 5 | report | 1.24 | 699 | 777 | 4.55 | 4.49 | 5.41 | 12.1 |
| 6 | store | 1.26 | 502 | 506 | 2.27 | 2.27 | 2.58 | 11.4 |
| 7 | for | 25.7 | 41 | 138 | 1.14 | 1.14 | 1.20 | 10.6 |
| 8 | used | 1.82 | 709 | 112 | 1.14 | 1.14 | 1.17 | 10.3 |
| 9 | away | 0.49 | 335 | 227 | 2.27 | 0.33 | 0.30 | 9.04 |
| 10 | being | 1.05 | 1032 | 308 | 4.55 | 3.09 | 2.69 | 8.71 |
| 11 | address | 1.13 | 261 | 499 | 13.6 | 1.67 | 1.19 | 7.16 |
| 12 | around | 0.77 | 263 | 124 | 3.41 | 0.32 | 0.21 | 6.37 |
| 13 | action | 0.70 | 1191 | 602 | 7.96 | 7.75 | 4.04 | 5.22 |
| 14 | section | 1.00 | 1169 | 500 | 9.09 | 8.96 | 4.73 | 5.28 |
| 15 | amount | 0.40 | 267 | 350 | 5.68 | 0.58 | 0.30 | 5.21 |
| 16 | real | 1.29 | 848 | 396 | 6.82 | 6.82 | 3.48 | 5.11 |
| 17 | return | 0.89 | 679 | 718 | 7.96 | 7.32 | 3.49 | 4.77 |
| 18 | review | 1.47 | 671 | 690 | 12.5 | 11.1 | 5.47 | 4.93 |
| 19 | over | 1.99 | 88 | 353 | 1.14 | 1.09 | 0.50 | 4.61 |
| 20 | open | 1.04 | 63 | 1569 | 2.27 | 1.56 | 0.84 | 5.37 |

^a 10^0

^b 10^{-1}

^c 10^{-2}

Table 6

Accumulated learning costs $\sum c_i$ of learning strategies for different accumulated frequency percentages. In the process of learning words, with more new words are learned, the accumulated frequency increase, we record the accumulated learning costs required for each strategy to achieve every 10% of total frequency. Before 40% frequency, learning cost of $S^{weight*}$ is slightly higher than the frequency first strategy $S^{frequency}$. After reaching 80% frequency, $S^{weight*}$ turns to be the best performer using least accumulated learning costs even less than the frequency based strategy $S^{frequency}$. It is worth noting that only top 10% words with highest frequencies contribute to over 80% of total frequency, this means $S^{frequency}$ only perform well for the small fraction of top words but fail to $S^{weight*}$ for the rest 90% words. Overall, $S^{weight*}$ is a more better choice.

| $\sum f_i$ | $S^{frequency}$ | S^{cost} | $S^{dictionary}$ | S^{random} | S^{weight} | $S^{weight*}$ |
|------------|-----------------|------------|------------------|--------------|--------------|---------------|
| 10% | 0.12 | 2.45 | 1351.75 | 1571.99 | 16.79 | 5.38 |
| 20% | 0.50 | 7.84 | 3274.87 | 3240.54 | 44.87 | 13.05 |
| 30% | 2.22 | 31.98 | 4901.69 | 4992.76 | 89.55 | 45.23 |
| 40% | 11.07 | 80.39 | 5960.49 | 6059.52 | 147.99 | 87.7 |
| 50% | 25.96 | 215.71 | 7050.07 | 7215.94 | 181.43 | 114.46 |
| 60% | 53.12 | 348.72 | 7663.57 | 8136.98 | 215.35 | 161.08 |
| 70% | 97.44 | 520.4 | 8374.67 | 8838.39 | 249.58 | 190.4 |
| 80% | 172.22 | 655.49 | 9043.68 | 9382.06 | 282.03 | 217.22 |
| 90% | 399.05 | 694.78 | 9113.82 | 9744.21 | 324.44 | 249.64 |
| 100% | 3705.19 | 4016.91 | 9522.49 | 10004.18 | 3624.33 | 3533.18 |

Table 7

Accumulated learning costs $\sum c_i$ of learning strategies for different accumulated weight percentages. It is easy to see that $S^{weight*}$ is significantly better than any other strategies in the whole learning process.

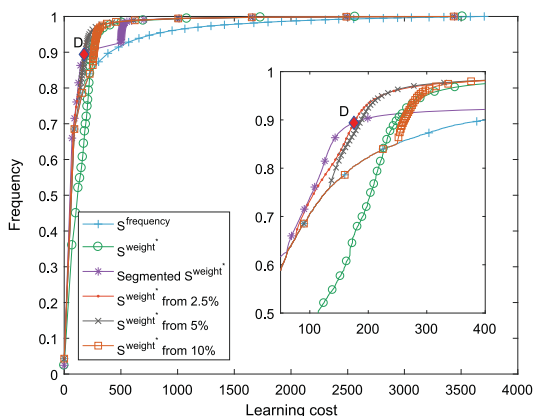
| $\sum W_i$ | $S^{frequency}$ | S^{cost} | $S^{dictionary}$ | S^{random} | S^{weight} | $S^{weight*}$ |
|------------|-----------------|------------|------------------|--------------|--------------|---------------|
| 10% | 11.89 | 24.77 | 1425.89 | 1602.94 | 1.12 | 0.62 |
| 20% | 17.98 | 91.17 | 3118.37 | 2975.63 | 3.81 | 2.39 |
| 30% | 29.11 | 148.59 | 3974.38 | 4292.05 | 7.53 | 5.21 |
| 40% | 48.59 | 213.88 | 5271.71 | 5696.7 | 12.62 | 9.48 |
| 50% | 62.79 | 279.89 | 6658.78 | 6728.99 | 21.24 | 15.49 |
| 60% | 89.84 | 349.39 | 7650.05 | 7715.69 | 32.98 | 24.8 |
| 70% | 121.15 | 415.97 | 8192.36 | 8396.3 | 52.58 | 39.24 |
| 80% | 176.57 | 489.79 | 8376.86 | 9077.41 | 83.34 | 61.27 |
| 90% | 321.07 | 530.66 | 8906.02 | 9667.63 | 135.95 | 96.77 |
| 100% | 3705.19 | 4016.91 | 9522.49 | 10004.18 | 3624.33 | 3533.18 |

the rest 97% are learned using optimized weight first strategy, similarly, we also include another two cases (b) $S^{weight*}$ from 5% and (c) $S^{weight*}$ from 10%, in which 5% (2476 words) and 10% (4952 words) are learned according to frequency then by the optimized weight first strategy. We carried calculations using both two new strategies. To make the new strategies more

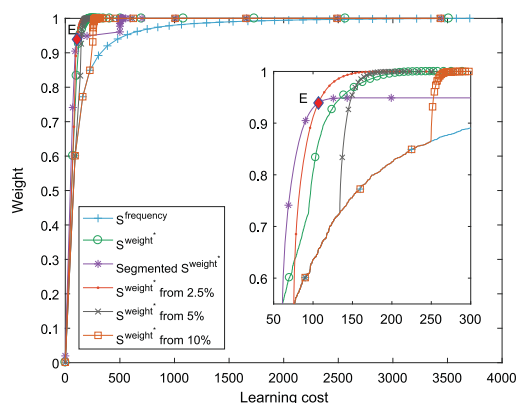
Table 8

Accumulated learning costs $\sum c_i$ of learning strategies for the number of accumulated learned words percentages. S^{weight^*} works excellently from the very beginning and when the completeness comes to 90%, i.e. 44 574 out of the total 49 527 words, the learning cost is only 943.25 out of 3,533.18.

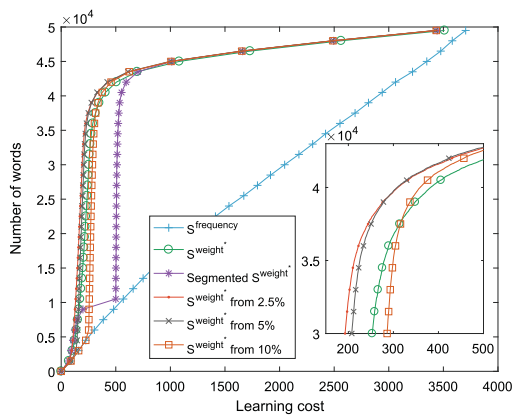
| $ w_{learned} $ | $S^{frequency}$ | S^{cost} | $S^{dictionary}$ | S^{random} | S^{weight} | S^{weight^*} |
|-----------------|-----------------|------------|------------------|--------------|--------------|----------------|
| 10% | 249.71 | 92.48 | 1705.05 | 1883.31 | 210.55 | 128.29 |
| 20% | 526.36 | 215.88 | 3183.52 | 3544.92 | 261.95 | 169.06 |
| 30% | 837.66 | 319.36 | 4396.23 | 4963 | 290.63 | 196.49 |
| 40% | 1202.20 | 411.28 | 5474.89 | 6144.53 | 314.53 | 217.14 |
| 50% | 1604.15 | 493.86 | 6482.9 | 7130.03 | 344.31 | 235.19 |
| 60% | 2017.87 | 561.34 | 7416.87 | 7966.22 | 395.39 | 252.22 |
| 70% | 2434.22 | 631.93 | 8100.74 | 8626.8 | 488.38 | 276.87 |
| 80% | 2860.11 | 736.7 | 8664.82 | 9199.05 | 758.76 | 369.99 |
| 90% | 3304.13 | 1283 | 9073.36 | 9644.66 | 1561.87 | 943.25 |
| 100% | 3705.19 | 4016.91 | 9522.49 | 10 004.18 | 3624.33 | 3533.18 |



(a) Accumulated Learning Cost $\sum c_i$ – Accumulated Frequency $\sum f_i$.



(b) Accumulated Learning Cost $\sum c_i$ – Accumulated Weight $\sum W_i$.



(c) Accumulated Learning Cost $\sum c_i$ – Accumulated number of learned words $|w_{learned}|$.

Fig. 6. Comparison of segmented strategies.

comparable, we select frequency first strategy $S^{frequency}$ and optimized weight first strategy S^{weight^*} as the standard methods. The results are very interesting. We plotted the results in Fig. 6.

We find that the strategy of S^{weight^*} from 2.5% is the most efficient: (1) In Fig. 6(a), though the Segmented S^{weight^*} strategy performs better before point D, it only happens when the cost is less than 175.51 which takes no more than 5.077% of all. Besides, at point D, which means with the same cost, 24 027 words have been learned using S^{weight^*} from 2.5% strategy while only 8659 words are learned by using Segmented S^{weight^*} strategy. (2) In Fig. 6(b), similarly, at point E, it costs only 106.89. However, in the following of more than 96.90% learning cost, the S^{weight^*} from 2.5% strategy is the best. (3) In Fig. 6(c), we can easily figure out that the strategy of S^{weight^*} from 2.5% is the best since the beginning.

6. Conclusion and discussion

Vocabulary building is important in English learning. It is challenging for learners to memorize English words and normally requires tremendous efforts. However, there is a lack of purposely optimized word learning strategy, or order, for learners to adopt in order to achieve high efficiency with least effects to overcome the learning costs. The frequency-based strategy does not have the least learning costs. English words are connected with each in two ways. Those structurally similar words are connected by sharing same components, while those words which are antonyms, synonyms, or related are associated with each other. These structural and meaning connections carry useful information of how a word can influence other words. In this paper, by constructing the word network using the structural and meaning information, we use the topological information from the word network with the frequency information to design learning strategies. Our quantitative results suggest that the weight first strategy S^{weight} and optimized weight first strategy $S^{weight*}$ can significantly outperform the naive strategies such as purely frequency first strategy $S^{frequency}$, the cost first strategy S^{cost} , the dictionary based strategy $S^{dictionary}$, or the random strategy S^{random} . This research provides a quantitative method of words learning using both topological information and frequency information. However, this initial approach is limited in many ways. It might cause some confusion if we have learned a word first then its synonyms since it is relatively not simple to tell the difference. In the future, we think we can introduce different weights for similar words, synonyms, antonyms, related words rather than treat them with same weights. We understand that the result is still not ready for practical use. However, we hope this work can serve as a starting point to attract scholars of language teaching, computational linguistics, corpus-based linguistic, network science, and data science for further researches on quantitative learning strategies.

Acknowledgments

The authors thank the editor and three anonymous reviewers for their valuable comments and suggestions for this paper. Their helpful input greatly improved the quality of this work. The authors also appreciate faculty members of School of Foreign Language, UESTC for their constructive discussions.

References

- [1] D. Graddol, *English Next*, British Council, 2006.
- [2] Ethnologue: Languages of the World (2016).
- [3] Oxford University Press, Oxford dictionaries, accessed 10-May-2016 (May 2016) [cited 10.05.2016]. URL <http://www.oxforddictionaries.com>.
- [4] X. Zhang, P. Yujiao, X. Xiaoqi, The correlation study on the relationship between the depth of vocabulary knowledge and comprehensive ability, *Studies in Literature and Language* 9 (1) (2014) 94–110. <http://dx.doi.org/10.3968/5230>.
- [5] S. Wang, An empirical study on the role of vocabulary knowledge in EFL listening comprehension, *Theory and Practice in Language Studies* 5 (5) (2015) 87–97.
- [6] D. August, M. Carlo, C. Dressler, C. Snow, The critical role of vocabulary development for English language learners, *Learning Disabilities Research & Practice* 20 (1) (2005) 50–57. <http://dx.doi.org/10.1111/j.1540-5826.2005.00120.x>.
- [7] Committee of College English Test Band 4 and 6, Outline of college English test band 4.
- [8] Committee of College English Test Band 4 and 6, Outline of college English test band 6.
- [9] Y. Minhong, *TOEFL Vocabulary-Word Root+Associative Memory-Portable Edition (Chinese Edition)*, Chinese Edition, Beijing Language and Culture University Press, 2011.
- [10] Y. Minhong, *IELTS Vocabulary Root + to Associate Memory Method - Chaos Order Portable Version, Chinese Edition*, Zhejiang Education Publishing House, 2011.
- [11] Office of the Ministry of Education, Annex of college English curriculum teaching requirements (2007).
- [12] Y. Zheng, L. Cheng, Test review: College English test (CET) in China, *Language Testing* 25 (3) (2008) 408–417. <http://dx.doi.org/10.1177/0265532208092433>.
- [13] K. Haastруп, B. Henriksen, Vocabulary acquisition: Acquiring depth of knowledge through network building, *International Journal of Applied Linguistics* 10 (2) (2000) 221–240. <http://dx.doi.org/10.1111/j.1473-4192.2000.tb00149.x>.
- [14] P. Meara, Teaching vocabulary, *System* 12 (2) (1984) 185–186. [http://dx.doi.org/10.1016/0346-251X\(84\)90029-0](http://dx.doi.org/10.1016/0346-251X(84)90029-0).
- [15] J.H. Hulstijn, M. Hollander, T. Greidanus, Incidental vocabulary learning by advanced Foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words, *The Modern Language Journal* 80 (3) (1996) 327–339.
- [16] Y. Gu, R.K. Johnson, Vocabulary learning strategies and language learning outcomes, *Language Learning* 46 (4) (1996) 643–679. <http://dx.doi.org/10.1111/j.1467-1770.1996.tb01355.x>.
- [17] N. Schmitt, D. Schmitt, Vocabulary notebooks: Theoretical underpinnings and practical suggestions, *ELT Journal* 49 (2) (1995) 133–143. <http://dx.doi.org/10.1093/elt/49.2.133>.
- [18] L. Cengizhan, Vocabulary learning strategies: A case of Edirne Anatolian high school, *Procedia - Social and Behavioral Sciences* 15 (2011) 1870–1874. <http://dx.doi.org/10.1016/j.sbspro.2011.04.018>. 3rd World Conference on Educational Sciences – 2011.
- [19] A. Nacera, Languages learning strategies and the vocabulary size, *Procedia - Social and Behavioral Sciences* 2 (2) (2010) 4021–4025. <http://dx.doi.org/10.1016/j.sbspro.2010.03.634>. Innovation and Creativity in Education.
- [20] K. Hong-Nam, A.G. Leavell, Language learning strategy use of ESL students in an intensive English learning context, *System* 34 (3) (2006) 399–415. <http://dx.doi.org/10.1016/j.system.2006.02.002>.
- [21] P.Y. Gu, Vocabulary learning in a second language: Person, task, context and strategies, *TESL-EJ* 7 (2) (2003) 1–25.
- [22] R.L. Oxford, R.C. Scarcella, Second language vocabulary learning among adults: State of the art in vocabulary instruction, *System* 22 (2) (1994) 231–243. [http://dx.doi.org/10.1016/0346-251X\(94\)90059-0](http://dx.doi.org/10.1016/0346-251X(94)90059-0).
- [23] A.U. Chamot, Issues in language learning strategy research and teaching, *Electronic Journal of Foreign Language Teaching* 1 (1) (2004) 14–26. <http://dx.doi.org/10.1017/S0261444808005612>.
- [24] T. Huckin, J. Coady, Incidental vocabulary acquisition in a second language: A review, *Studies in Second Language Acquisition* 21 (02) (1999) 181–193. <http://dx.doi.org/10.1017/S0272263199002028>.

- [25] D. Brodsky, Spanish vocabulary: An etymological approach, University of Texas Press, 2008.
- [26] Y. Huang, Learning Spanish Words Through Etymology and Mnemonics, Xlibris, 2016, p. 286.
- [27] J.C. Loach, J. Wang, Optimizing the learning order of chinese characters using a novel topological sort algorithm, PLOS ONE 11 (10) (2016) 1–17. <http://dx.doi.org/10.1371/journal.pone.0163623>. <http://dx.doi.org/10.1371%2Fjournal.pone.0163623>.
- [28] X. Yan, Y. Fan, Z. Di, S. Havlin, J. Wu, Efficient learning strategy of chinese characters based on network approach, PLoS ONE 8 (8) (2013) 1–7. <http://dx.doi.org/10.1371/journal.pone.0069745>, arXiv:1303.1599.
- [29] J. Li, J. Zhou, Chinese character structure analysis based on complex networks, Physica A: Statistical Mechanics and its Applications 380 (1–2) (2007) 629–638. <http://dx.doi.org/10.1016/j.physa.2007.02.059>.
- [30] C.-C. Wang, M.-L. Wei, Y.-L. Chang, H.-C. Chen, Y.-L. Chung, J.-F. Hu, Advanced learning chinese characters strategy based on the characteristics of component and character frequency, in: CogSci 2014 Proceedings, 2014, pp. 3067–3071.
- [31] S. Arbesman, S.H. Strogatz, M.S. Vitevitch, Comparative analysis of networks of phonologically similar words in English and Spanish, Entropy 12 (3) (2010) 327. <http://dx.doi.org/10.3390/e12030327>.
- [32] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
- [33] D.J. Watts, Small Worlds: The Dynamics of Networks Between Order and Randomness, Princeton university press, 1999.
- [34] S.N. Dorogovtsev, J.F.F. Mendes, Language as an evolving word web, Proceedings of the Royal Society of London B: Biological Sciences 268 (1485) (2001).
- [35] M. Markořová, Network model of human language, Physica A: Statistical Mechanics and its Applications 387 (2) (2008) 661–666. <http://dx.doi.org/10.1016/j.physa.2007.09.027>.
- [36] S. Zhou, G. Hu, Z. Zhang, J. Guan, An empirical study of Chinese language networks, Physica A: Statistical Mechanics and its Applications 387 (12) (2008) 3039–3047. <http://dx.doi.org/10.1016/j.physa.2008.01.024>.
- [37] G.A. Wachs-Lopes, P.S. Rodrigues, Analyzing natural human language from the point of view of dynamic of a complex network, Expert Syst. Appl. (ISSN: 09574174) 45 (2016) 8–22. <http://dx.doi.org/10.1016/j.eswa.2015.09.020>.
- [38] P. Tang, T.W. Chow, Mining language variation using word using and collocation characteristics, Expert Syst. Appl. 41 (17) (2014) 7805–7819. <http://dx.doi.org/10.1016/j.eswa.2014.05.018>.
- [39] U. Shoaib, N. Ahmad, P. Prinetto, G. Tiotto, Integrating multiwordnet with italian sign language lexical resources, Expert Syst. Appl. 41 (5) (2014) 2300–2308. <http://dx.doi.org/10.1016/j.eswa.2013.09.027>.
- [40] Y. Li, Small-world patterns in Chinese phrase networks, Chinese Science Bulletin 50 (3) (2005) 286. <http://dx.doi.org/10.1360/982004-392>.
- [41] R. F. i. Cancho, R.V. Solé, The small world of human language, Proceedings of the Royal Society of London B: Biological Sciences 268 (1482) (2001) 2261–2265. <http://dx.doi.org/10.1098/rspb.2001.1800>.
- [42] R. Ferrer i Cancho, R.V. Solé, R. Köhler, Patterns in syntactic dependency networks, Phy. Rev. E 69 (5) (2004) 051915. <http://dx.doi.org/10.1103/PhysRevE.69.051915>. <http://link.aps.org/doi/10.1103/PhysRevE.69.051915>.
- [43] M. Medeiros Soares, G. Corso, L. Lucena, The network of syllables in Portuguese, Physica A: Statistical Mechanics and its Applications 355 (2) (2005) 678–684. <http://dx.doi.org/10.1016/j.physa.2005.03.017>.
- [44] G. Peng, J.W. Minett, W. S.-Y. Wang, The networks of syllables and characters in Chinese, Journal of Quantitative Linguistics 15 (3) (2008) 243–255. <http://dx.doi.org/10.1080/09296170802159488>. <http://www.tandfonline.com/doi/abs/10.1080/09296170802159488>.
- [45] T.T. Hills, M. Maouene, J. Maouene, A. Sheya, L. Smith, Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? Psychological Science 20 (6) (2009) 729–739. <http://dx.doi.org/10.1111/j.1467-9280.2009.02365.x>. <http://www.ncbi.nlm.nih.gov/pubmed/19470123>, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4216730>, <http://pss.sagepub.com/lookup/doi/10.1111/j.1467-9280.2009.02365.x>.
- [46] J. Ke, Y. Yao, Analysing language development from a network approach, Journal of Quantitative Linguistics 15 (1) (2008) 70–99. <http://dx.doi.org/10.1080/09296170701794286>. <http://www.tandfonline.com/doi/abs/10.1080/09296170701794286>.
- [47] M. Steyvers, J.B. Tenenbaum, The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth, Cognitive Science 29 (1) (2005) 41–78. <http://dx.doi.org/10.1207/s15516709cog2901.3>. <http://www.ncbi.nlm.nih.gov/pubmed/21702767>, <http://doi.wiley.com/10.1207/s15516709cog2901.3>.
- [48] A.E. Motter, A.P.S. de Moura, Y.-C. Lai, P. Dasgupta, Topology of the conceptual network of language, Phy. Rev. E 65 (6) (2002) 065102. <http://dx.doi.org/10.1103/PhysRevE.65.065102>. <http://link.aps.org/doi/10.1103/PhysRevE.65.065102>.
- [49] M. Zolfagharkhani, R.G. Moghadam, The effect of etymology instruction on vocabulary learning of upper- intermediate EFL Iranian learners, Canadian Social Science 7 (6) (2011) 1–9. <http://dx.doi.org/10.3968/j.css.1923669720110706.180>.
- [50] N. Schmitt, P. Meara, Researching vocabulary through a word knowledge framework, Studies in Second Language Acquisition 20 (1997) 17–36. <http://dx.doi.org/10.1017/S0272263197001022>.
- [51] E. Partridge, Origins: A short etymological dictionary of modern English, Routledge, 2006.
- [52] T.F. Hoad, The Concise Oxford Dictionary of English Etymology, Oxford University Press Oxford, 1993.
- [53] D. Tufiş, D. Ştefănescu, Experiments with a differential semantics annotation for wordNet 3.0, Decis. Support Syst. 53 (4) (2012) 695–703. <http://dx.doi.org/10.1016/j.dss.2012.05.026>. <http://linkinghub.elsevier.com/retrieve/pii/S0167923612001376>.
- [54] M. Cecchini, H. Aytug, G.J. Koehler, P. Pathak, Making words work: Using financial text as a predictor of financial events, Decis. Support Syst. 50 (1) (2010) 164–175. <http://dx.doi.org/10.1016/j.dss.2010.07.012>. <http://www.sciencedirect.com/science/article/pii/S0167923610001181>.
- [55] B. Furlan, V. Batanović, B. Nikolić, Semantic similarity of short texts in languages with a deficient natural language processing support, Decis. Support Syst. 55 (3) (2013) 710–719. <http://dx.doi.org/10.1016/j.dss.2013.02.002>. <http://www.sciencedirect.com/science/article/pii/S0167923613000614>.
- [56] Dictionary.com, LLC, Dictionary.com online dictionary, accessed 10-May-2016 (May 2016) [cited 10.05.2016]. <http://www.dictionary.com>.
- [57] Dictionary.com, LLC, Thesaurus.com online dictionary, accessed 10-May-2016 (May 2016) [cited 10.05.2016]. <http://www.thesaurus.com>.
- [58] Merriam-Webster, Inc, Merriam-Webster online dictionary, accessed 10-May-2016 (May 2016) [cited 10.05.2016]. <http://www.merriam-webster.com>.
- [59] Interapple, Inc, Webster-dictionary online dictionary, accessed 10-May-2016 (May 2016) [cited 10.05.2016]. <http://www.webster-dictionary.org>.
- [60] T. Brants, A. Franz, Web 1T 5-gram Version 1 LDC2006T13. DVD.
- [61] L. Lu, Z.-K. Zhang, T. Zhou, Deviation of Zipf's and Heaps' laws in human languages with limited dictionary sizes., Sci. Rep. 3 (2013) 1082. <http://dx.doi.org/10.1038/srep01082>. arXiv:1202.2903.
- [62] M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks, 2009.