A Contrastive Analysis between English Vocabulary Profile and College English Wordlist

Dongyun Sun

College of Foreign Languages and Literatures, Fudan University, Shanghai, China

Abstract—**This paper makes a contrastive analysis between *English Vocabulary Profile (EVP)* and *College English Wordlist* (CEW) from the aspects of philosophy, vocabulary, criteria and word frequency. Results show that the latter has a numerical advantage in the mere term of vocabulary, but these words are mainly passive ones. By contrast, the former attaches greater importance to the number of active words and usage, and is more concerned with learners' actual word power. This study sheds light on the reform of college English teaching and the preparation of the vocabulary syllabus.**

Index Terms—**English vocabulary profile, vocabulary learning, active words, passive words**

I. OVERVIEW

The study of learner vocabulary lists dates back to as early as 1588, when Timothy Bright published an "island vocabulary" of 559 English words which could be used to cover the meanings of 6,000 (Fox, 1979, p.65). However, the empirical study of vocabulary size for learners did not start until the 20th century, when Edward J. Thorndike compiled *A Teacher's Word Book* (1921), *A Teacher's Word Book of 20,000 Words* (1932), *A Teacher's Word Book of 30,000 Words* (1944). In addition, the famous linguist CK Ogden (1930) introduced the basic English wordlist consisting of 850 words, while Michael West (1953) introduced a *General Service List* of 2000 basic words.

In an era of science quantification, the mechanistic view that learners can easily surmount the language barrier by a grasp of grammar and a sizeable vocabulary seems virtually self-evident to second-language teachers. It is no wonder that training courses promising quick expansion of vocabulary have always remained a mainstay in various language training programs across the world. In the past five decades or so, however, a paradigm shift towards empiricism marked by the rise of corpus linguistics gave rise to a new boom of vocabulary lists, notable among which is Mark Davies' series of wordlists based on his Corpus of Contemporary American English (COCA)[1]. The corpus-driven methodology has also branched into empirical analyses of language units beyond the word level which include, inter alia, collocations and word chunks. Those so-called "pre-fabricated chunks" (Bishop, 2004; Boers et al., 2006) can be automatically extracted from the corpus by the frequency of occurrence. The ability to use language chunks is also deemed an important indicator of language competence.

In the same vein, the College English Teaching Steering Committee of China also launched *College English Curriculum Requirements* which conveys a College English Wordlist (7,821 words), of which the active vocabulary being 2,404. English Major Teaching Syllabus Glossary Working Group launched a TEM-4/8 Vocabulary (2004) of 13,000 words. These quantitative indicators that are harvested from first-hand language materials prove useful in guiding our teaching. On the one hand, language education policymakers can use these wordlists to determine the level of proficiency of language learners as an important indicator; on the other hand, the boom of overseas education in China witnessed Chinese students cramming words in preparation for TOEFL, GRE or IELTS. It is thus no wonder that books or training courses promoting various mnemonic skills of words have been embraced by learners with blind craze. Some English learning websites and online dictionaries have launched English word mnemonic functions and applications; to adapt to the era of mobile Internet, vendors have also released a plethora of word mnemonic apps that can be installed into smartphones. This exam-centered learning approach that aspires to instant success of word learning often leads to over-reliance on the vocabulary. As a result, the student's negative vocabulary can grow rapidly, but their positive vocabulary is disproportionally small in comparison. In other words, although students can recognize a large number of words, they can barely read professional articles, for their active vocabulary is still very poor. The results of the study of the learner corpus show that Chinese learners command limited types of word collocations while errors in collocations abound (Deng, 2005, p.9), over-reliance on high-frequency vocabulary (Deng, 2007, p.17), colloquial style in written communication which does not improve in proportion to the overall level of English competence (Wen, Ding & Wang, 2003; Liu, 2005).

In addition to the abovementioned problems exposed by scholars based on their study of learners corpora, learners themselves are not satisfied with the efficiency of their foreign language learning. Therefore complaints on the "time-

---

[1] Mark Davies produced 5,000 to 60,000 word vocabulary words and 100,000 word vocabulary, of which 50 word vocabulary can be downloaded free, see http://www.wordfrequency.info/ .

consuming inefficiency" in English learning arose one after another, with maladies labelled "dumb English" or "high score but poor competence" and various other accusations on an endless litany. However, the excessive concern for vocabulary and glossaries cannot be attributed to foreign language teachers, nor is it the intention of policy makers. The general consensus of the foreign language community is that intensive study of words in the hope of cracking vocabulary may show short-term benefits of improvement but cannot improve the vocabulary competence (Gui, 2006, p.62). If such hasty efforts are not complemented by a proportionate amount of aural or visual input of language materials, language attrition may follow suit after the examination. As Gui (2006, p.64) points out, whatever approach to vocabulary learning should be based on language in use. Only the use of language, including listening, speaking, reading, writing can help learners consolidate their knowledge of vocabulary. If you do not use the words you learn, they will soon be forgotten. Therefore when learners try to memorize words without using them, they gain immediate benefits at the cost of long-term knowledge. Although many learners can pass language exams, they usually find that their verbal or written communication skills remain as poor as before.

Therefore it is necessary to interpret the English Vocabulary Profile by University of Cambridge ESOL Examinations to see how an empirical approach to vocabulary in use may reveal the authentic vocabulary competence of English learners.

II. INTRODUCTION TO ENGLISH VOCABULARY PROFILE (EVP) AND COLLEGE ENGLISH WORDLIST (CEW)

A. The English Profile and the Common European Framework of Reference (CEFR)

English Profile was jointed developed in 2007 by the University of Cambridge ESOL Examinations, Cambridge University Press, British Council, Cambridge University, University of Bedfordshire and English UK under the active support from the Council of Europe. Currently the project is still under way. The EVP project is already launched, while the English Functions Profile, English Grammar Profile and other modules are still being developed. The most distinctive feature of EVP is that this is a corpus-driven glossary based on Cambridge Learner Corpus. The corpus brings together hundreds of thousands of test papers around the world by students who take Cambridge English examinations. From this huge corpus of more than 45 million words, researchers extract words, concepts and phrases that reflect the general vocabulary of learners of different levels (Good, 2010, p.114).

For each level of English competence there is one glossary that is divided into two versions, namely the British English version and the American English version. The glossary is composed of words, phrases, phrasal verbs and idioms, which reflect the actual language competence of learners who are qualified for each level of the CEFR. A major feature is that words and phrases are assigned to different levels by the frequency of their meaning, as shown below:

Figure 1. The word "degree" in four different levels in EVP

This example shows that the verb "degree" is divided into four separate entries based on its four different meaning, each classified into a separate CEFR ranging from B1 to C2. To help learners, the site also provides example sentences from the dictionary and sentences written by EFL learners, ending with tags indicating the level and the geographical region of the candidate. These examples from the learner corpus can well reflect the true linguistic competence of the learners.

It should be explained that the level A1 to C2 corresponds exactly to the Common European Framework of Reference (CEFR). The CEFR is a project developed by Council of Europe. It divides language learner/user's language proficiency into six levels of three classes, namely, C2 Mastery, C1 Effective operational proficiency, B2 Vantage, B1 Threshold, A2 Waystage and A1 Breakthrough. Grade C corresponds to a proficient user, while the class B reflects an independent user. The class A1 and A2 collectively reflect a basic user. This standard has been widely implemented around the world. Chinese linguists have begun to study the issues of aligning various language tests to the CEFR. Since the Cambridge English examinations are already aligned with the CEFR, the English competence of candidates who pass the Cambridge English (Advanced) Exam should be considered C1 on the CEFR list[2].

The current research is performed by comparing the CVP glossary with the CEW glossary developed by the Chinese team of college English professors. The CEW glossary was released in 2007 as the appendix of the document College English Curriculum Requirements issued by the Ministry of Education of China, exerting an increasingly broad impact on the English education of China, because it has an important guiding significance to college English teaching and testing.

B. Comparability and the Significance of a Comparative Study

The comparability is well justified by the resemblance between the two projects. Both the CVP and the CEW were developed to determine the English competence of EFL learners by some of the most distinguished scholars in the two countries. The current analysis is based on the following considerations:

1. The learner orientation

Both CEW and CVP are oriented towards EFL learners. The CEW project is oriented towards college-level EFL learners of China, as a benchmark for the evaluation of students' English competence in teaching and testing. The CVP project is developed by top-notch EFL scholars who built a learner corpus composed of Cambridge English exam test papers. It remains unknown whether the former was built in a corpus-driven approach, but these two projects share the same orientation, and the target group of students are closely intertwined and overlapped.

2. Levels of English competence

Cambridge English exams are composed of five levels, each corresponding to one of the levels ranging from A2 to C2, a system which is perfectly compatible with the CEFR system.

A2	Key
B1	Preliminary
B2	First
C1	Advanced
C2	Proficiency

Figure 2. The scale of levels of Cambridge English examinations.

By contrast, the CEW is designed for college students in China. Although without the benchmark of "can-do" lists in College English Curriculum Requirements (CECR) or any official alignment between the College English Test (CET) Band 4 or Band 6 with the CEFR scale, the author observes that most college English learners in China can attain the level of B on the CEFR scale and a few can attain C. This is also testified by three-level classification of teaching requirements in the CECR document. To illustrate the comparability between the "can-do" lists of the CEFR document and the "can-do" descriptions of the CECR, this paper uses the example of written skills:

---

[2] See http://www.cambridgeenglish.org/exams/advanced for a detailed description of the alignment of Cambridge English exams with the CEFR.

TABLE 1.
"CAN-DO" DESCRIPTORS OF WRITTEN COMPETENCE BETWEEN CECR AND CEFR

CECR "can-do" descriptors (Competence of written expression) (Ministry of Education of China, 2007) (My translation)	CEFR "can-do" descriptors (Written Interaction) (Council of Europe, 2001, p.83)	CEFR "can-do" descriptors (Written Production) (Council of Europe, 2001, p.61)
Competence of written expression (**Advanced**): can write short reports and articles in English in one's own field or discipline; can freely express personal views in written form; can write a descriptive or expository article of at least 200 words within 30 minutes, with a clear view, rich content, a clear structure and good logic	I can express myself with clarity and precision, relating to the addressee flexibly and effectively in an assured, personal style.	**C2**. I can write clear, smoothly flowing text in an appropriate style. I can write complex letters, reports or articles, which present a case with an effective logical structure, which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.
		C1. I can express myself in clear, well-structured text, expressing points of view at some length. I can write detailed expositions of complex subjects in an essay or a report, underlining what I consider to be the salient issues. I can write different kinds of texts in a style appropriate to the reader in mind.
Competence of written expression (**Upper**): can express personal views on general topics; can write the English abstract of a paper of one's own field or discipline; can write a short paper on one's own field of study; can describe charts and diagrams with words; can produce a short composition of at least 160 words within 30 minute, with a complete idea, a clear view, good structure and smooth expressions.	**B2**. I can write letters highlighting the personal significance of events and experiences.	**B2**. I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view.
	B1. I can write personal letters describing experiences and impressions.	**B1**. I can write straightforward connected text on topics, which are familiar, or of personal interest.
Competence of written expression (**General**): can accomplish general writing assignments; can describe personal experience, feelings, emotions and events; can write for practical purposes of generic genres; can write a short composition of at least 120 words on a general topic or outline within 30 minutes, with a complete idea and a clear central idea, can use the right words to achieve coherence; can master basic writing skills.	**A2**. I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something.	**A2**. I can write a series of simple phrases and sentences linked with simple connectors like "and", "but" and "because".
	A1. I can write a short, simple postcard, for examples sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	**A1**. I can write simple isolated phrases and sentences.

This table shows that the "can-do" statements of the CECR are divided into three levels while those of the CEFR are divided into six levels. In spite of the differences in classification and description, we can find that there is a rough correspondence in between. For example, the highest levels of written competence of both are quite similar with each other, both covering reports and articles. However, the CECR's requirements are obviously lower than those at the C2 level of the CEFR. Therefore this paper tentatively aligns the Advanced level of written communication of the CECR to a level between B2 and C1 of the CEFR.

3. The size of glossaries

According to Capel (2010: 5), the sizes of vocabulary of CEFR A1 to B2 are as follows:

CEFR A1	601
New words at CEFR A2	925
New words at CEFR B1	1,429
New words at CEFR B2	1,711

The grand total of words from A1 to C2 is 6,970 headwords, so we can estimate that the new Grade C should be composed of 2,304 words (Capel, 2012: 4).

The source of Grade A and Grade B is slightly different from that of Grade C. The former derived entirely from the answer sheets of test papers while the Grade C reference vocabulary is collected entirely from a corpus of native English speakers. Capel (2010, 2012) explains that in Grade A and Grade B, there is no drastic difference between the learner's receptive knowledge and output vocabulary, and the modern communicative approach to language teaching encourages students to use what they have learned in actual use. Especially in this Internet era, students have far better opportunities of using foreign language, so researchers did not distinguish between passive vocabulary and active vocabulary for the levels below B2. By contrast, for the levels starting from C1, the receptive knowledge may be more extensive than the actual output of the vocabulary range, and learners have learned to use the logical skills to guess the actual meanings of words by the context, so they consulted the frequency information of corpora of native English speakers and academic English vocabulary during their preparation of Grade C glossary.

III. A CONTRAST BETWEEN EVP AND CEW

A. Data Collection

The EVP project is freely available at http://vocabulary.englishprofile.org/ to all interested language teachers and researchers. Subscription is free to all who register for the project. However, the glossary can only be retrieved in dynamic webpages through manual queries. Therefore the author had to scroll through all pages from A1 to C2 and copy all the content of each standalone meaning and is listed at webpage. The HTML webpages are cleaned manually and imported into a spreadsheet document. Thus we arrived at a complete glossary of the EVP.

B. A Contrastive Analysis of the Criteria of Selection

First, the difference in the criterion of entry separateness.

The most obvious difference between words in the CEW and the EVP is the criterion of entries. As mentioned above, the words of the EVP are listed as separate entries by each standalone meaning and are listed at different levels by their frequency of occurrence. The benefit of this approach is that it can amply reflect the degree of mastery and ability of the learner. In contrast, the CECR promulgated by China's Ministry of Education in 2007 adopts a much simpler and traditional approach to the entries. Each standalone word as an entry appears only once in the CEW, however many meanings it has. In some cases such as the homographs, words are separated. For example, the homograph "bank" is divided into bank1 and bank2. The latter is a traditional approach to vocabulary that treats words in a lexicographical manner. By contrast, the former is the latest novelty thanks to the prevalence of corpus linguistics.

Second, the difference in the handling of idioms and collocations.

The difference in the selection of idioms and collocations reflect different approaches to language learning. The CEW reflects a rather traditional approach to the compilation of wordlists. While phrases constitute an integral part of the wordlist, collocations never made their way into it. By contrast, the EVP embraces collocations and daily word combinations, reflecting a more liberal approach to vocabulary learning. Here are some of the examples from the EVP:

I can't tell you how ...
taste good/bad/sweet
reasonably priced
Oh no!
have been meaning to do

These collocations or phrases can hardly be found in the entries of traditional dictionaries or wordlists. This makes it much more difficult to compare the two wordlists. Therefore this paper makes a tentative comparison of the two wordlists at the word level rather than the phrase level.

Third, the proportion of active vs. passive vocabulary.

It was clearly specified in the CECR that of the 7,676 words and phrases in the CEW, 2,367 belong to active vocabulary. However, there is no description as to the approach of preparation to this wordlist. By contrast, as aforementioned, the EVP wordlist is composed mainly of the active vocabulary extracted from test papers. Only C1 and C2 were prepared based on the corpora of native English speakers.

C. Part-of-speech (POS) Comparison

Since a POS comparison can reveal the different attitudes to the compilation of wordlists, the author performed POS tagging and retrieval using the software program Tree Tagger (Liang, 2010). Results are shown below:

TABLE 2.
COMPARISON OF EVP AND CEW BY POS

POS Codes	DEFINITIONS	EVP	CEW
CC	Coordinating conjunction	11	10
CD	Cardinal number	27	28
DT	Determiner	13	11
FW	Foreign word	2	2
IN	Preposition or subordinating conjunction	61	56
JJ	Adjective	1391	1497
JJR	Adjective, comparative	8	4
JJS	Adjective, superlative	9	3
LS	List item marker	0	0
MD	Modal	10	11
NN	Noun, singular or mass	3430	4523
NNS	Noun, plural		32
NP	Proper noun, singular	2	36
NPS	Proper noun, plural	0	0
PDT	Predeterminer	1	1
POS	Possessive ending	0	0
PP	Personal pronoun	17	18
PP $	Possessive pronoun	7	6
RB	Adverb	*510*	*217*
RBR	Adverb, comparative	10	4
RBS	Adverb, superlative	0	1
RP	Particle	0	3
SYM	Symbol	1	0
TO	To	1	2
UH	Interjection	1	0
VB	Verb, base form	1	1
VV	verb, base form	569	808
VVD	verb, past tense	89	25
VVN	verb, gerund / participle	*116*	*25*
VVG	verb, past participle	*121*	*37*
VVP	verb, present, non-3rd p.	244	268
VVZ	verb, present 3d p. sing.	*34*	*2*
WP	wh-determiner	4	4
WDT	wh-pronoun	3	3
WP $	possessive wh-pronoun	1	1
WP	verb, base form	4	4
WRB	verb, past tense	7	7

As shown in the above table, the total vocabulary of the EVP is lower than that of the CEW, so the numbers of nouns and verbs are justifiably less than those of the latter. However, it's interesting to note that, in spite of the gap between the EVP and the CEW, there are a few exceptions in which the former far outnumbers the latter. The most prominent case is the number of adverbs (RB)—the number of adverbs in the EVP is about twice as much as that of the latter. Other significant variations that deserve our attention are the different derivations of verbs: past participle, third person singular and other forms (all these deviations are marked in italicized bold letters in the table). This phenomenon is quite interesting. It was observed that the adverbs of "-ly" suffices included in the EVP are as high as 365, while the CEW included only 76. It should be noted that the conventional approach to lexicography is that the majority of adverbs are not included as separate entries in the learner's dictionaries, but rather in the part of definition or example sentences under the entries of their adjective forms. For example, the word "importantly" in the EVP can be treated as a separate entry in *Longman Dictionary of Contemporary English (LDOCE)*; however in *Oxford Advanced Learner's English-Chinese Dictionary (OALD)*, *Collins COBUILD English Dictionary (COBUILD)* and most other dictionaries, it is only found under the entry "important." It seems that this approach of the EVP is aligned with that of LDOCE, that is, emphasis on the description of the learner's actual vocabulary competence.

The various inflections of verbs, however, derive from the fact that the EVP incorporates a large number of idioms and fixed combinations which entail the necessary inflections to reflect the actual use of words. For example, the third person singular verb form (VVZ) "appears" can be found in the entry "it appears (that)" (CEFR B2). It also shows great attention to phrases, idioms and collocations in the EVP. Although a far cry from the traditional practice of Thorndike, it reflects the new approaches of a new era in which the words are not treated in isolation but in combination with other words in the context.

D. Comparative Analysis of the Unique Word

Unique words are a good indicator of the differences between two wordlists. They can be extracted by following Feng's approach (2010, p.216). In this paper, the author used the computer program Concordance 3.2 (Watt, 2004). Results of this comparison show an interesting difference between the CEW and the EVP:

Although the EVP (British English portion) has a vocabulary of 6,935 while the CEW has 7,676, with the gap of only 741 in between. However, the number of unique words in the CEW is 2,572. In other words, these 2,572 words cannot be found in the EVP. Conversely, the number of unique words in the EVP is 1,875. So what are the causes of such great difference?

To examine this difference, the author chose the unique words of Letter A of the EVP and the CEW. Though limited in quantity, this case study contributes to understanding of the principles of preparation. In the case of the EVP, most unique words are derivations that are not routinely included in the entries of traditional dictionaries, especially the adverbs ending with "-ly" and the adjectives ending with "-ed" (usually deriving from verbs). The number of the former is 19 while the latter numbers 17, together accounting for nearly half of the 82 unique words of the EVP in the part of Letter A. In contrast, the 206 unique words of the CEW seem more "demanding", reflecting a more advanced level of vocabulary in use. Those words like "abbey", "abduct", "abreast", "accomplish" are obviously advanced words with much higher level of difficulty. From the part of Letter A, the level of difficulty of the CEW is generally higher than the overall requirements of the EVP. And since the EVP is perfectly compatible with the CEFR, it seems that some words in the CEW have surpassed the level of CEFR C2.

E. Analysis of Causes

1) Corpus selection. EVP is mainly based on the learner corpus of foreign candidates for the Cambridge English examinations. Those foreign learners' active vocabulary is limited, but their actual passive vocabulary may be much larger than the former. The CEW, by contrast, is based on college English textbooks in China and other published wordlists, both of which being basically built upon the corpora of English native speakers.

2) The competence orientation. EVP embodies the plurilingual language and pluricultural competence orientation, emphasizing the "can do" indicators of learners in the real communication environment. As a result, this wordlist is not built for the pursuit of advanced vocabulary or the reading competence in written communication, but rather focuses on the verbal communication and cultural experience of learners.

Of course, given the differences between the EVP and the CEW, a contrastive of the Letter A part is still inadequate, calling for more means to be incorporated for further in-depth analysis.

IV. SUMMARY AND DISCUSSION

EVP embodies the learner competence-oriented foreign language teaching philosophy. As the latest research achievement of the international academic community, its main features include separate entries of words based on meaning and word class, a large number of idioms and regular combinations, authentic example sentences retrieved from the learner corpus, etc. These deserve full attention and in-depth study from the foreign language community.

The EVP project is based on the Cambridge learner corpus from Cambridge English Examination testpapers. The examinations per se are not a natural environment for language in use, and candidates are supposed to write on prescribed topics, hence it is unlikely for them to produce their most natural writing or oral output. The EVP team is currently building corpora based on other learner corpora, including classroom assignments, extracurricular assignments, classroom discussions, day-to-day chats, based on data from 16 primary and secondary schools and individual teachers worldwide (Good, 2010). Although some of the passive words at the C1 and C2 levels were extracted from the vocabulary of Academic Word List (Coxhead, 2000), it is still insufficient. According to the provisions of CEFR, C2 should be equivalent to 8.5-9 points of IELTS, and with very few candidates reaching the point of 9. The 6,000 words and phrases in EVP may mislead learners into believing that they could get 9 points in IELTS if they can master these words, but in fact it is obviously impossible.

Of course, the use of corpus tools in this study also has some limitations. On the one hand, the corpus tool of this study can only reveal the frequency of words at the word level, but cannot fully consider the idioms and fixed collocations; on the other hand, in order to facilitate research, the author performed POS tagging with Tree Tagger. Given the 3% error rate of automatic POS tagging, there is understandably a small margin for errors in this quantitative research.

In short, the EVP is a useful reference wordlist, and its elaborate description of the actual vocabulary competence of learners sheds light on future efforts. However, it has to be noted that EVP as a descriptive research cannot be used directly to guide the compilation of teaching materials and language exam research and development, and proper caution should be made to prevent students from rote memorization of the EVP wordlist, for this approach deviated from the original intention of the researchers and is thus not conducive to the normal language learning process.

REFERENCES

[1] Bishop, H. (2004). Noticing formulaic sequences—A problem of measuring the subjective. *LSO Working Papers in Linguistics*, 4, 15-19.

[2]   Boers, F., J. Eyckmans, J. Kappel, H. Stengers & M. Demecheleer. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language teaching research*, 10(3), 245-261.
[3]   Capel, A. (2010). A1-B2 vocabulary: insights and issues arising from the English Profile Wordlists project. *English Profile Journal*, 1 (1), 1-11.
[4]   Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal,* 3(1), 1-14.
[5]   Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.
[6]   Coxhead, A. (2000). The Academic Word List. http://www.victoria.ac.nz/lals/resources/academicwordlist/default.aspx. (Accessed 15 March, 2017).
[7]   Deng, Y. (2007). Review of Learner Corpora and SLA Research. *Foreign Language World*, 1, 16-21.
[8]   Deng, Y., T Wang (2005). Statistics in Collocation Extraction and Computer Implementation. *Technology Enhanced Foreign Language Education*, 5, 26-29.
[9]   Fox, L. (1979). On acquiring an adequate second language vocabulary. *Journal of Basic Writing*, 2(3), 68-75.
[10]   Good, M. (2010). Meet the English Profile Wordlists: describing what learners Can Do. *Dictionaries: Journal of the Dictionary Society of North America*, 31 (1), 113-117.
[11]   Gui, S. (2006). Aspects of English Vocabulary Learning - An Answer. *Foreign Language World*, 1, 57-65.
[12]   Liang, M. (2010). Tree Tagger. National Research Centre for Foreign Language Education, Beijing Foreign Studies University.
[13]   Liu, X. (2005). Chinese university English learner written in colloquial tendency. *Higher Education Research Institute of Technology*, 24: 109-112.
[14]   Ministry of Education of China. (2007). College English Curriculum Requirements. http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/cmsmedia/image//UserFiles/File/2008/05/06/2008050611/20080506 11_163068.doc. (accessed 1 March, 2017).
[15]   Ogden, C. K. (1930). Basic English: A General Introduction with Rules and Grammar. London: Paul Treber & Co., Ltd.
[16]   Watt, R. J. C. (2004). Concordance (3.2 edn) (software), www.concordancesoftware.co.uk/ (Accessed 11 November 2012).
[17]   Wen, Q., Ding Y. & Wang W. (2003). Chinese Students' English Writing Colloquial trend—a high level of English Learners Corpus analysis. *Foreign Language Teaching and Research*, 35, 268-274.
[18]   West, M. (1953). A general service list of English words. London: Longman, Green & Co.

Dongyun Sun was born in Anhui Province, China. She received her PhD from Fudan University in January, 2015. She is currently a lecturer in the College of Foreign Languages and Literatures, Fudan University.
Her research interests include educational linguistics and translation teaching.