A Corpus-based Empirical Study on the Use of *Get* by English Majors

Li-Xia ZHANG[1,a,*]

[1]Inner Mongolia University of Finance and Economics, Hohhot, China, 010070

[a]2008zhanglixia@163.com

*Corresponding author

Keywords: corpus-based, collocation, semantic prosody.

Abstract. The study of English lexical competence has been the concern of many linguists and foreign language teachers in recent years. With a corpus-based contrastive interlanguage analysis, this study reveals the differences of English word *get* used by Chinese EFL learners compared with native speakers in terms of frequency, collocation and semantic prosody.

Introduction

In recent decades, the importance of vocabulary in language teaching and learning has been realized. Vocabulary has rapidly changed in status from a neglected aspect of language learning to an area of growing research and publication. Wilkins writes that without grammar, little can be conveyed, without vocabulary, nothing can be conveyed. McCarthy also points out that "No matter how well the student learns grammar, no matter how successful the sounds of L2 are mastered, without words to express a wider range of meaning, communication in a L2 just cannot happen in any meaningful way"[3]. Mastering the vocabulary of a language does not consist only in gathering the lexical knowledge. Rather, we argue that it also rests on the development of lexical competence.

Theoretical Foundations

Collocation

Hoey defines collocation as "the relationship a lexical item has with items that appear with greater than random probability in its textual context" [1]. That is, words are collocates of each other if, in a given sample of language, they are found together more often than their individual frequencies would predict. Words which stand in such a relationship can be said to predict one another because the presence of one makes the presence of the other more likely than it would otherwise be. Sinclair refers to collocation as "the occurrence of two or more words within a short space of each other in a text" [2], this could logically refer to co-selection between lexical or grammatical items. Stubbs defines collocation as: "a lexical relation between two or more words which have a tendency to co-occur within a few words of each other in running text" [7]. This is the sense in which collocation will be used in the present paper.

Semantic Prosody

Louw points out that semantic prosody is the consistent aura of meaning with which a form is imbued by its collocates. Partington defines semantic prosody as "the spreading of connotational colouring beyond single word boundaries" [4]. Semantic prosody may be broadly classified into three types: positive, neutral, and negative. When the semantic prosody of a word is positive, most, if not all, of its collocates have positive values. When the semantic prosody of a word is neutral, its

collocates include words of both positive and negative values. When the semantic prosody of a word is negative, most, if not all, of its collocates exhibit negative values.

Data Collection and Analysis

Data Description

Chinese Learner English Corpus (CLEC) is one of the largest existent Chinese EFL learner corpora. This is a million-word corpus constructed by Shanghai Jiaotong University and Guangdong Foreign Studies University. Gui and Yang report that error locations and error types in this corpus were tagged using machine-aided manual tagging. The CLEC consists of written samples collected from almost all levels of school learners of English in China, ranging from high school students to university English majors at different skill levels and is divided into five sub-corpora: high school students (ST2), junior and senior non-English majors (ST3 & ST4), and junior and senior English majors (ST5 & ST6), representing Chinese English learners at different stages. The sampling involves a diversified and varied coverage of both student essays written at tests and different assignments completed at various learning situations. The interest of the current thesis is on the English near-synonyms discrimination competence of Chinese College English majors, only two subcorpora of CLEC are used here, that is, the English-major college student subcorpora (ST5 &ST6) of CLEC. The native language corpus, i.e. Freiburg-LOB Corpus of British English (i.e. FLOB) is selected as a reference corpus. This is an update of the Lancaster-Oslo-Bergen corpus (LOB), which is composed of about one million tokens of running text in total. FLOB contains 500 texts of about 2,000 words each.

Data Collection

There are some basic points that we need to spell out from the start. One is the distinction between tokens and types, which applies to any count of the words in a text. The number of tokens is the same as the total number of word forms, which means that individual words occurring more than once in the text are counted each time they are used. On the other hand, the number of types is the total number of different words forms, so that a word which is repeated many times is counted only once. Our attention is restricted to content words, but the problem is that these words come in a variety of forms. The word *get* is selected since it ranks top in frequency.

Data Analysis

Frequencies of Get in CLEC (ST5 & ST6) and FLOB

At a first step in our analysis, the author wants to check whether English majors have a tendency to overuse or underuse *get* in comparison with the native speakers. With the corpora and concordance software, related concordance lines with *get* are extracted. The results are given in Table 1.

Table 1 Frequencies of get in CLEC (ST5&ST6) and FLOB (Per 100,000 words)

Lemma	ST5 (214,510 words)	ST6 (226,106 words)	FLOB (1,007,004 words)
	Token/Per 100,000 words	Token/Per 100,000 words	Token/Per 100,000 words
get	278/129.6	315/139.32	116/11.52

As far as the lemma *get* is concerned, the English majors' use is about five times of that of native speakers. From the above Table 1 of data from ST5, ST6 and FLOB, the conclusion is that English majors overuse *get* to a large extent.

Collocations and Semantic Prosodies of *Get* in FLOB

In this section, collocations and semantic prosody features of *get* will be studied. As to collocations of *get*, the author mainly considers their right collocations with nouns or noun phrases. Therefore, collocations and semantic prosodies of *get* only considered for two colligations: "V+O" and "V+O+ prep". The following tables present the data of collocates and semantic prosodies of get.

Table 2 The collocation and semantic prosody of *ge*t in the FLOB corpus

Types	Noun collocations	Occurrence	Percentage	Semantic prosody
1	money, wage, income, etc.	20	18.5	positive
2	job, chance, etc.	16	14.8	positive
3	aid, help, support, freedom, etc.	16	14.8	positive
4	idea, impression, answer, etc.	38	34.3	neutral
5	degree, education, diploma, etc.	5	4.6	neutral
6	others	15	13	neutral

Table 2 shows that there is a higher proportion of Type 4 with a neutral semantic prosody, constituting 34.3% in the FLOB. Type 1 accounts for 18.5% with a positive semantic prosody in the FLOB. The proportions of Type 2 and Type 3 are the same (14.8%) in the FLOB, showing positive semantic prosody. Type 5 with a neutral semantic prosody is not used very frequently by native speakers, constituting only 4.6% in the FLOB. Therefore, the collocates of *get* have two kinds of semantic prosodies: positive semantic prosody and neutral semantic prosody. Collocates such as *money*, *pound*, *job*, *chance*, *help* are showing positive semantic prosody, and collocates such as *idea*, *answer*, *degree*, *education* are showing neutral semantic prosody.

Table 3 The collocation and semantic prosody of get in ST5&ST6

Types	Noun collocations	Occurrence	Percentage	Semantic prosody
1	job, money, income, salary, etc.	69	26.1	positive
2	degree, grade, knowledge, etc.	52	19.7	neutral
3	punishment,warning, retribution, etc.	12	4.5	negative
4	permission, reward, support, etc.	43	16.3	positive
5	chance, opportunity, answer, idea, etc.	32	12.2	neutral
6	thing, anything, topic, view, etc.	56	21.2	neutral

The above Table 3 shows that the proportion of Type 1 is 26.1% with a positive semantic prosody. Type 2 accounts for 19.7% with a neutral semantic prosody, which is higher than the proportion (4.6%) found in the FLOB. Type 3 constitutes 4.5% showing negative semantic prosody, which is not found in the FLOB. The proportion of Type 4 is 16.3% with a positive semantic prosody. Type 5 and type 6 make up 12.2% and 21.2% respectively in ST5&ST6 with neutral semantic prosody.

Based on the above data discussion, we may conclude that there are some distinctions between English majors and native speakers in the use of *get* in terms of frequency, collocation and semantic prosody. There are mainly three reasons for this.

Firstly, there is a lot of evidence to show that second language vocabulary learning is influenced by first language vocabulary, which is called L1 transfer. *Get* has the same translation "dedao" in Chinese. We cannot neglect the existence of L1, simply because L2 learners are no longer in a linguistic vacuum, which means this is a linguistic system. L1 is already there in their heads. With L1

as the background in their mind, and L2 learner is trying to master a new linguistic system. An L2 learner is somehow pulled by two competing systems, consciously or unconsciously, L1 will come into the picture of SLA process. The impact of L1 transfer on lexical development in L2 is straightforward. Given the presence of the established L1 lexical system, L2 learners, in particular Chinese EFL learners, may tend to rely on L1 system in learning new words in a second language. Because the meanings of a L2 word can be understood through their L1 translation, the learner's language processor or language acquisition device may be less motivated to pay attention to the contextual cues for meaning extraction.

Secondly, English has many words that are similar, and dictionary definitions often characterize such words as identical meaning. The patterns of use and semantic prosodies for these words, however, are often very different. Therefore, insufficiencies in CE dictionaries are another reason for overusing, underusing or misusing some English words by English majors.

Thirdly, during second language learning, especially learning second language in the classroom, the constraint is the poverty of input in terms of both quantity and quality. Chinese EFL learners often lack sufficient and highly contextualized input in the target language in the classroom. This often makes it extremely difficult for Chinese learners to extract and create collocation and semantic prosody specifications about a word and integrate such information into that word.

Implications of Vocabulary Teaching and Learning

Vocabulary ability constitutes various kinds of knowledge one has to possess in order to use a word appropriately. Such definition can be found in Richards [5], who proposes eight "assumptions concerning the nature of lexical competence". Vocabulary ability is seen as much more than knowing the form and meaning of a word. It also means to know, for example, the association between a word and other words, the likelihood a word may occur in a linguistic context, and the limitations imposed on the use of a word according to variations of function and situation. Similarly, Nation considers vocabulary ability as consisting of four dimensions of knowledge form, position, function and meaning.

When we bring the corpus into classroom, we have to teach the student several things. Tognini-Bonelli, E. proposes some methods of teaching vocabulary with corpus. Firstly, some collocates with the node should be noticed. This becomes obvious when the concordance is alphabetized: N-1, N-2, N-3, etc. on the left co-text of the node, N+1, N+2, N+3, etc. on the right co-text of the node. Secondly, within this initial level of co-selection at word level, we have to teach the students to query the patterns further in the light of grammatical categories. The third is at the semantic and functional levels. It is a further stage in abstraction compared with the collocational and colligational levels. Sometimes the frequent collocation of certain words may help the students to identify the node. This is where the identification of the semantic prosody takes place. It is the most important stage and the most useful in classroom activities and it is the final step bringing together the other two stages. A lot of materials can be assembled using the concordances and asking the students to evaluate them in terms of formal collocations and colligations, and regroup them according to certain functional tasks.

References

[1] Hoey, Michael.Patterns of Lexis in Text. Oxford: Oxford University Press, 1991.

[2] Jones, S. & Sinclair, J. M. English lexical collocations. A study in computational linguistics[J].

Cahiers de lexicologien24(1974)15-61.

[3] McCarthy, M. Vocabulary. Oxford: Oxford University Press, 1990.

[4] Partington, A. Patterns and meanings: Using corpora for English language research and teaching. Amsterdam: John Benjamins, 1998.

[5] Richards, J.C. The role of vocabulary teaching[J]. TESOL Quarterly, 1976, 10, 77-89.

[6] Sinclair, J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.

[7] Stubbs, M. Two quantitative methods of studying phraseology in English[J]. International Journal of Corpus Linguistics, 7(2002)215-244.

[8] Wilkins, D. A. Vocabulary Linguistics in Language Teaching. London: Edward Arnold Ltd, 1972.