An Empirical Study on the Distribution of Vocabulary of College English Textbooks

Feng Juan

Basic education department of Southwest Petroleum University, Nanchong Sichuan 637100 China

ABSTRACT.   This study analyzes and evaluates the textbooks based on the four sets of college English textbook corpus on account of the word family, and discusses their inter-discourse lexical growth patterns, lexical density and word families of CET4 and CET6. The coverage rate and the inter-discourse vocabulary repetition rate of the word family tables in each textbook are take into account in evaluation. The results show that the Brunet model has a good fit to the vocabulary growth curve of the four sets of textbooks. This model can be used to predict the vocabulary coverage of the four sets of textbooks for natural texts of any length outside the textbook. The overall difference in vocabulary density of the four sets of textbooks is not very large, but some textbooks are not arranged according to the difficulty of each book; four sets of textbooks can not completely cover the CET4 and CET6 family table, moreover, two-word family table has lower inter-discourse vocabulary in the four sets of textbooks. On the basis of qualitative analysis, using the corpus as a means to quantitatively analyze the vocabulary of the 21st Century College English Applied Comprehensive Course, and to calculate the relatively objective evaluation results, I hope to bring some new ideas to the current textbook evaluation. And provide some reference for the design and preparation of the textbook.

KEYWORDS: Corpus; College English; Lexical analysis

1. Introduction

Usually when evaluating a textbook, we will conduct subjective evaluations based on certain standards or evaluation lists from the aspects of textbook writing guidelines, textbook content, and style. Considering that an important aspect of college English teaching is to help students expand their vocabulary, this study uses the corpus approach to objectively analyze and evaluate four sets of general English textbooks. Vocabulary is the driving force of language output, but the difficulties encountered by learners in learning and using language are mostly related to vocabulary. Studies have shown that although Chinese college students have received many years of English education, the limitations of their vocabulary knowledge are still the main constraints that hinder their effective English input and

output.[1]Therefore, how to expand students' vocabulary and improve their vocabulary acquisition level has become a problem that cannot be ignored in foreign language teaching and research. On the other hand, the word family can reflect the intrinsic relationship between vocabulary, strengthen the vocabulary grammar awareness of the students, and help the learners to expand the vocabulary and reduce the vocabulary learning burden by means of affixes. Therefore, the word family concept is used to guide the vocabulary. For English learning in a classroom setting, the textbook is the primary medium for learners to reach the target language. Chinese English learners' vocabulary acquisition mainly depends on the college English textbooks they use. Therefore, this study selects the most widely used College English Teaching Course, the New Age Interactive English Reading and Writing Course, the New College English Comprehensive Course, and the New Vision College English Reading and Writing Course. These four sets of textbooks (replaced by T1, T2, T3, and T4 respectively) are used to analyze the distribution of word families in the following three questions to evaluate their effects on student vocabulary acquisition: one is four the textual vocabulary growth model of the textbook; the second is its lexical density; the third is the coverage of the fourth and sixth vocabulary and the inter-discourse vocabulary repetition rate of the two vocabularies in the textbook. I hope this study can provide some reference for vocabulary development, textbook writing and promotion of Chinese English learners' vocabulary acquisition.

2. Qualitative Research on Teaching Materials

According to the "Teaching Requirements for College English Courses", the textbook is an important guarantee for achieving the teaching objectives defined by the syllabus, and the cultivation of language application ability is the key point. The textbook should provide the best language samples and systematic and targeted language practice activities for classroom teaching, so that students can lay a solid foundation in language, develop language application skills and improve cultural literacy. [4] The formation of language ability is based on basic language knowledge and skills, cross-cultural awareness and communication skills. According to the description of the "Requirements", and with reference to the above-mentioned foreign mainstream evaluation systems and the evaluation principles of textbooks by relevant domestic scholars, the author conducts qualitative analysis on the textbooks from the following aspects: subject selection, language input and output, and the wide selection of content. The qualitative analysis of this study refers to certain teaching material evaluation criteria, and compares the requirements of textbook design and writing one by one, and draws the following conclusions.

(1) The theme is novel, the material selection is unique, the focus of contemporary college students is grasped, and students' interest in learning language is enhanced. The unit theme of this set of textbooks is determined by the authors after extensively soliciting students' opinions through questionnaires. According to the development needs of the times, the theme and selection of each unit can attract students' great attention and lead to heated discussion. Make classroom teaching lively. Each of the 1-4 volumes offers 8 different themes, covering a wide range and

reflecting the spirit of the times. For example, the second unit of Book 1 is about the topic of "Campus Love", which is a topic of "Campus Love". [5]The author has arranged the topic of first love in the form of composition in front of this unit. Many students have reviewed their own excitement or sadness or sadness. In the past emotional experience, I also learned how to deal with the relationship between campus love and learning life from Text B.

(2) Provide sufficient language input and output preparation to inspire students to introduce new knowledge through reserve knowledge. The reading and understanding of text is a complex process in which text interacts with the images in the reader's mind. When introducing new knowledge from the knowledge structure that students already have, it is especially important to provide context and to mine the knowledge stored in the learner's mind. This series of textbooks fully provides and mines the knowledge stored in the students' minds in theme import and practice design, making learning a continuation process from old to new. For example, in the Starter section of each unit, based on the vocabulary that the student already has and the subject of the unit, the theme is also imported in the form of the student's favorite video. The video of the Writing section of the exercise also helps to motivate the students, export the existing language knowledge reserves, and successfully complete the writing practice.[6] In order to extend the practical application of language output and language, each unit has added a unique part of workshop, requiring students to complete a project-based writing and summarization through mutual aid and cooperation. In this way, the learner cultivates and improves the language application ability in the continuous practice. As Hedge pointed out that the language communication content includes language ability, pragmatic ability and discourse ability, cultivating these abilities requires a lot of practice and practice, and the teaching materials should create conditions. To provide students with a large number of language materials with real communication scenarios, from this point of view, this set of textbooks meets these requirements, and strives to focus on developing the practical application of students' language in all aspects.

(3) The textbook provides a digital college English teaching platform, which is conducive to students' independent learning. An integrated teaching model is proposed, which is based on computer and classroom English teaching mode, and emphasizes that one of the teaching objectives is to "enhance the ability of students to learn independently to adapt to the needs of social development and international communication in China". Fudan University Press and textbook writers jointly created a digital learning platform for college English teaching based on paper materials.[7]The "application type" textbook includes self-learning module, teaching assistant module, teaching evaluation module, network management module and interaction module between teachers and students, which helps teachers to obtain feedback information comprehensively, objectively and accurately, and improve teaching management. Students can also adjust their learning strategies in a timely manner to improve their learning efficiency.

3. Research Design

3.1 Research corpus

The corpus based on this study is four sets of college English textbooks, and Table 1 is the basic information about the relevant materials.

Table 1 Basic data of four sets of teaching materials

Teaching material	Number of articles	Article length	Total number of symbols
T1	64	332 ~930	43 439
T2	64	730 ~2 008	71 874
T3	80	320 ~ 917	58 715
T4	80	554 ~ 980	65 573

3.2 Data analysis

The researchers used Visual Foxpro 60 to write a computer program to extract the word family. The steps are as follows which show in Figure 1:
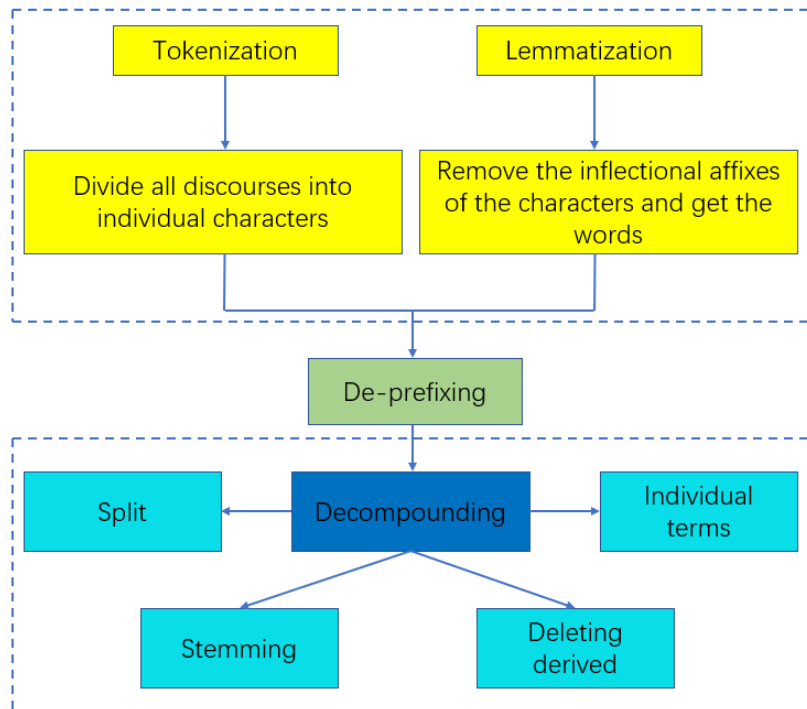

Figure 1 Data analysis procedure

Tables 2 and 3 are the processed data of the two vocabularies and the four sets of textbooks respectively.

Table 2 Statistics of the fourth- and sixth-level word family tables

Word family table	Class symbol	Term	Word family
CET4	4 795	4 432	3 366
CET6	6 395	5 950	4 338

Table 3 Total vocabulary of the four sets of textbooks

Teaching material	Shape	Class symbol	Term	Word family
T1	43 439	6 193	4 163	2 911
T2	71 874	8 301	5 461	3 654
T3	58 715	7 564	4 963	3 417
T4	65 573	7 798	5 158	3 597

3.3 Vocabulary density statistics of textbook vocabulary

The token is similar to the "word" we say daily. The type refers to any unique form in the corpus text. In a text, a repeated character can only be recorded as a class. The type-token ratio (TTR) is a common method for measuring the lexical density in text, and the standardized type-token ratio (STTR) is a measure. Nation [8] has mentioned that the vocabulary density index of modern written English is 1:2.4, which is 0.4176. The higher the vocabulary density, the more difficult the text is. The lexical density of spoken text is lower than the lexical density of written text. Spoken language is less than 40% and written is greater than 40%. According to the "Requirements", college students' reading speed for general topics is 70 words per minute. Therefore, the lexical density of written texts in English textbooks (ie, the index of text difficulty) should be around 40%, and the vocabulary density should be considered in textbook articles.

Table 4. Statistical results of TTR and STTR

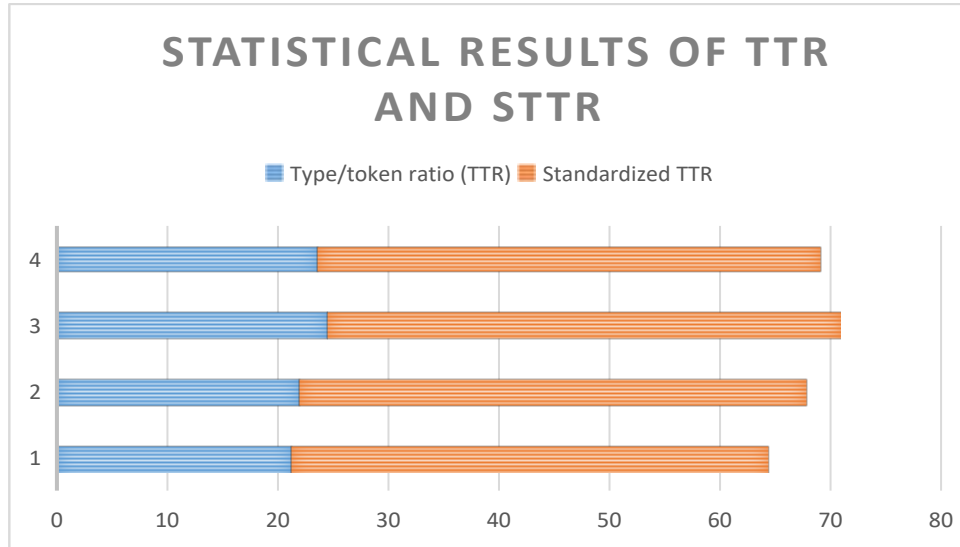|	Book 1	Book 2	Book 3	Book 4	All Books(1-4)
Tokens (running words) in text	12,543	12,919	15,418	16,403	57,283
Types (distinct words)	2,605	3,131	3,596	3,567	7,748
Type/token ratio (TTR)	21.09	21.89	24.49	23.52	13.66
Standardized TTR	43.13	45.85	46.3	45.47	45.35

Figure 2. Statistical results of TTR and STTR

Table 4 is about the statistics of TTR and STTR in the 1-4 volumes of textbooks. These two indicators reflect the difficulty and density of the words in the text. The larger the index value, the greater the difficulty. It means that the greater the difficulty and density of the new words, the more difficult it is for students to learn the text. From Figure 2, it can be found that the TTR values of the 1-4 volumes of the application type are all above 20, the STTR value is above 43, and the TTR of the whole set of teaching materials is 45.53, which indicates that the vocabulary of the textbook is too difficult and the density of the new words is too large. The total number of Types covered by the 1-4 books is Book 1 (2605); Book 2 (3131); Book 3 (3596); Book 4 (3567); STTR are Book 1 (43.13); Book 2 (45.85); Book 3 (46.3); Book 4 (45.47), it can be seen that there is a gradual trend in the distribution of the overall STTR value of 1-3 books, that is, 3 is greater than 2 and greater than 1, indicating "21st Century" 1-3 The difficulty of the texts is basically reasonable, basically follows the law of language learning, and reflects the principle of easy to difficult and gradual progress in the preparation of textbooks. However, it can be seen from the total number of words and the STTR value that the difficulty of Book 4 is lower than that of Book 3, and the students of Book 4 are about to face the CET4. Therefore, the author suggests that teachers should add more vocabulary to the teaching of CET4. The vocabulary and extension exercises are more adequate for students to actively respond to the CET4.

3.4 The average length of words, sentences and texts

The amount of words in each sentence represents the complexity of the textbook structure. Table 5 shows the average length of words, sentences, and texts in the

Applied 1-4 textbooks using Wordsmith 5.0 software. In general, the average length of the texts is increasing from Book 1 to Book 4, but the data also indicates that the average length of the words in Book 4 is lower than the first three volumes, consistent with the previous discussion on STTR. Form Figure 3, Book 3 is lower than the other three volumes in terms of the average length of the sentence, indicating that the textbook does not have a more appropriate arrangement in terms of sentence length. At the same time, compared with the length of the sentence in the reading comprehension of the 4th and 6th grade exams, the 1-4 volumes are lower than the average sentence length of the 4th and 6th grades: the fourth grade 19.01, the sixth grade 20.21 [6], fail to reach the fourth and sixth grade exams. I hope this will attract the attention of the school, and if necessary, adjust and supplement the corresponding reading training.

Table 5. Statistical results of the average length of words, sentences, and texts

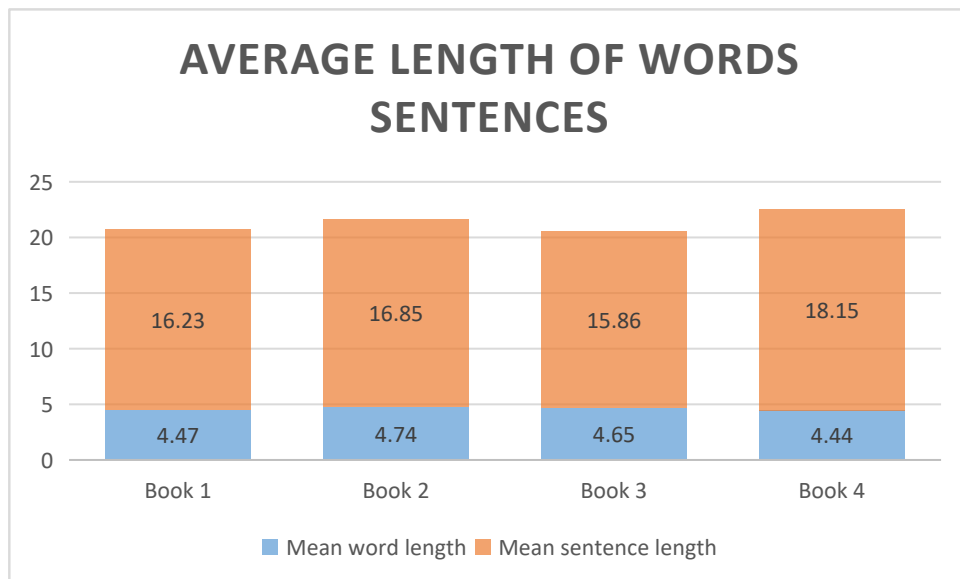|	Book 1	Book 2	Book 3	Book 4	All Books(1-4)
Mean word length	4.47	4.74	4.65	4.44	4.57
Mean sentence length	16.23	16.85	15.86	18.15	16.77
Mean text length	772	799	956	1018	886

Figure 3 Statistical results of the average length of words and sentence

4. Conclusion

The results of this paper show that the word family concept can be better used to study the vocabulary distribution characteristics of the four sets of textbooks compared to the class and the word. The research data in this paper indicates that the

theoretical value range of the vocabulary coverage of the two sets of textbooks for English natural texts of 2000 words is 85% to 93%. The overall difference in vocabulary density between the four sets of textbooks is not very large, but in some textbooks, the difficulty of the four volumes is not incremental. This shows that we should pay attention to the difficulty of the article in the process of writing the textbook, and should make full use of the existing technical means to deal with the selected materials. In this way, the order of the units and the order between the books and the books can be arranged more scientifically and reasonably to ensure the gradual progress of teaching. In addition, the four sets of textbooks can't completely cover all the CET4 and CET6, and the ratio of the CET4 and CET6 grades in the four sets of textbooks is five times or more. According to the current situation, the knowledge of English vocabulary of Chinese college students is still the main constraint that hinders their understanding and output of English. College English course materials take on the responsibility of enriching students' vocabulary knowledge. Therefore, improving the vocabulary coverage of the CET4 and CET6 of textbooks, especially improving their repetition rate in the intertext, will help students' vocabulary acquisition and further enrich their vocabulary knowledge, thus further improving their English proficiency. Finally, most vocabularies are formulated in terms of terms, so some derivatives are listed as separate terms, such as: care, careful, careless, which is easy to cause repeated vocabulary teaching, and is not conducive to students through word formation Improve their ability to recognize derivations. The word family can solve this problem and effectively reduce the burden of learners' vocabulary learning. Teachers can teach students affix knowledge according to their English level and help them remember words efficiently.

This study mainly analyzes and evaluates the vocabulary distribution characteristics of four sets of college English textbooks from the perspective of family. Other aspects of vocabulary research, such as the growth model of new words and the collocation of core vocabulary, are not covered in this paper. In addition, vocabulary is only one aspect of textbook writing, and there are many other factors that need to be considered, such as genres, topics, etc., so more research is needed to conduct a more comprehensive analysis and evaluation of textbooks.

References

[1]Zhang M . An Empirical Study of Applying Associative Method in College English Vocabulary Learning.[J]. English Language Teaching, 2014, 7(6).

[2]Xu X . Study on the Effect of Dictionary Use on Second Language Incidental Vocabulary Acquisition— An Empirical Study of College English Vocabulary Learning Strategy[J]. Journal of Language Teaching & Research, 2010, 1(4):519-523.

[3]Maghsoudi M , Talebi S H , Mirkamali F . The Impact of Different Tasks on Incidental Vocabulary Acquisition Regarding Different Types of Dictionary[J]. Procedia - Social and Behavioral Sciences, 2014, 9(8):1056-1061.

[4]Roth, Daniel. Morphemic Analysis as Imagined by Developmental Reading Textbooks: A Content Analysis of a Textbook Corpus[J]. Journal of College Reading and Learning, 2017,47(1): 26-44.

[5]Liu C L . A Review of Effect of Different Tasks on Incidental Vocabulary Acquisition[J]. Higher Education Studies, 2015, 5(5).

[6]Zou D , Xie H , Rao Y , et al. A Comparative Study on Various Vocabulary Knowledge Scales for Predicting Vocabulary Pre-Knowledge[J]. International Journal of Distance Education Technologies, 2017, 15(1):69-81.

[7]Heidari-Shahreza M A , Tavakoli M . The effects of repetition and L1 lexicalization on incidental vocabulary acquisition by Iranian EFL Learners[J]. Language Learning Journal, 2016, 44(1):17-32.

[8]Tian L , Macaro E . Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: A Lexical Focus-on-Form study[J]. Language Teaching Research, 2012, 16(3):367-391.