HOW CORPUS-BASED APPROACH CAN CONTRIBUTE TO THE STUDY OF SEEMINGLY SYNONYMOUS WORDS

Lu Lu
Capital Normal University

Abstract
In the English language, there are many words which are considered synonymous in that dictionaries and thesauruses often define them in identical meanings. Under such circumstances, EFL learners are confronted with a big difficulty in acquiring these words of similar meanings. They often regard synonymous words as equivalents and use them alternately, which sometimes result in ambiguity or awkwardness of their language. However, with the advent of corpora and concordancing programs, a new way of investigating and learning English synonymous words has emerged. This paper explores collocational patterns and semantic prosody of two seemingly synonymous verbs, gain and obtain, in two native speaker corpora, the Brown Corpus and the LOB Corpus, in order to show how these “equivalencies” can be misleading because synonymous words are typically used in different ways. The study also examines EFL learners’ behavior in using synonymous words by investigating the Chinese Learner English Corpus (CLEC). From the results of the study, it can be seen that Chinese EFL learners have not yet acquired a full understanding of the usage of gain and obtain with respect to collocation and semantic prosody. Their undersuses and overuses of some grammatical forms and lexical patterns of the two target synonymous words might be influenced either by their mother tongue or by the registers of their English writing or by both. The paper finally discusses the implications of corpus-based study in vocabulary teaching and learning in general.

Key words
corpus-based approach, synonymous words, collocation, collocation, semantic prosody

1 Introduction

Lexical competence has recently been identified as one of the most significant predictors to general language ability, but it is also admitted by most learners to be a big challenge in language learning. With the advent of corpora and concordancing programs, a new view of teaching and learning vocabulary has emerged. Compared with traditional vocabulary learning tools, dictionaries, corpus-assisted vocabulary learning approach has some advantages. On the one hand, corpus can offer authentic language data and thus create an authentic language learning environment. On the other hand, unlike dictionaries which mainly offer sentence-based examples of word entries, corpus provides examples, concordances of target words in a context either at the sentence level or beyond the sentence level, such as displaying the complete text where the target word appears. In addition, concordance lines are typically displayed in a KWDI keyword-in-context way, which can easily arouse learners’ awareness of the grammatical patterns and the usage of the target words. In this sense, learners may learn not only word meanings but also word usage in specific contexts.

At present, one big lexical problem existing in EFL learners’ writing is that they often pick up words randomly from their “word bank” provided that the word meanings are in accord with those in their mother tongue. Some Chinese EFL learners use both big and large to refer to a quantity or amount of
something [for example] a big number of or a large number of. From the learners' viewpoints since big and large are synonymous words and both of them mean 'a great number' in Chinese [da shu mu] can be translated into either a big number of or a large number of in English. The present study adopts a corpus-based approach to explore the usage of synonymous words in terms of coligation and semantic prosody, showing how these "synonymous" words are used by native speakers. In addition, the paper discusses the way that EFL learners use synonymous words by investigating the Chinese Learner English Corpus (CLEC). It also provides some implications of corpus-based study in vocabulary teaching and learning.

2. Coligation, collocation, and semantic prosody

Nowadays two concepts are often referred to in vocabulary learning, the scope of vocabulary knowledge and the depth of vocabulary knowledge. Pu 2003 [139]. The former notion refers to the number of words one has known, while the latter one deals with the extent to which one has grasped the usage of words. Pu 2003 [139]. Apparently, the analysis of synonymous words is to help students with the depth of vocabulary knowledge. By the depth of vocabulary knowledge, there are two components: 1) mastery of the core grammatical patterns of words, 2) mastery of the typical combinations of words. Pu 2003 [139]. These two concepts are explained in terms of "coligation" and "collocation" in this paper.

The notion of coligation is first discussed by Firth, referring to "the inner relation of grammatical categories in syntactic structure". Firth 1957 [99]. Used to describe the grammatical patterns of words, coligation is an abstraction of systematic grammatical relations between grammatical categories. For example, the word order is often used in the patterns of "V N" and "V that." Here "V" stands for a verb, and "N" refers to a noun, a nominal phrase or a pronoun. "That" represents a that clause. The patterns of "V N" can accept a proposal and "V that" can accept that her leaving is a possibility. These are regarded as two typical coligations of the verb "accept".

The other component of the depth of vocabulary knowledge, collocation, has been studied for at least fifty years. Once again, Firth was one of the pioneers who defined "collocation." According to Firth, 1957 [99], "collocations are actual words in habitual company." In this paper, "collocation" is considered as the habitual co-occurrence of words, especially nouns, verbs, adjectives, and adverbs within certain coligations. Pu 2004 [18]. As far as the verb "accept" is concerned, in its grammatical pattern, "V N", it habitually co-occurs with such nouns as offer, view, and proposal, and these nouns are considered as significant collocates of accept in its coligation of "V N". From the above discussion, one can see that collocation is closely related to coligation. In fact, collocation refers to the habitual co-occurrence of words within certain coligations.

Shifting from forms to meanings, some corpus-based studies demonstrate that within certain grammatical forms, there are always semantic relations between words and their collocates, and among the collocates themselves. Stubbs 2002 [225]. The semantic atmosphere arising from the interaction between words and their collocates and among the collocates themselves is referred to as "semantic prosody." Louw 1993 [157]. According to Louw 2000 [57], semantic prosody is "a form of meaning which is established through the proximity of a consistent series of collocates whose primary function is to express speakers' or writers' attitude or evaluation." In general, semantic prosody can be classified into three kinds: positive, neutral, and negative. It might be argued that the semantic prosody that belongs to a word is the result of the interplay between the word and its collocates in specific contexts. For example, the research of Stubbs 1995 shows that the verb "cause" has a strongly negative semantic prosody since more than 90% of its collocates connote negative meanings, such as cause an accident, cause a crisis, cause delay, and so forth.

So far, one can see that coligation, collocation, and semantic prosody demonstrate the grammatical forms' lexical patterns and affective meanings of a word. Yet what is important to note here is that before the advent of corpora and concordancer, these phenomena have hardly been noticed. They are right under our noses, but it is corpora and concordancer that bring them to our attention.
3. Corpus data and methodology of the present study

3.1 Corpus data

Since the objective of the research is to examine the usage of seemingly synonymous words as well as to investigate EFL learners' uses of them, the foremost task is to find enough representative language data to fulfill this purpose. Three principal corpora were employed in this study: the Brown Corpus, the LOB Corpus, and the Chinese Learner English Corpus (CLEC). Each corpus contains approximately one million words. The Brown Corpus and the LOB Corpus are famous native speaker corpora while the CLEC is a learner corpus. The subcorpora of the CLEC mainly consist of the English writing of the Chinese high school students and college students in the forms of test composition, class work and homework. The concordancer used is WordSmith Tools 4.0.

3.2 The words to be investigated

In consideration of the research objective mentioned in 3.1, it is neither necessary nor possible to cover all English synonymous words in this study. A sample has to be selected. Here we choose synonymous verbs gain and obtain as research samples. In the Longman Dictionary of Contemporary English, the definitions of the two verbs gain and obtain are almost identical. The core meaning of "gain" is "to obtain or achieve something you want or need" and "obtain" means "to get something that you want especially through your own effort." The Chinese meaning of the two words is the same. "huò de" In fact, as far as the definitions and examples provided by the dictionary are concerned, it is hard to discriminate between gain and obtain when they are used as verbs. Accordingly, gain and obtain are the typical English synonymous words and the right research samples for this study.

3.3 Methodology

After the sampled synonymous words have been determined, the concordancer WordSmith Tools 4.0 was used to make concordances of GAIN and OBTAIN in the Brown Corpus, the LOB Corpus and the CLEC so as to find out a true reflection of native speaker and EFL learners' behavior on seemingly synonymous words. Since this study only examined GAIN and OBTAIN used as verbs, those improper concordance lines such as the ones with GAIN used as nouns were crossed out. Then the acceptable concordances were scrutinized and classified by human researcher in terms of grammatical patterns in order to establish the collocations of GAIN and OBTAIN in the three corpora respectively. The next step was to analyze the collocation and semantic prosody of the target words. Being common English verbs, GAIN and OBTAIN have such a large number of concordances that it indicates a big difficulty for human researcher to examine the concordance lines one by one. In addition, the aim of this study is not to attain the exact figures of each target word's collocates but to seek for some tendencies. There is no need to examine so many concordances to see the tendencies. Therefore, an upper limit was set so that only 60 concordance lines of GAIN and OBTAIN in each corpus were randomly chosen for a more detailed study — most importantly, the target words' collocates and their semantic prosodies.

4. Results

This section focuses on the salient features of native speakers' and Chinese learners' behavior on GAIN and OBTAIN and explores the underlying reasons for the features. In the first part, the two synonymous words are examined in the native speaker corpora and in the second part they are investigated in the learner corpus.

4.1 The usage of GAIN and OBTAIN in the native speaker corpora

4.1.1 Collocation

In the Brown Corpus and the LOB Corpus the grammatical patterns of GAIN can be summed up as "V from A" and "V N" and "V N from A." As for OBTAIN, it is basically used in two collocations "V N" and "V N from A." Due to our limited space, some concordances of GAIN and OBTAIN randomly chosen.

1. Here capital letters are used to represent different verb forms of gain and obtain including gain, gains, gained, gaining, obtain, obtains, obtained and obtaining.
from the Brown Corpus are displayed as follows:

Concordances of **GAIN** in the Brown Corpus
g to consider procedures by which they may gain more insight into themselves and better under
day [he told his parishioners] when "I did not gain from you far more than I ever gave to you"
mises" One is that they were established [or gained eminence] under pressure provided by the
mined [which sped the process of gaining equality] or perhaps hindered it [only his
The result says Jolliffe ["The child gains the feeling that food is the purpose of life""]

Concordances of **OBTAIN** in the Brown Corpus
ce height is [if] and h can be eliminated to obtain an equation for the total potential energy of
chosen a certain vocational program have obtained a job as a consequence of the training [the
derstanding] and evidently [from the quarrel] obtained an unfavorable impression of Morgan [i
mised in mortality if they ate of the tree [he obtained mortality instead] The wages of sin is de

2 Collocation and semantic prosody

As stated earlier [collocation is the lexical realization of the constituent elements] in a coligation [Pu [2003 [123 [ The collocational behavior and the semantic prosody of **GAIN** and **OBTAIN** are only
considered for two coligations [" V N" and " V N from " On the basis of the 60 randomly chosen
concordance lines of each word in the corpus the following four tables present the data of collocates and
semantic prosodies of **GAIN** and **OBTAIN** in the Brown Corpus and the LOB Corpus]

### Table 1 The collocation and semantic prosody of **GAIN** in the Brown Corpus

<table>
<thead>
<tr>
<th>Noun collocates</th>
<th>Semantic categories</th>
<th>Occurrence</th>
<th>Percentage</th>
<th>Semantic prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>importance [support] success</td>
<td>30</td>
<td>50</td>
<td>positive</td>
</tr>
<tr>
<td>2</td>
<td>understanding [experience]</td>
<td>13</td>
<td>21</td>
<td>positive</td>
</tr>
<tr>
<td>3</td>
<td>other meanings or ambiguous meanings</td>
<td>12</td>
<td>20</td>
<td>neutral</td>
</tr>
</tbody>
</table>

### Table 2 The collocation and semantic prosody of **GAIN** in the LOB Corpus

<table>
<thead>
<tr>
<th>Noun collocates</th>
<th>Semantic categories</th>
<th>Occurrence</th>
<th>Percentage</th>
<th>Semantic prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>importance [support] success</td>
<td>25</td>
<td>41</td>
<td>positive</td>
</tr>
<tr>
<td>2</td>
<td>understanding [experience]</td>
<td>10</td>
<td>17</td>
<td>positive</td>
</tr>
<tr>
<td>3</td>
<td>other meanings or ambiguous meanings</td>
<td>15</td>
<td>25</td>
<td>neutral</td>
</tr>
<tr>
<td>4</td>
<td>unfavorable meanings</td>
<td>1</td>
<td>1</td>
<td>negative</td>
</tr>
</tbody>
</table>

### Table 3 The collocation and semantic prosody **OBTAIN** in the Brown Corpus

<table>
<thead>
<tr>
<th>Noun collocates</th>
<th>Semantic categories</th>
<th>Occurrence</th>
<th>Percentage</th>
<th>Semantic prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>importance [support] success</td>
<td>11</td>
<td>18</td>
<td>positive</td>
</tr>
<tr>
<td>2</td>
<td>understanding [experience]</td>
<td>5</td>
<td>8</td>
<td>positive</td>
</tr>
<tr>
<td>3</td>
<td>statistical data or related information</td>
<td>11</td>
<td>18</td>
<td>neutral</td>
</tr>
<tr>
<td>4</td>
<td>other meanings or ambiguous meanings</td>
<td>27</td>
<td>45</td>
<td>neutral</td>
</tr>
<tr>
<td>5</td>
<td>reduction death conviction</td>
<td>6</td>
<td>10</td>
<td>negative</td>
</tr>
</tbody>
</table>
Table 4: The collocation and semantic prosody of OBTAIN in the LOB Corpus

<table>
<thead>
<tr>
<th>Noun collocates</th>
<th>Semantic categories</th>
<th>Occurrence</th>
<th>Percentage</th>
<th>Semantic prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td></td>
<td>4</td>
<td>6.97%</td>
<td>Positive</td>
</tr>
<tr>
<td>Support</td>
<td></td>
<td>3</td>
<td>5.2%</td>
<td>Positive</td>
</tr>
<tr>
<td>Success</td>
<td></td>
<td>29</td>
<td>48.3%</td>
<td>Neutral</td>
</tr>
<tr>
<td>Other meanings</td>
<td></td>
<td>24</td>
<td>40.7%</td>
<td>Neutral</td>
</tr>
<tr>
<td>or ambiguous</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>meanings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the above tables one can learn some characteristics of GAIN and OBTAIN in native speaker corpora with respect to collocation and semantic prosody.

1. GAIN has a strongly positive semantic prosody in both corpora with 71% of the noun collocates in the Brown Corpus and 59% of the noun collocates in the LOB Corpus showing a “good” meaning. The examples of the noun collocates in its colligations of “V N” and “V N from” include success, popular acclaim, confidence, liberty, etc. Except for one “bad” noun collocate found in the LOB Corpus, there are no other negative noun collocates of GAIN appearing in either of the corpora. Indeed, a bank of noun collocates with positive senses are bound to the target verb GAIN and thus, GAIN seems to show a favorable semantic prosody in native speaker corpora.

2. Table 3 and 4 demonstrate that what people are going to obtain might be something pleasant such as success and insight, something neutral such as evidence and product, or something unpleasant such as reduction and conviction. In the Brown Corpus, there is a neutral and negative noun collocates of OBTAIN constitute 26% and 36% of all the collocates respectively. Although no negative noun collocates are found in the LOB Corpus, there are only 11% of the noun collocates showing good senses while 88% indicate neutral meanings. The statistical data clearly show that OBTAIN has a neutral semantic prosody in the two native speaker corpora.

3. It is worth noting that OBTAIN has a strong tendency towards co-occurring with the nouns indicating statistical data. In fact, 58% and 48% of the noun collocates in the Brown Corpus and the LOB Corpus respectively are such words as equation, statistics, spectra, measurement, etc. This phenomenon points to a fact that OBTAIN might often be used in academic articles or research papers and its right noun collocates are associated with statistical data or results. In contrast, GAIN might be barely used in such contexts.

4. The usage of GAIN and OBTAIN in the learner corpus

This part focuses on the characteristics of learner use of GAIN and OBTAIN in the CLEC. The methodology adopted in this section is the same as the one employed to study the two target words in the native speaker corpora. The research findings are as follows:

1. **Colligation**

   The colligations of GAIN and OBTAIN in the CLEC are identical in “V N” and “V N from”. None of the concordance lines of GAIN is found in such colligation as “V from”. All which is different from the native speaker corpora as far as the colligation of GAIN is concerned. The following are some concordance lines of GAIN and OBTAIN picked up randomly from the CLEC.

   Concordances of GAIN in the Chinese Learner English Corpus:

   d3 [I we should devote ourselves to it and gain some success by our working. In society today etc.] Thus we can learn enough from society and gain much experience and get ready for going to

   ople from the lower classes must learn how to gain equal economic status first through every da

   nd learn to use np7 [computer in order to gain more knowledge and keep up with the so

   mostly men and if a women np3 [I wants to gain success in society she must sacrifice a lot of
Concordances of **OBTAIN** in the Chinese Learner English Corpus
and universities must make sure that students obtain all the necessary information before they
as of people[10] to obtain support and to cooperate[10] Thus[10] when we g
fect sweat and 1 percent gift". That means obtaining success requires industry[12] If you want to s

t 2 Collocation and semantic prosody

The following two tables can be drawn up to sum up learners' behavior on **GAIN** and **OBTAIN** with
respect to collocation and semantic prosody

**Table 5** The collocation and semantic prosody of **GAIN** in the CLEC

<table>
<thead>
<tr>
<th>Noun collocates</th>
<th>Semantic categories</th>
<th>Occurrence</th>
<th>Percentage</th>
<th>Semantic prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> importance</td>
<td>support success etc</td>
<td>33</td>
<td>55%</td>
<td>positive</td>
</tr>
<tr>
<td><strong>2</strong> understandi</td>
<td>experience knowledge etc</td>
<td>17</td>
<td>28%</td>
<td>positive</td>
</tr>
<tr>
<td><strong>3</strong> other meanings or ambiguous meanings</td>
<td>10</td>
<td>16%</td>
<td>neutral</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6** The collocation and semantic prosody of **OBTAIN** in the CLEC

<table>
<thead>
<tr>
<th>Noun collocates</th>
<th>Semantic categories</th>
<th>Occurrence</th>
<th>Percentage</th>
<th>Semantic prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> importance</td>
<td>support success etc</td>
<td>25</td>
<td>41%</td>
<td>positive</td>
</tr>
<tr>
<td><strong>2</strong> understandi</td>
<td>experience knowledge etc</td>
<td>22</td>
<td>36%</td>
<td>positive</td>
</tr>
<tr>
<td><strong>3</strong> other meanings or ambiguous meanings</td>
<td>13</td>
<td>21%</td>
<td>neutral</td>
<td></td>
</tr>
</tbody>
</table>

From the above tables, one can see that learners use **GAIN** and **OBTAIN** almost in an identical way.
The right noun collocates of the two verbs and the semantic atmosphere they create in the contexts are rather similar.
Indeed 83% and 78% of the noun collocates of **GAIN** and **OBTAIN** in the CLEC show favorable senses.
In other words, both **GAIN** and **OBTAIN** demonstrate strongly positive semantic prosody in the CLEC.

Through investigating the reasons why learners use **GAIN** and **OBTAIN** in the same way, one can find that the meanings conveyed in the two words are rather similar in learners' other tongue Chinese.
**GAIN** and **OBTAIN** are both equivalent to the Chinese phrase huo de. On the one hand unlike
its English translations of **GAIN** and **OBTAIN**, huo de does not indicate an unfavorable affective
meaning in Chinese. There are a multitude of Chinese phrases concerned with huo de such as huo de
chong gong, huo de zhi shi, huo de zhi shi, huo de zhi shi, huo de zhi shi, huo de zhi shi,
**conviction**. Accordingly, influenced by their mother tongue, Chinese EFL learners rarely use negative
noun collocates with either **GAIN** or **OBTAIN**.

Similarly, due to the lack of such Chinese expressions as huo de gong shi, huo de jie guo, or huo de
jie guo, some of the learners employ English phrases like obtain formulas or obtain results. However, we do find learners use the phrase of get results. This is because the Chinese equivalent of GET is de dao and de dao jie guo get results is fairly
common in Chinese. When learners translate "de dao" into English straightforwardly, their first choice
would possibly be a simpler word GET and thus get results. Yet, there might be other reasons for no such expressions as obtain formulas or obtain results appearing in the CLEC.

As a matter of fact, most of the compositions collected in the CLEC deal with campus life, personal views or social phenomena, and very few of them are academic or research papers. Lack of the register of academic writing might be the other reason for no such expressions as obtain data or
obtain results in the CLEC.

To sum up in terms of coligation, collocation and semantic prosody, native speakers and Chinese EFL learners use two seemingly synonymous words GAIN and OBTAIN in quite different ways. As for native speakers, the verb GAIN is often used in three grammatical patterns “V from [n]” “V N” and “V N from” and the verb OBTAIN is used in two patterns “V N” and “V N from” [n]. However, Chinese EFL learners often use these two verbs in the same coligation “V N” and “V N from” [n]. It reveals that Chinese EFL learners have not acquired the coligation of “V from [n]” which is yet often used by native speakers. With respect to collocation and semantic prosody, native speakers treat GAIN and OBTAIN differently. Most right noun collocates of GAIN show favorable senses which results in the generally positive semantic prosody of GAIN in the two native speaker corpora. However, the collocates of OBTAIN are fairly complicated. With diverse senses: positive, neutral and negative in native speaker corpora, OBTAIN creates a generally neutral semantic prosody. Compared with native speakers, Chinese EFL learners consider these two verbs almost identical. Most collocates of both GAIN and OBTAIN in the CLEC show positive meanings. In addition, in the native speaker corpora, it is found that the verb OBTAIN is often accompanied by the nouns with the senses of statistical data and results whereas there are no such findings in the learner corpus.

5 Conclusion

This study has explored the collicative patterns, collocational behavior and semantic prosody of two seemingly synonymous verbs GAIN and OBTAIN in a corpus-based approach. It has also investigated some Chinese EFL learners’ uses of these two verbs by adopting the Chinese Learner English Corpus (CLEC). From the results of this case study, it can be seen that the collocation, collocation and semantic prosody of words are observable when concordance lines are analyzed in detail. As they may have different collicative patterns, collocational behavior and semantic prosody, seemingly synonymous words are normally not interchangeable under some circumstances. Based on this study, it is also becoming increasingly noticed that lexis and syntax are selected and form and meaning are closely related. Both the grammatical properties and lexical features of words ought to be taken into consideration when teachers teach words, students learn words and one distinguishes synonymous words. Therefore, in the vocabulary class, especially in the class of advanced EFL learners, word study should involve not only word meanings but also word usage such as the core grammatical patterns and the lexical features of words. It is apparent that language corpora can enhance the quality of vocabulary teaching and learning in these two aspects. With adequate and authentic language material provided by the corpora, learners will be exposed to the authentic language environment and find out the proper and idiomatic usage of words from the contexts. In this way, when writing compositions, EFL learners may choose words carefully rather than casually putting down the words which come to their pens.

References


20