A Corpus-Based Study on the Employment of Verb Keep between St3 and St4 in CLEC and Brown

Ning Zhang

Office of Basic Teaching Department, Weihai Campus of Shandong Jiaotong University, Weihai 264209, Shandong, China

Correspondence should be addressed to Ning Zhang; 201147@sdjtu.edu.cn

Received 22 June 2022; Revised 3 August 2022; Accepted 23 August 2022; Published 19 September 2022

Academic Editor: Yanyi Rao

Copyright © 2022 Ning Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The present paper focuses on the frequently used verb keep by Chinese non-English majors and native speakers based on three corpora, namely, Brown, St3, and St4 in CLEC. Brown stands for a corpus of native speakers, whereas St3 and St4 in CLEC stand for Chinese non-English majors who take CET-4 and CET-6 tests, respectively. This paper employs the contrastive analysis theory. The thesis investigates the commonly used keep senses by Chinese learners in CLEC compared with those senses in Brown by native speakers. This current research also employs a large amount of computer instruments like SPSS, AntConc, Microsoft Excel, and Wordsmith. Furthermore, the research aims to find out whether Chinese learners’ performance on verb keep gets improvement with the promotion of their English proficiency. The results of the paper show that Chinese non-English majors do not have a good command of the frequently used senses of keep by native speakers. There exists overuse and underuse of some senses. Nevertheless, with the improvement of English proficiency, Chinese learners’ employment of keep is becoming much closer to that of the native speakers. According to the research, some pedagogical reflections to improve English teaching and learning of verbs like keep are also discussed in the paper. Blended teaching can be applied before class, and microlectures can be provided online. Keep phrasal verbs, idioms, and collocations can be included in microlectures, while class teachers and students can collaboratively draw mind maps of verb keep. After class, more assignments can be provided online, such as multiple choice, blank filling, translation, and so on.

1. Introduction

In the acquisition of second language, vocabulary plays an unquestionably important part. The predicate is indispensable in making up a sentence. Meanwhile, the predicate of a sentence must be a verb. Jiang and Zhong [1] stated that different types of verbs constitute different sentence patterns. The most changeable and active of all parts of speech are exactly verbs in English. Different tenses depend mainly on different kinds of verbs. Therefore, verbs bring great difficulties to language learning, especially with regard to high-frequency verbs.

“Keep” is chosen as a representative of verb in this paper. “Keep” is considered to be one of the most usually used words in BNC English, as is shown in Longman Dictionary of Contemporary English [2].
1.2. Research Significance. The significance of this thesis exists in that it complements previous studies on verbs. This makes a new contribution to the study of verb *keep*. These two aspects are discussed in detail in section 1.3.

1.3. Theoretical Background. The contrastive analysis is a major theory used in linguistic analysis. The contrastive analysis stands for the comparison between the output of native speakers and the output of second-language learners; however, the components that are distinct from their mother tongue are quite hard for them.

A corpus-based study of “Keep” by Wang Ying [4] is a previous study of *keep*. Wang Ying’s research is based on corpora LOB and SLESSON. LOB corpus is a British English corpus modeled on the proportion of Brown corpus. SLESSON is a corpus based on a set of English textbooks used in high school in China by He Anping. As a result, Wang’s study mainly focuses on the study of textbooks instead of the study of authentic use by Chinese learners. Only twelve senses and three patterns of sense 1 are investigated in Wang Ying’s paper.

The author concentrates on the authentic Chinese learners’ employment of *keep* instead of textbooks’ employment. Some new pedagogical reflections, such as blending class, microlecture, and mind map, are provided in EFL learning. In addition, the fourteen senses are more specific than the twelve senses by Wang Ying.

2. Research Methodology

Three corpora are involved in the research. Moreover, this research is carried out with the help of computer instruments, such as AntConc, SPSS, Microsoft Excel, and Wordsmith.

2.1. Corpora. The corpora involved in this study are Brown and CLEC. The Chinese Learner English Corpus (CLEC) [5] is the first authoritative learner corpus in China with a total of 1.1 million words. CLEC’s data are general genre English compositions of Chinese learners at different levels. It consists of five subcorpora, in which St3 represents the compositions of Chinese non-English majors in CET-4 and St4 represents the compositions of Chinese non-English majors in CET-6. St3 and St4 are test compositions written by CET-4 and CET-6 testees. Because they are authentic language users, St3 and St4 are appropriate for studying Chinese learners’ second-language development patterns.

Brown corpus is a modern native speakers’ universal corpus that is also computer readable and quite authoritative. The corpus is composed of 1 million words of texts in American English in 1961. Therefore, the size of Brown is compatible with CLEC.

2.2. Instruments. The data needed for this study are the sentences in which the verb *keep* appears. In order that the data needed in this paper can be available, the application of computer software is much more convenient and efficient. AntConc, SPSS, Microsoft Excel, and Wordsmith are used in this study [6]. AntConc is used to extract the required tested words. The chi-square value is tested by SPSS software package. Microsoft Excel is employed to calculate and graphically display results. This thesis employs Wordsmith to test the type-token ratio (TTR) of the three corpora [7].

3. Data Collection and Research Procedure

The fourteen *keep* meanings are evidently shown in *Oxford Advanced Learner’s English-Chinese Dictionary* [8]. According to this dictionary, all *keep* in the three corpora are checked to determine what meaning they belong to, and each meaning’s frequency and percentage are calculated. If necessary, the results are shown in tables and figures. The meaning distributions among the three corpora, between native speakers and Chinese learners, and between St3 and St4 learners are compared. The chi-square test is conducted in order to discover whether there is a significant difference in the use of *keep*. This paper analyzes the characteristics of Chinese learners’ use of *keep* in detail.

The detailed processing of data collection, research findings, and explanations for the outcome are listed below. The author compared the employment of *keep* in Brown and in CLEC. In this thesis, quantitative and qualitative analyses together with the possible reasons of the results are provided in detail.

3.1. *Keep* Occurrences in the Three Corpora. Keep frequencies in Brown and in St3 and St4 in CLEC are computed to discuss the use of verb *keep* with regard to overuse and underuse. Table 1 [9] shows the results.

In Table 1, it is quite evident that *keep* frequency in St3 and St4 is 206 and 124, respectively, which are much lower than the *keep* frequency in Brown. Nevertheless, the employment percentage of *keep* in the three corpora shows that the percentage of the employment of *keep* in St3 is remarkably higher than the percentage in Brown. On the contrary, the percentage of the employment of *keep* in St4 is significantly lower than the percentage in St3 and at the same time narrowly higher than the percentage in Brown. The remarkable difference in overall *keep* frequency across the three corpora can be attributed to the difference in the size of the three corpora. As a consequence, we utilize the chi-square test to research whether there is a significant difference between Chinese learner’s second-language corpora and mother-tongue corpus. In Table 2 [9], *keep* distributions across the three corpora are provided.

According to Zhou Shijie [6], the critical value of the chi-square test is 3.84. That means one degree of freedom at 5% level. In Table 2, we can see the chi-square value is up to 66.200 between Brown and St3. The chi-square value is remarkably larger than 3.84. As a consequence, it is safe to arrive at the conclusion that the difference in *keep* frequency
in terms of Brown and St3 is significant. St3 English learners employed the verb *keep* much more frequently than English native speakers. The most evident reason for this result may be that the range of vocabulary grasped by St3 learners is quite limited. Consequently, St3 learners are liable to use *keep* much more frequently when they cannot figure out other appropriate substitutes for *keep*.

Similarly, in Table 3, the chi-square test between St4 and Brown is carried out using exactly the same procedure as Table 2.

In Table 3, it is quite apparent that with regard to St4 and Brown, the chi-square value is 1.768. The number is much lower than the critical value of the chi-square test 3.84. Consequently, it is safe to conclude that the difference is significant among Chinese EFL learners between St3 and St4. It is distinct that compared with St4 learners, St3 learners are inclined to employ *keep* much more frequently.

As we can see from the above three tables, namely, Tables 2–4, Chinese EFL learners have a tendency to employ verb *keep* much more frequently compared with English native speakers, especially when it comes to the St3 corpus.

Table 1: *Keep* in the Brown, St3, and St4.

<table>
<thead>
<tr>
<th>Corpora</th>
<th>Frequencies</th>
<th>Corpora words</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>518</td>
<td>1015537</td>
<td>0.051</td>
</tr>
<tr>
<td>St3</td>
<td>206</td>
<td>209043</td>
<td>0.0985</td>
</tr>
<tr>
<td>St4</td>
<td>124</td>
<td>212855</td>
<td>0.0583</td>
</tr>
</tbody>
</table>

According to the same procedure as Tables 2 and 3, the chi-square test with regard to Chinese learners is conducted in the following Table 4.

In Table 4, it is apparent that the outcome of the chi-square test is 21.869. The number is much larger than the critical value of the chi-square test 3.84. Consequently, it is safe to conclude that the difference is significant among Chinese EFL learners between St3 and St4.

As we can see from the above three tables, namely, Tables 2–4, Chinese EFL learners have a tendency to employ verb *keep* much more frequently compared with English native speakers, especially when it comes to the St3 corpus.
exhibits that the range of vocabulary used in St4 is a little less limited compared with St3 non-English majors. As a result, we can arrive at the conclusion that keep takes up a large ratio in overall Chinese learners’ English output. Consequently, it results in the overuse of the high-frequency verbs like keep in Chinese learners’ corpora.

To summarize the above analysis, Chinese learners are liable to employ verb keep too frequently compared with mother-tongue learners as a result of their limited range of vocabulary. Nevertheless, in the meantime, with the promotion of non-English majors’ English proficiency, St4 learner’s employment of keep is apt to be narrowly nearer to mother-tongue speakers than St3 learners.

3.2. Distribution of Different Keep Senses in the Three Corpora.

This section aims to investigate whether there are differences in the use of different meanings of keep across the three corpora and on the employment of verb keep if higher-proficiency English learners that are represented by St4 non-English majors are apt to become more native-like than lower proficiency St3 non-English majors. Hornby [8] supplied fourteen senses of keep in his dictionary.

(1) continue to be in the specified condition or position; remain or stay; ~ (on) doing sth; continue to move in the specified direction.

E.g., She has the ability to keep calm in an emergency.

(2) cause sb/sth to remain in the specified condition or position.

E.g., These gloves will keep your hands warm.

(3) detain or delay (sb); ~ sb from sth/doing sth.

E.g., You’re an hour late; what kept you?

(4) continue to have (sth); retain; look after sth for sb; retain sth; have (sth) in a particular place; store; retain (sth) for future use or reference.

E.g., Please keep me a place in the queue.

(5) own and manage (a shop, restaurant, etc.)

E.g., Her father kept a grocer’s shop for a number of years.

(6) own and look after (animals) for one’s use or enjoyment

E.g., She keeps dogs in her apartment.

(7) have (sth) regularly on sale or in stock.

E.g., “Do you sell Turkish cigarettes?” “I’m sorry, we do not keep them.”

(8) not reveal (a secret)

E.g., Can you keep a secret?

(9) (of food) in good condition.

E.g., “Do you sell Turkish cigarettes?” “I’m sorry, we do not keep them.”

(10) be in the specified state of health.

E.g., “How are you keeping?” “I’m keeping well, thanks.”

(11) make written entries in (sth); write down (sth) as a record.

E.g., She kept a diary for over twenty years.

(12) provide what is necessary for (sb); support (sb) financially.

E.g., He scarcely earns enough to keep himself and his family.

(13) guard or protect (sth); protect sb from (sth).

E.g., May the Lord bless you and keep you.

(14) be faithful to (sth); respect or observe.

E.g., I have an appointment to keep at 3 pm.

According to the 14 meanings listed above, all verb keep occurrences appearing in Brown, St3, and St4 are categorized and calculated, respectively. Then, the author calculates the percentage of each meaning in the three corpora, respectively. The percentages of these fourteen senses distributed in the three corpora are shown in Table 6. The order of the fourteen senses is listed in accordance with the sense frequencies from high to low appeared in Brown [9].

According to Table 6, both Chinese non-English major learners and English native speakers have a tendency to use the first four senses more frequently. It is quite apparent that when it comes to the rest ten senses, Chinese non-English major learners seldom use them. Figure 1 is provided according to Table 6 to offer a much more vivid presentation by means of Microsoft Excel.

As is quite evident in Figure 1 that Chinese non-English major learners and English native speakers are different

Table 5: TTR across the three corpora.

<table>
<thead>
<tr>
<th>Corpora</th>
<th>St3</th>
<th>St4</th>
<th>Brown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>7757</td>
<td>8648</td>
<td>42579</td>
</tr>
<tr>
<td>Token</td>
<td>232541</td>
<td>241969</td>
<td>1015537</td>
</tr>
<tr>
<td>TTR</td>
<td>3.34</td>
<td>3.57</td>
<td>4.19</td>
</tr>
</tbody>
</table>

Table 4: Outcome of keep distribution’s chi-square test between St3 and St4.

<table>
<thead>
<tr>
<th>Value</th>
<th>Df</th>
<th>Asymp. sig. (2-sided)</th>
<th>Exact sig. (2-sided)</th>
<th>Exact sig. (1-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>21.869</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Continuity correction</td>
<td>21.357</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Likelihood ratio</td>
<td>22.082</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Fisher’s exact test</td>
<td>21.869</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Linear-by-linear association</td>
<td>422228</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Computed only for a 2 × 2 table, b. 0 cells (0%) have expected count less than 5. The minimum expected count is 163.54.
from one another with regard to the proportion of each verb keep sense. Compared with English native speakers, Chinese non-English major learners’ use of keep is excessive with regard to sense 1 and sense 4. On the contrary, Chinese non-English major learners are inclined to underuse sense 2 and sense 3 significantly, especially when it comes to sense 2. According to Table 6 and Figure 1, the following two subsections offer a detailed analysis of the data and results.

### 3.2.1. Respective Analysis of the Data across the Three Corpora

At first glance, compared with the first four senses of keep, the other ten senses account for merely a small proportion of all the senses among the three corpora. Among the last ten senses, two outstanding senses that are sense 11 and sense 14 merely appear in Brown. In the last ten senses, Chinese learners only use three senses once, which are sense 8, sense 12, and sense 14. With respect to sense 5, nobody in the three corpora has ever used it.

St3 non-English major learners only use five senses of the verb keep, which are sense 2, sense 1, sense 4, sense 3, and sense 8 ranging according to the use frequency. Sense 8 of keep only appears once in St3. The percentage of sense 1 in St3 is remarkably higher than that of Brown, which is nearly triple of Brown. However, the percentage of sense 4 in St3 is slightly higher than that of Brown, which is about 1.5% more than Brown. The percentage of sense 2 in St3 is significantly lower than that of Brown, which is about 30% lower compared with Brown. The percentages of sense 3 in St3 and St4 are slightly lower than that of Brown.

The top four senses most commonly used by St3 learners are sense 1, sense 2, sense 4, and sense 3, which account for 99% of all keep occurrences in St3. From the data, St3 learners seem to use these four senses more frequently than the other ten senses, especially sense 1. Similarly, these four senses also rank the top four in St4 and Brown, but in quite different order. The order in Brown is sense 2, sense 1, sense 4, and sense 3 in terms of frequency. The most prominent sense in St3 is sense 1, the percentage of which is almost three times of Brown.

Although the overall verb keep percentage in St4, which is 0.0583, is not much different from that in Brown, which is 0.51, as is shown above in Table 1, and the frequency of each keep sense in St4 and Brown is quite distinct. St4 non-English major learners mainly use the first four senses, which are sense 1, sense 2, sense 4, and sense 3 ranging according to frequency, whereas sense 12 and sense 14 only appear once in St4. In the first four senses, the percentage of sense 1 in St4 is more than twice of that in Brown. With regard to the percentage of sense 4, it is about 10% higher in St4 than in Brown. Nevertheless, the percentage of sense 2 in Brown almost doubles that in St4. Furthermore, sense 3 in Brown was 2.4% higher than the frequency in St4. All in all, the top four most commonly used senses in St4 are sense 1, sense 2, sense 4, and sense 3 according to frequency, which account for 98% of all keep occurrences in St4.

With regard to the two learner corpora St3 and St4, the first four senses in St3 and St4 and their order are consistent, which are sense 1, sense 2, sense 4, and sense 3. When it comes to the other ten senses, St3 and St4 non-English major learners seldom use them. However, compared with St3, the percentage of sense 1 and sense 2 in St4 was closer to that of Brown. This may signify that Chinese learners’ English level gets improvement with the advancement of their English learning. Nevertheless, no matter how high their English level
level is, Chinese learners are still likely to become closer to native speakers, whereas it is quite difficult for them to be exactly the same as English native speakers.

3.2.2. General Analysis of the Results. The above results show that although the verb keep has been learned in a comparatively early stage of English learning, Chinese non-English majors still do not comprehensively master the fourteen meanings of the verb keep as English native speakers do. Some senses like sense 1, sense 2, sense 4, and sense 3 are quite frequently used, while others are rarely used. The fact that Chinese teachers and learners have always emphasized the first four senses since they began to learn English. Therefore, they are inclined to use keep with regard to the first four senses. Third, even after being admitted into college, students are likely to use words that they are quite familiar with and reluctant to memorize and employ unfamiliar and difficult vocabulary. As a result, they can save energy and reduce the burden of memorizing new words. Fourth, with regard to the last ten senses that Chinese learners seldom use, Chinese learners may be are lacking in adequate input. However, it does not necessarily mean that Chinese learners do not know these senses. The reason is likely to be that Chinese learners do not attach great importance to these ten senses. On the contrary, they prefer to use words or expressions other than keep. Finally, the overuse of sense 1 in St3 may be correlated with the corpus to some extent. Many St3 learners use “so in order to keep fit” repeatedly in one composition. The materials in St3 are all CET-4 compositions or compositions of similar levels, and one of them is “health.” The title of the composition may contribute to limit learners’ vocabulary output and expression. There are as much as 40 “keep fit” in St3 in all.

The overemphasis on the first four senses, the unfamiliarity of the other ten senses, the limited range of vocabulary grasped by Chinese learners, the reluctance to memorize new vocabulary, and the limitations of the topics of compositions in St3 and St4 corpora can be attributed to the difference in the employment of keep senses across the three corpora.

Generally speaking, with the improvement of English proficiency, Chinese learners’ use of keep is much closer to that of native speakers.

The present corpus-based study of verb keep can provide some implications for EFL teaching and learning in China, especially for the teaching and learning of English verbs. English verbs deserve more attention during vocabulary teaching and learning. Students should try to enlarge the depth and breadth of their understanding of vocabulary. Remembering as many words as possible can greatly help students to express themselves accurately and appropriately. At the same time, because the choice of words and sentences can demonstrate students’ English level to a great extent, teachers are also supposed to direct the students to differentiate various similar terms, enrich their wording and phrasing, and offer profound example illustrations so as to expand students’ words and phrases. Employing high-frequency words as much as possible is bound to get low grades and does not deserve excellence.

Blended teaching can be applied in verb teaching. Before class, microlectures on verb keep can be provided online. The sequence of the different senses of keep can be shown in accordance with the frequency of different senses in Brown from high to low. Keep phrasal verbs, idioms, and collocations can be included in microlectures. Students ought to learn these materials before class and know the different senses of keep, phrasal verbs, idioms, and collocations. In class, teachers and students can draw mind maps of verb keep collaboratively. After class, more assignments can be provided online, such as multiple choice, blank filling, translation, and so on.

Data Availability
The data that support the findings of this study can be obtained from the author upon reasonable request.

Conflicts of Interest
The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References


