Vocabulary Analysis of English Major Textbooks Based on Range Software

Wei Zhou
Dalian University of Science and Technology
467512256@dlvtc.edu.cn

Abstract—Range is a program that provides vocabulary data. It can automatically measure the vocabulary produced by students' work, and through the comparison of the excellent degree of students' work, formulate the direction of subsequent teaching work. This can provide teachers with a new vocabulary analysis method for the text of English majors. This article summarizes the current situation of text vocabulary in English majors. The author discusses the results of this research and the content of the discussion from three aspects: statistics of the vocabulary coverage of the textbook and the vocabulary matching degree required by the syllabus, statistics of vocabulary recurrence rate, vocabulary learning opportunities, and case analysis.

1. INTRODUCTION
Textbooks are the core teaching materials and the actual manifestations of the syllabus. They are also an important basis for students to acquire knowledge and teachers to teach. From the perspective of the characteristics of language textbooks, textbooks can usually become a content carrier for a certain learning stage. This also includes specific language knowledge and language skills. It can provide resources and tools for teaching and become a teaching guide for teachers. It can also be seen from this that textbooks play an irreplaceable role in the development of actual English teaching work.

2. CURRENT VOCABULARY OF TEXTBOOKS FOR ENGLISH MAJORS
2.1 The Status Quo of Research on Text Vocabulary of English Majors in China
In recent years, China has edited dozens of applied teaching materials for English majors, but only a few English majors have been widely used after subsequent revisions. In addition, Chinese domestic scholars have a very narrow scope of research in the field of English majors, mainly focusing on foreign language teaching materials evaluation standards and compiling methods and methods. Secondly, from the perspective of research content, the existing textbooks do not present the characteristics of gradient and depth, but focus on the research history of English textbooks and the introduction of the textbook evaluation system. More importantly, the main research methods of the textbooks are lacking. In general, the analysis method used is qualitative analysis. From the development process of the past ten years, it can be seen that many textbook users have summed up many new experiences from teaching practice, and have comprehensively summarized the characteristics and deficiencies of the textbooks. But even in the "China English Proficiency Rating Scale", the text description of each level of ability classification lacks corresponding quantitative standards. On the whole, domestic research on the evaluation of English majors mainly focuses on macro or qualitative aspects. This also makes the assessment of English majors lack of objective inspection standards. More importantly, teachers often choose the type of teaching
materials based on word-of-mouth communication among groups, or it is difficult for leaders and publishers to promote the targeted characteristics of teaching materials [1].

2.2. Research Status Abroad
In many areas abroad, the education department will encourage teachers to adapt or self-edit textbooks according to their own and student needs. Moreover, teachers also have the right to decide on textbooks, and teachers often do not carry out systematic evaluation of textbooks issued by some publishing houses. Inconsistent with the research basis of foreign English textbooks, Chinese English textbooks are mainly based on the syllabus, and they are responsible for the syllabus. For example, in 2000, the English Teaching Steering Committee of Colleges and Universities formulated the "English Teaching Syllabus for Colleges and Universities". It clearly points out the specific teaching requirements and level instructions for English majors, which also involves quantitative standards for vocabulary teaching. In fact, foreign languages mainly provide learners with appropriate corpus input. With the help of a reasonable teaching process, learners can learn to master various forms of corpus, so as to achieve language output. The input corpus is formed according to the form of listening and reading. At this point, the reading session will be very important. Vocabulary is the basic content that carries the meaning of language, which will have an impact on reading comprehension. Therefore, the actual material reading is the process by which the text of the textbook has an impact on English teaching. In order to facilitate the research, this article chooses two sets of English major textbooks, "New English Textbook" and "Comprehensive Course", and chooses the first volume of each textbook to read the text for analysis, in order to solve the following problems. First, whether the text of the textbook contains vocabulary consistent with the requirements of the syllabus. Second, whether the text of the textbook can provide students with corresponding vocabulary learning opportunities [2].

3. Research Method
The main program used in this study is the Range program. The software application is based on the BNC corpus. 14,000 word families are selected from 100 million words according to word frequency, and every 1000 word families is one word level. Among them, the 100 word families involved in the first level belong to high-frequency word families. From the second level to the 14th level, the word frequency is in a state of decreasing. The main basis of these 14-level theoretical research on word level is that learners master vocabulary in order of word frequency from high to low. Besides, there are many proper noun vocabularies and interjection vocabularies involved in the entire BNC corpus. These vocabularies can also be designed in the analysis conditions in actual software analysis. What’s more, BNC represents the most authoritative British English corpus in the next stage. Among them, written corpus accounted for 90% and spoken corpus accounted for 10%. This situation can maintain mutual agreement with the textual features of the textbook based on written language corpus in this research. The specific analysis results can also be compared with the quantitative vocabulary indicators in the "Outline" and the students' existing vocabulary, so as to solve specific problems.

The specific research process is as follows. First of all, the self-built small corpus of target textbooks selected by this research is the reading text text1 and text2 in the selected textbooks. In addition, the content of "New Edition" mainly includes 30 articles, the number of symbols is 15,241, and the "Comprehensive" contains 28 articles. Among them, the number of pictograms is 24384, which can be used as the main corpus content of this research analysis. In the study of English word symbols, the main content involved are English words, letters, and abbreviations with space breaks, but punctuation marks are not involved. Secondly, researchers should do a good job in self-building small corpora to ensure that various interference factors in the analysis process of Range corpus can be eliminated. After the pre-processing work, the number of symbols is 15078 in "New Edition" and 24369 in "Comprehensive". In the actual pretreatment process, the main content involved is as follows. Firstly, spell checking. It checks the spelling errors of individual words during the conversion process. Meanwhile, in order to match British English in the corpus vocabulary, people also need to convert American English vocabulary to avoid software identifying such vocabulary as "out-of-thesaurus". Secondly, hyphen replacement. Under
normal circumstances, the Range software will incorporate hyphenated compound words into the "out-of-thesaurus" list. It should be noted that most of the words in these compound words are the first 14 levels of vocabulary. In consequence, people should conduct a comprehensive analysis of the corpus and replace all hyphens with space symbols [3].

After the above operations are completed, researchers can use the "Vocabulary Test" tool developed by Nation and Beglar to carry out the vocabulary test work for freshmen. The number of test personnel is 129. The development of the measurement tool is mainly based on the "14-level BNC1000 word family table", which can maintain the overall development of the design work.

4. RESEARCH RESULTS AND DISCUSSION CONTENT

4.1. Statistics on the Vocabulary Coverage of Textbooks and the Vocabulary Matching Degree Required by the Syllabus

S. P. Nation believes that vocabulary coverage is mainly the proportion of vocabulary known to readers in the total number of words in the text. This is also the best measure to judge whether the text can be fully understood. Generally speaking, the impact of vocabulary coverage on reading comprehension will vary with the type of text. However, researchers M. Hu and others also stated that vocabulary coverage and reading comprehension can show a clear linear relationship, and our actual reading comprehension level will also increase with the increase of vocabulary coverage. In contrast to all second language learners, if someone can recognize more than 95% of the vocabulary in the text, their reading comprehension ability is high. But if they want to better improve their personal abilities, their vocabulary knowledge should be kept above 98%. In other words, in the actual textbook text, if there are 4 new words in 50 words, the specific vocabulary coverage rate is 92%. At this point, students can also understand the text with the help of a dictionary or a teacher.

4.1.1. Realistic Data Analysis

It can also be seen from the actual research data of lexicon that the vocabulary coverage rate of the most frequent 1000-level vocabulary ranges from 78% to 81%. The second and third high-frequency word levels accounted for 8% to 9% and 3% to 5% respectively. The proportion of the 6th and 9th word levels under the intermediate frequency is 2%. The low frequency 10th to 14th word level accounts for 1% of the text ratio [4].

The "Outline" puts forward specific requirements for the language knowledge and ability of English majors, which also includes the mastery of vocabulary. For example, English majors should master 4000 to 5000 vocabulary after completing their first-year studies. This also includes 2,000 in the middle school period, and they need to correctly use the basic collocation of 2,000 to 2,500. It can also be seen from this that when the course of a semester is completed, the vocabulary of the students needs to be maintained at least around 4000. Another 1000 vocabulary increases can be acquired in the second semester of study. In order to ensure that students’ vocabulary mastery can meet the requirements after the first semester, teachers should ensure that the vocabulary in the first book textbook should reach 92% to 92% from the first word level to the fourth word level based on the above-mentioned vocabulary research data information 95% coverage. Only in this way can it be ensured that students have sufficient vocabulary learning and understandable language input.

It can also be seen from the specific lexical research process that educators need to study the vocabulary of the text. For example, there are 7 symbols in It is easy to say it correctly, but there is a problem of it repetition, and there are only 6 parts of speech. It can also be seen from this that it is very scientific to calculate the total vocabulary based on parts of speech.

After obtaining the data, we can find that the minimum vocabulary coverage rate of the two sets of textbooks at the 1st and 4th word level is 92%. The proportion in the "New Edition" is 82.75%, that is, 4.16% of the first word level + 21.33% of the second word level + 9.04% of the third word level + 48.22% of the fourth word level, and the proportion of "Comprehensive" is 82.05. If all the parts of speech in the two sets of textbooks are counted, only "Comprehensive" can reach more than 4000. On the surface,
"New Edition" and "Comprehensive" basically meet the vocabulary requirements of the textbook text in the outline. However, the vocabulary coverage of the two sets of textbooks is low, which makes the overall teaching and learning work more difficult. More importantly, the proportion of the two sets of textbooks in the 4th word level has increased significantly. It can also be seen from this that in order to ensure that students master more vocabulary within a limited time, the textbook designers have virtually abandoned 3000 high-frequency words in the text, mainly for the first collocation and the third word level. If the textbook does not provide corresponding language input for high-frequency word learning, only students who have just met the entrance requirements of the 'Syllabus' will find it difficult to reach another level of teaching after one semester. The correct and proficient application of 2000 to 2500 basic word collocations Requirements. English majors can also increase their exposure to high-frequency words through other professional courses and extracurricular reading. However, for some students with poor foundations, the text with low coverage of high-frequency words and vocabulary cannot meet the basic learning foundation, which is more difficult [5].

4.1.2. Ideal Angle Analysis
When teaching, the staff should also start from the perspective of vocabulary coverage. The results show that the two sets of textbooks studied in this article are more suitable for students with more than 3000 vocabulary to carry out language learning activities. From the perspective of total vocabulary alone, "Comprehensive" is in a state of compliance.

This type of textbook can combine the English vocabulary level of high school graduates with the vocabulary requirements of the "Outline" in the process of writing. More importantly, the "New Standards for Middle School English Curriculum" requires high school students to master 3000 words and 400 to 500 idioms after graduation. In contrast, most graduates who meet this requirement are concentrated in economically developed regions. The specific teaching level in many regions of China has not yet reached this standard.

4.2. Vocabulary Recurrence Rate Statistics and Vocabulary Learning Opportunities
Related research shows that if students encounter new words in reading, they will have a great impact on vocabulary learning. If you encounter it only once, it will not produce any learning opportunities. When the number of times increases, the potential learning opportunities increase greatly. Researcher S. Rott said that there is a high probability of learning a new word after six repeated occurrences. R. Waring et al. indicated through research that non-purpose learning requires at least 20 recurrences to learn new words. The textbook text is mainly based on classroom teaching, which can achieve purposeful learning. For this reason, the statistical group distance of the recurrence rate of this study is set to 10 or more. It can also be seen from the statistics of the recurrence rate that the lexical measurement unit mainly adopts the word family rather than the part of speech in the previous research question. Generally speaking, there are mainly prefixes and inflections in the word family. When students learn the words in a word family, they can also learn other words in the word family. For example, after learners learn to be safe, they can also understand safe, safer and unsafe in text reading. Therefore, in the study of statistical input corpus vocabulary recurrence rate, word family is the best measurement unit. According to the previous data analysis, the high proportion of word categories and word families are concentrated in the 4th word level. As a result, people need to make statistics on the recurrence rate of the two sets of textbooks at the word-level word family to make it clear whether the text of the textbook can provide students with appropriate vocabulary learning opportunities [6].

When carrying out this research, the quantitative analysis of vocabulary of "New English Course" and "Comprehensive Course" was mainly based on the application of Range software to test the vocabulary coverage rate and word family recurrence rate. By comparing the research results with the vocabulary teaching requirements in the "Syllabus", we can clarify the text vocabulary characteristics of the two sets of textbooks. The results show that the proportion of intermediate frequency vocabulary has been increased in the two sets of textbooks, which is in line with the first-level requirements of vocabulary teaching in the "Syllabus" for English majors. However, due to the low recurrence rate, relying solely on
the text of the textbook cannot meet the vocabulary requirements in the Outline. Simultaneously, the vocabulary coverage rate of high-frequency words is very low, and the fourth word level is relatively high, which proves that the vocabulary teaching entrance requirements in the "Syllabus" have been ignored. This is also an overestimation of students' enrollment level, which makes it difficult for students with poor foundations to grow. When teaching, in addition to designing learning opportunities for low and medium recurrence vocabulary, teachers should also provide students with some high-frequency word input corpus learning materials. So as to realize the effective consolidation of students' learning foundation. The existence of this kind of situation has increased a great teaching burden for educators, making teaching efficiency drastically reduced. However, if you analyze it from the perspective of vocabulary analysis, a lot of vocabulary exercises are designed after the "Comprehensive" unit, which can facilitate the development of teachers' teaching work.

4.3. Case Analysis
Range software can only recognize plain text files during work. Educators need to organize the two assignments, name and save them separately. When the Range software is running, we need to click the text box and enter the saved file name. In addition, the Range software can perform in-depth analysis of the word frequency in the text. It can clarify the frequency and amplitude in alphabetical order, and analyze it in the order of word frequency. Range software can also generate study analysis files from the analysis results, which can be opened in the form of "notepad". The analysis of actual vocabulary distribution should point out that those words in the basic vocabulary list exceed three basic words. The words involved in this student work are mainly distributed in the first two units. It can also be seen from the comparison that the 1000 words commonly used in the exercise and the academic words are far from mastered. To this end, educators must also speak clearly about the frequency and distribution of words and get specific information.

5. Conclusion
To sum up, the actual corpus in the two sets of textbooks selected in the study of this article needs to be improved. For this reason, in the follow-up research, the staff also need to prepare two sets of textbooks to understand the differences between the textbooks and to ensure a good connection between the textbooks. In the meantime, in the research of actual English major textbooks, quantitative grading should provide directions for subsequent textbook data and teaching reform research.

References
[1] Liu Yuting. Research on the Application of Educational Game Software in College English Teaching——Taking Monopoly English Learning Platform as an Example [J]. Computer Products and Circulation, 2019(03):213.
[2] Wang Ban. On the application of WeChat software in college English teaching [J]. Campus English, 2019(09): 35.
[3] Zeng Xin, Zhang Nan. The role and function of "English interesting dubbing" software in teaching aids——Take the oral teaching of college English as an example [J]. Journal of Chengdu Normal University, 2019, 35(01): 48 -51.
[4] Liu Dongmei. On the use of vocabulary memory software in college English vocabulary teaching [J]. Talent, 2019(01): 102-103.
[5] Zhou Jinxiang. The application of word APP software in college English vocabulary classroom teaching[J]. Think Tank Times, 2018(41):264-265.