A contrastive analysis of the use of verb "keep": A corpus-based study between CLEC and Brown

CHEN Fang, MEI Zhong-fang

(College of Foreign Languages, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: This is a corpus-based research focusing on the high-frequency verb "keep" used by Chinese non-English majors and native speakers. The corpora involved in the paper are Brown which stands for native speakers, Students 3 and Students 4 in CLEC (Chinese Learner English Corpus) which stand for Chinese non-English majors. The paper tries to investigate the similarities and differences between them and put forward the possible reasons. At last, some suggestions of ELT (English Language Teaching) are given.

Key words: "keep"; use; contrastive analysis

1. Introduction

Vocabulary plays an important role in Second Language Acquisition (SLA). English verbs are the most active and variable among all parts of speech, which impose great difficulty on language learning, especially on high-frequency verbs. Therefore, it is very important for EFL (English as a Foreign Language) learners to have a good command of these high-frequency verbs. However, although most of the high-frequency verbs are taught at a comparatively early stage of instruction, they are still not fully acquired by Chinese learners. EFL learners may still commit errors on the use of these verbs, not to say that their English can be native-like.

According to *Longman Dictionary of Contemporary English* (Summers, 2004, p. 1063), "keep" is one of the most frequently used words in oral and written English in BNC (the British National Corpus). This paper aims to investigate the characteristics of the "keep" used by Chinese non-English majors (St3 and St4 in CLEC) in comparison with native speakers (Brown).

The authors tend to find out the similarities and differences on the frequency of "keep" employed by Chinese learners and native speakers. And the possible reasons are put forward.

2. A corpus-based research on verb "keep"

2.1 Research method

The study is a corpus-based and computer-aided study. Two corpora are employed: a learner corpus—St3 and St4 in CLEC and a native speaker corpus—Brown.

CHEN Fang, lecturer of College of Foreign Languages, Qingdao University of Science and Technology; research field: linguistics and didactics.

MEI Zhong-fang, lecturer of College of Foreign Languages, Qingdao University of Science and Technology; research field: language teaching.

2.2 Data collection and research procedure

2.2.1 Instrument

The data needed in the paper are sentences where verb "keep" appears. To acquire the data, computer software is applied for both convenience and efficiency. Antconc, SPSS, Microsoft Excel and Wordsmith are utilized for the research. Antconc is applied to extract the desired occurrences of the tested words and other relevant items. Software Package SPSS is adopted to test chi-square value. Microsoft Excel is used for calculation and vivid presentation of the results. Wordsmith is employed to carry out TTR (Type-Token Ratio) of the three corpora in the research.

2.2.2 Data collection

Two sub-corpora—St3 and St4 are chosen as the sample corpora of learner language in the study. St3 represents Chinese non-English majors who have taken CET-4 and St4 stands for Chinese non-English majors who have taken CET-6. Thus, St4 stands for a higher proficiency in comparison with St3. Brown corpus contains 500 samples with each of about 2,000 continuous written English. What's more, the texts are sampled from 15 different text categories. As the subjects in CLEC can represent English learners in China, according to YANG (2002) and GUI (2004), the learner corpus can be used objectively and is compatible with native speakers' corpus.

As the study is on verb "keep", "keep" should be first sorted out from the corpora. First, Antconc is applied to sort out all sentences containing "keep", and its other tokens "keeping", "kept" and "keeps", in the 2 corpora—St3 and St4. Because "keep" and "keeping" can also be used as nouns, so the noun forms of "keep" and "keeping" are extracted from the total "keep" occurrences. Besides, the compounds consisting of the 4 tokens are also picked out manually, such as "well-kept". Because of the imperfection of the corpora, there exist some repeated lines in the corpora. They are all extracted manually from the total occurrences, so that only one of them is left. According to the above principles, the total number of "keep" occurrences in Brown is 518. In St3, the number is 206, and in St4, it is 124. Consequently, all effective "keep" occurrences are sorted out. Since all occurrences have been sorted out, the following part is devoted to the detailed research procedure with the help of the above instruments.

2.2.3 Research procedure

Distributions of overall "keep" occurrences in the three corpora are examined and their frequencies and percentages are calculated. It includes the comparison of the use of verb "keep" by native speakers and Chinese learners. The outcomes are presented in tables and figures if necessary. The characteristics of Chinese learners' use of "keep" are analyzed.

3. Result and discussion

3.1 Distribution of overall "keep" occurrences among the three corpora

To investigate the use of verb "keep" in terms of overuse and underuse, the frequencies of "keep" occurrences in Brown and St3, St4 in CLEC are calculated.

Table 1   Distribution of verb "keep" across the corpora

Corpus	Keep frequency	Corpus (size/words)	Percentage (%)
Brown	518	1015537	0.051
St3	206	209043	0.0985
St4	124	212855	0.0583

According to Table 1, the total frequencies of verb "keep" in St3 and St4 are lower than that in Brown. However, the percentage "keep" takes in every corpus shows that "keep" used in St3 is soundly larger than that in Brown, while the percentage of "keep" used in St4 is much smaller than that in St3 but slightly larger than that in Brown. The differences in overall frequency might be the result of different sizes of the corpora. Therefore, chi-square tests are performed to investigate whether the differences among the learner corpus and the native corpus are significant. Table 2 shows chi-square tests between the distribution of "keep" in Brown and St3.

Table 2   Chi-square test on distribution of "keep" across Brown and St3

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson chi-square	66.200	1	0.000		
Continuity correction	65.399	1	0.000		
Likelihood ratio	57.500	1	0.000		
Fisher's exact test				0.000	0.000
Linear by linear Association	66.200	1	0.000		
N of valid cases	1225304				

Note: Computed only for a 2×2 table. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 123.64.

The critical value for all chi-square tests in the present study 3.84 for one degree of freedom at 5% level. According to Table 2, the chi-square value of "keep" across Brown and St3 is 66.200, which is much larger than 3.84. Consequently, it is safe to say that there is a significant difference in the frequency of verb "keep" across Brown and St3. St3 group employs verb "keep" more frequently than native speakers do.

Using exactly the same procedure chi-square value across Brown and St4 is carried out in Table 3.

Table 3   Chi-square test on distribution of "keep" across Brown and St4

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson chi-square	1.768	1	0.184		
Continuity correction	1.632	1	0.201		
Likelihood ratio	1.718	1	0.190		
Fisher's exact test				0.191	0.101
Linear by linear Association	1.768	1	0.184		
N of valid cases	1229034				

Note: Computed only for a 2×2 table. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 111.25.

As is shown in Table 3, chi-square value of "keep" across Brown and St4 is 1.768, which is smaller than 3.84. Therefore, there is no significant difference in the frequency of verb "keep" across Brown and St4. As St4 group is non-English majors who have taken CET-6 while St3 group is non-English majors who have taken CET-4, the tentative reason might be with the improvement of English proficiency, second language learners do develop their ability of using verb "keep" and their use of "keep" is more and more closer to that of the native speakers, at least in terms of frequency. However, there is no significant difference does not mean there is no difference. In comparison with Brown, St4 group still has a minor overuse of verb "keep".

Similarly, chi-square test across St3 and St4 is performed in Table 4.

Table 4   Chi-square test on distribution of "keep" across St3 and St4

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson chi-square	21.869	1	0.000		
Continuity correction	21.357	1	0.000		
Likelihood ratio	22.082	1	0.000		
Fisher's exact test				0.000	0.000
Linear by linear Association	21.869	1	0.000		
N of valid cases	422228				

Note: Computed only for a 2×2 table. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 163.54.

According to Table 4, chi-square value of "keep" across St3 and St4 is 21.869, which is much larger than 3.84. Thus there is a significant difference in the frequency of verb "keep" across St3 and St4. St3 learners overuse verb "keep" a lot in comparison with St4 learners.

From the above analysis, Chinese English learners—St3 and St4 groups tend to overuse verb "keep" in comparison with native speakers. The lexical density, or the authors can say TTR (Type-Token Ratio) can help to explain the phenomenon. The TTR of the 3 corpora is listed in Table 5 using software Wordsmith.

Table 5   TTR in Brown, St3 and St4

Corpora	Brown	St3	St4
Type	42579	7757	8648
Token	1015537	232541	241969
TTR	4.19	3.34	3.57

It is obvious that Brown has the large range of vocabulary among the 3 corpora. The lowest TTR 3.34 indicates a limited range of vocabulary used in St3, while the TTR of St4 is much lower than that of Brown and a little higher than that of St3. Consequently, some high-frequency verbs like "keep" take too much portion in Chinese learners' English. Therefore, the overuse of high-frequency verb like "keep" occurs.

Chinese learners tend to overuse verb "keep" in comparison with native speakers out of their limited range of vocabulary. However, with the improvement of their English proficiency, St4 learners' use of "keep" is much closer to that of native speakers in comparison with St3 learners.

4. Conclusion and pedagogical implication

From the above analysis, people can know that St3 and St4 learners both tend to overuse verb "keep" in comparison with native speakers out of Chinese learners' limited range of vocabulary. However, as to the degree of the overuse, St3 learners rank the top. It shows that with the improvement of English proficiency, Chinese learners' English is much closer to that of the native speakers.

Generally speaking, Chinese non-English majors tend to overuse verb "keep" in comparison with native speakers. St4 learners' overall use of verb "keep" is closer to that of the native speakers with 0.0583% in comparison with 0.0985% by St3 learners.

The reason for Chinese non-English majors of overuse of verb "keep" may due to the poor vocabulary grasped by Chinese learners, which can be identified from TTR. Thus, they try to make full use of high-frequency verbs they are more familiar with including "keep" even when it is not their turn.

English verbs should be paid more attention. Students should try hard to enlarge their vocabulary in depth and width. Keeping many words in mind can help the students to express themselves exactly. Teachers should also instruct the students to distinguish the synonyms and expand their vocabulary, especially in compositions, for vocabulary can partly illustrate the students' English proficiency. Employment of too many high-frequency verbs does not deserve a high mark.

References:

GUI, S. C.. 2004. A cognitive model of corpus-based analysis of Chinese learners' errors of English. *Modern Foreign Languages*, *2*, 129-139.

GUI, S. C. &YANG, H. Z.. 2003. *CLEC*. Shanghai: Shanghai Foreign Language Education Publication.

JIANG, Z. R. & ZHONG, Y. F.. 2002. The connection between English verbs and pattern structure. *Journal of Chuanbei Educational College*, *4*.

LI, X. H.. 2004. A corpus-based comparison study of collocation behavior of "effect". *Foreign Language Teaching*, *6*.

Summers, D.. 2004. *Longman dictionary of contemporary English*. London: Pearson Education Limited.

YANG, H. Z.. 2002. *An introduction to corpus linguistics*. Shanghai: Shanghai Foreign Language Education Publication.

ZHANG, N.. 2009. *A contrastive analysis of the use of the verb "keep"*. (Master Paper, DaLian Maritime University)

(Edited by Chris and Suky)