

# A Review of Vocabulary Research in English Textbooks Based on Corpora

Shuxian Li

Received: 03 Jul 2024; Received in revised form: 05 Aug 2024; Accepted: 12 Aug 2024

©2024 The Author(s). Published by TheShillonga. This is an open-access article under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>)

## Abstract

*With the promulgation of high school New Curriculum Standards, various new textbooks have emerged, and the presentation of knowledge content for vocabulary that has a profound impact on English learning in these textbooks has also changed. Corpus linguistics often uses corpus software as a research tool to efficiently search and extract language patterns from numerous language examples. The research process is objective and the results are scientific. Therefore, more and more scholars are using corpus tools to study vocabulary in English textbooks. This study will summarize the current status of vocabulary research in English textbooks based on corpus and look forward to the development trends in this field. The study aims at revealing the distribution and development patterns of vocabulary in various English textbooks, provide inspiration for future textbook compilation, and provide reference for future vocabulary evaluation research in English textbooks.*

**Keywords—** *Corpus based, English textbooks, vocabulary, review*

## I. INTRODUCTION

Without grammar, there is very little content that can be expressed; without vocabulary, nothing can be expressed" (Wilkins 1972). Cunningsworth (2002) also stated that vocabulary knowledge is the key to mastering grammar knowledge. Indeed, for any language, if one wants to understand, explain, read, and write English contents, vocabulary as the basic language unit is an important foundation, and solid learning and accumulation of it is the first step in building the English language building.

In today's era of intelligence and networking, the channels for students to acquire and master vocabulary have been greatly expanded. From multimedia platforms to various books, students can have first-hand access to vocabulary information. However, in today's educational environment, for the vast majority of Chinese English learners, they cannot have a real and authentic language environment in the process of English acquisition, and the input channels for English vocabulary are limited. In this situation, English classrooms are the main way for learners

to acquire vocabulary, and textbooks are the most important source of vocabulary input. Therefore, the arrangement, selection, and presentation of vocabulary in textbooks have a profound impact on the vocabulary acquisition and even the entire language acquisition of English learners in China. Since the 1980s, with the widespread use of corpora, more and more scholars have used corpora to study vocabulary, and corpora have become a key basis for evaluating vocabulary in textbooks. Corpus, which can compare a large number of textbooks with reference corpora, has become an important tool for vocabulary evaluation in textbooks. It has the characteristics of objectivity, efficiency, and comprehensiveness. Under the theoretical guidance of systems science, using corpora to analyze vocabulary from discourse is an inevitable choice for language research and foreign language teaching (Li Xiuhong, Ding Ge, 2022).

Traditional vocabulary research often focuses on language learners, exploring the breadth and depth of their vocabulary. However, there is currently a limited amount of vocabulary research in the academic community that

focuses on textbooks, and a recognized framework for textbook vocabulary research has not yet been found. However, through reading literature, it is found that the study of textbook vocabulary is fundamentally based on both breadth and depth. The research on the vocabulary breadth of foreign language textbooks by domestic and foreign scholars focuses on the amount of vocabulary presented in the textbooks and the coverage rate with the reference corpus (Kim 2017). Therefore, this study will review the vocabulary research of previous corpus-based English textbooks from the perspectives of vocabulary breadth and depth, providing reference significance for the development of related textbook vocabulary research in the future.

## **II. A REVIEW OF RESEARCH ON THE VOCABULARY BREADTH OF ENGLISH TEXTBOOKS BASED ON CORPORA**

Breadth of vocabulary knowledge, also known as vocabulary quantity. Nation (2003) proposed that one dimension of vocabulary knowledge is vocabulary breadth knowledge. Ma Guanghui (2016) pointed out that in the field of language acquisition research, vocabulary breadth refers to the number of vocabulary that learners master in one language. He Anping (2009) pointed out that breadth research aims to understand whether textbooks provide the most basic vocabulary forms, and this research can be understood and analyzed by studying the formal coverage of textbook vocabulary on corpus basic words and curriculum standard vocabulary. Through this vocabulary breadth study, researchers can investigate whether English textbooks implement the curriculum objectives and concepts, as well as the degree and way in which textbooks scientifically and reasonably present high and low frequency words in English. Therefore, this chapter will focus on exploring previous research on the breadth of vocabulary presented in textbooks, mainly involving the quantity and frequency of vocabulary presented in textbook texts, as well as the repetition of word frequency tables related to native speakers.

### **2.1 Research on the Quantity of Textbook Vocabulary Based on Corpus**

In terms of studying the number of vocabulary in textbooks, Kim et al. (2017) used corpus tools to analyze

the latest vocabulary in high school English textbooks in South Korea and North Korea. The research results indicate that the vocabulary of the two textbooks is similar, but the number of different forms and variants of vocabulary in Korean English textbooks is twice that of Korean textbooks. Moser (2020) examined the five most popular foreign language textbooks in Arabia, clustering vocabulary using the MADAMIRA word form analyzer and comparing the top 3000 commonly used words in Arabic dictionaries with textbook vocabulary using AntWordProfiler. The study found that the number of vocabulary in textbooks for each word frequency band is relatively small. Yang and Coxhead (2020) used a corpus to study the vocabulary of New Concept Textbooks. This study found that although learners may encounter most high-frequency vocabulary in the book New Concept English, they need to know 3000 to 6000 words clusters to process these texts, and the fourth volume of the book requires 1000 more words clusters than the third volume to understand texts. The science related texts in this textbook have more vocabulary because they contain a large portion of low-frequency words that have only been used once and have a higher vocabulary than humanities texts. And there is an opportunity to learn mid-frequency vocabulary in this textbook.

### **2.2 Research on the Coverage of Textbook Vocabulary to Other Reference Corpus Based on Corpus**

Scholars are also enthusiastic about studying the coverage of textbook vocabulary to other reference corpora. At the end of the last century, Ljung sorted out the top 1000 high-frequency words in Swedish high school English textbooks and the COBUILD large general corpus. After comparison, it was found that most of the top 1000 words in the high school English textbook were semantically specific vocabulary, while in contrast, the top 1000 words in the COBUILD corpus were mostly semantically abstract vocabulary. On this basis, Ljung explored the missing vocabulary in the high school English textbook and found that those semantic abstract vocabulary that was not presented or insufficiently presented in the textbook were a very commonly used vocabulary in real communication, rather than unfamiliar and obscure vocabulary. This study suggests that high school English textbook compilation should include a larger proportion of non-narrative discourse in order to be consistent with the actual language

used (Ljung 1990). In addition, it also found that the vocabulary in this set of textbooks did not show a significant progressive relationship in terms of difficulty with grade growth.

Similar to Ljung's approach of using a universal large corpus for vocabulary comparison, Coniam (2004) compared early stage English textbooks in Hong Kong to the Bank of English (BoE), a large universal corpus. However, unlike Ljung who separately compiled the frequency of vocabulary presentation in textbooks and corpora, Coniam mainly used word frequency information from large corpora as a reference for evaluating vocabulary in textbooks. And, unlike Ljung who focused on missing high-frequency words in textbooks, Coniam paid more attention to the low-frequency words presented in the textbooks. He found that about one-fifth of the words in the textbooks are low-frequency words, which are less commonly used by native speakers. Through further analysis of the detected low-frequency words, Coniam found that the difficulty level of English textbooks in Hong Kong is slightly inverted. He believes that for some important low-frequency words such as home, nose, etc., they should be learned in lower grades, while for less common low-frequency words such as ticket, they should be learned in higher grades. Coniam's research findings indicate that while textbooks should present high-frequency words as much as possible, it should also be noted that the selection of vocabulary in textbooks should not be based on word frequency as a criterion.

In addition, Koprowski also applied the large universal corpus BoE to vocabulary research in 2005, but he focused more on vocabulary phrases in textbooks. By conducting a survey by placing textbook vocabulary and phrases in a large universal corpus, Koprowski found that about 14% of vocabulary and phrases in the textbook are not commonly used, and some even never appear in the reference corpus. Therefore, the selection of vocabulary and phrases in the textbook is subjective and experiential, and some phrases do not help improve students' language communication abilities.

In recent years, Norberg and Nordlund (2018) have studied vocabulary in seven Swedish primary school English textbooks. By comparing the vocabulary in the textbooks with those in the New General Service

vocabulary list and VP Kids corpus, they found that the textbooks presented a large number of low-frequency words that occasionally appear in everyday language. Rahmat (2021) used the Range corpus to study the vocabulary characteristics of Indonesian high school English textbooks, and found that the textbooks contain a large number of high-frequency words, accounting for 80% of the vocabulary in the Range corpus. Basaran (2022) examined the vocabulary of 30 German Foreign Language (GFL) textbooks based on a corpus and compared the core vocabulary with the top 2384 words in the vocabulary frequency table. The research results indicated that all core vocabulary in the textbook covered the top 2384 most frequently used words in the word frequency table. Nakayama (2022) examined the vocabulary of the new Japanese junior high school English textbook and compared the textbook vocabulary with a New Common Vocabulary List (NGSL) consisting of 2801 high-frequency words from regular English using corpus tools. The study found that this series of textbooks mainly consisted of vocabulary from NGSL, with a coverage rate of over 95%. However, it only covered a small portion of NGSL, with a coverage rate of less than 37%. In China, Zhao Yong (2003) referred to the BNC corpus to examine the core vocabulary of New Horizons College English, and found that 100% of the texts in volumes one to four covered the core vocabulary specified in the syllabus. Zhang Wei and Ma Guanghui (2007) referred to the frequency information of the general corpus, using the built-in vocabulary in the Range corpus software as a reference vocabulary, and compared the textbook "Experimental Textbook English for Compulsory Education Curriculum Standards (New Objectives)" with it. They found that high-frequency and low-frequency words that appeared in the large general corpus were presented in large quantities in the textbooks they studied. However, due to the wide coverage of the three reference vocabulary tables in the software, which differed from the high school English curriculum standards and actual learning needs, the presentation of these words in the textbook cannot guarantee a high degree of conformity with the curriculum standards vocabulary and high-frequency words. In addition, the study also found that some vocabulary in the textbook has a lower frequency of presentation and a narrower distribution span. However, it should be noted that it is not

reasonable for this study to use word frequency information from adult corpora as a reference corpus for vocabulary surveys in elementary school textbooks, as these English textbooks in the middle and low learning stages will consider the corresponding cognitive abilities and communicative contexts of the readers during preparation. Therefore, the word frequency of vocabulary in the textbooks is limited, and it is more appropriate to use appropriate peer reference corpora for comparative analysis.

In addition to referencing word frequency information from large general corpora, curriculum standard vocabulary is also an important reference corpus for vocabulary breadth research. For example, scholars Zhou Jialin and Li Qingsang (2013) compared the vocabulary in the People's Education Press and Foreign Language Research Press versions of high school English textbooks with the old curriculum vocabulary. The study found that about 93% of the old curriculum vocabulary was presented in the People's Education Press and Foreign Language Research Press versions of high school English textbooks, which means that these two sets of textbooks strictly follow the old curriculum. However, about one-third of the vocabulary in both textbooks is non-standard vocabulary, and some of the standard vocabulary exhibits low reproducibility and narrow distribution span. He Anping (2009) studied the coverage of vocabulary in the 2007 edition of the People's Education Press textbook to the Curriculum Standard and large corpora. The results showed that the textbook was consistent with both the Curriculum Standard vocabulary and the basic vocabulary extracted from several large general corpora. This study also indicated that corpus based methods for investigating and evaluating textbooks can help demonstrate the language characteristics of textbook compilation through a large amount of empirical data, and the optimization of textbook compilation and in-depth improvement of teaching can get some insights from these studies. Xie Jiacheng (2010) used Wordsmith corpus retrieval software to investigate the vocabulary presentation of primary and high school English textbooks. Research has found that many of the curriculum standard vocabulary and basic vocabulary in the primary and high school textbooks had a low frequency, especially the eighth-level vocabulary in the curriculum standard. Wang Xiaona (2018) compared and studied the coverage and distribution of vocabulary in

the Shanghai Education Oxford edition junior high school English textbook to the People's Education edition junior high school English textbook based on the vocabulary list specified in the new curriculum standard. She found that from the data of Range, the People's Education edition textbook was more in line with the requirements of the new curriculum standard's level five vocabulary than the Shanghai Education Oxford edition textbook. From the data of Wordsmith Tools 4.0, it can be seen that the transition between different textbook versions is smoother in the Shanghai Education Oxford version than in the People's Education Press version.

From the above research, it can be found that the research ideas in China on the breadth of vocabulary knowledge in textbooks is basically consistent with foreign research ideas, both focusing on the frequency of textbook vocabulary and the coverage of high-frequency vocabulary on large general corpora. But in addition, domestic research has also paid more attention to the coverage of textbook vocabulary on the curriculum standards. This indicates that the vocabulary research perspective in domestic textbooks is relatively comprehensive, representing that Chinese textbooks not only focus on authentic English acquisition, but also on accurate grasp of the curriculum, demonstrating the orderly development and scientific progress of English textbook compilation.

### **III. A REVIEW OF RESEARCH ON THE DEPTH OF VOCABULARY IN ENGLISH TEXTBOOKS BASED ON CORPORA**

A large general corpus is equally important for the in-depth study of vocabulary in textbooks. By referring to the semantic, grammatical, and collocation information of vocabulary in the large general corpus, the scientific and rational presentation of vocabulary depth in textbooks can be tested. The breadth research and depth research of vocabulary knowledge not only have similar paths, but also have a close relationship with each other, which can provide a research foundation for each other. Regarding the concept of deep vocabulary knowledge, vocabulary teaching expert Nation (2001) pointed out that vocabulary knowledge includes the following aspects in both receptive and productive dimensions: pronunciation, spelling, part of speech, semantic association, grammatical form,

collocation, and contextual constraints. Sinclair (2004) pointed out that different collocation words can form lexical items to represent basic meanings, including lexical collocation, grammatical collocation, semantic preference, and semantic rhyme. Scholar Ma Guanghui (2016) believes that vocabulary depth refers to the degree or quality of a learner's mastery of second language words, that is, the learner's mastery of multiple information and features of one word. In textbooks, vocabulary depth mainly refers to the collocation, grammar, semantics, context, etc. of the vocabulary involved in the textbook.

Although domestic and foreign scholars have different expressions of the concept of vocabulary depth knowledge, they all believe that vocabulary depth knowledge should include multiple aspects of vocabulary, such as form, semantics, grammar, collocation, context, etc. In addition, the "Curriculum Standards for General High School English (2017 Edition, Revised in 2020)" point out that students should understand the connotation and extension of specific word meanings in context, and learn the habitual collocation and expression of verb phrases. Therefore, in-depth research on vocabulary in English textbooks is of great significance for teacher teaching. This chapter mainly summarizes previous research on the depth of vocabulary knowledge presented in textbooks, which mainly involves the grammar collocation, vocabulary collocation, semantic preference, and semantic rhyme of some typical vocabulary in textbooks.

### **3.1 Research on the Vocabulary Difficulty of English Textbooks Based on Corpus**

In terms of the research on vocabulary difficulty in English textbooks, Lee (2008) used the vocabulary analysis function in NLP-TOOLS and three word lists to analyze the distribution of vocabulary in English textbooks for foreign college students. The study found that the vocabulary distribution in textbooks was between 12 grade textbooks and English short stories, and the difficulty of vocabulary increased from the first volume to the fourth volume. In China, Chen Xiaoxiao (2011) selected the Brown corpus as a reference corpus to study the vocabulary distribution and presentation of the New Horizon college English textbook. The author found that there was no significant difference in the overall distribution of vocabulary between the textbook and the native language corpus, but it clearly showed a trend

of easy first and then difficult, which was more in line with the law of vocabulary learning; The distribution of specific words in the text showed the same trend, initially similar to the native language. As learning progresses, the number of specific words in the text is significantly higher than in the native language corpus; The proportion of words appearing 5 to 15 times or more in the vocabulary of both corpora is relatively low, and the proportion of words appearing 5 to 8 times in the textbooks is even less than that in the native language corpus, which is not conducive to students' vocabulary learning. Kim and Lee (2017) studied the vocabulary by using high school English IA-II textbooks, the College Academic Ability Test (CSATS) English test, and EBS materials. The research results indicated that there were significant differences in vocabulary levels among the three corpora, and the vocabulary difficulty of EBS materials was higher than that of CSAT and textbooks. The author believed that EBS materials may bring excessive learning burden to students, and suggested maintaining a balance of vocabulary in various materials. Based on the theory of dynamic systems, He Anping (2015) used AntConc corpus software to examine the breadth and depth of the "MAKE" verb in English textbooks from primary school English textbooks to university English textbooks in China. The study found that the dynamic development trajectory of this word in English textbooks from primary school to university reflected the characteristics and rules of its depth of knowledge from simple to complex and from concrete to abstract. Song Xiaozhou (2016) used the Wordsmith corpus retrieval tool and found that the difficulty of vocabulary in various volumes of the Comprehensive Tutorial did not show a regular increase, but the overall difficulty of vocabulary was moderate. Wang Xiaona (2018) examined the Oxford and People's Education editions of junior high school English textbooks. She compared two sets of textbooks using Wordsmith and Range corpus retrieval software and found that the growth rate of type and token between the textbooks in the People's Education Press was faster than that in the Oxford edition. However, the transition of the Oxford edition of textbooks was smoother than that in the People's Education Press. Huang Kun (2018) used AntConc and Range corpus software to analyze the presentation of vocabulary in the Oxford edition of high school English textbooks. The study found that the

difficulty of vocabulary in textbooks did not show a trend from simple to difficult, which is slightly different from previous results. Previous research on textbooks has shown a trend of vocabulary difficulty developing from easy to difficult, which may be due to differences in the training objectives of the textbooks. But the grammar paradigm presented in the textbook conforms to daily norms. Chen Anni and Guo Aiping (2019) examined the difficulty level of the New Horizon College English textbook. They used the Coh Metrix corpus to analyze textbooks and found that as the number of the vocabular in college English textbooks gradually increases, the difficulty of vocabulary also increases significantly. In addition, Tang Meihua and Liang Maocheng (2021) conducted a study on the lexical complexity of college English textbooks, and the research results showed that the lexical complexity between each textbook reflected the principle of gradual progression. However, the textbooks need to be improved in distinguishing the complex gradients between each textbook, and further efforts are needed to be done to increase the lexical complexity step by step. Mek (2021) studied the vocabulary content of Türkiye A1 foreign language textbook based on the corpus. The author extracted the most common 1000 words in the textbook and compared them with the word frequency table generated by Aksan. The research results indicated that the similarity of vocabulary content presented by the two was not ideal, and the presentation of nouns in textbook vocabulary content was not sufficient. The diversity of adjective and verb content needs to be improved.

### **3.2 Research on whether English textbook vocabulary can present the most commonly used semantics and typical usage**

Sinclair&Renouf (1988) conducted a survey on the depth of vocabulary knowledge in textbooks and found that the high-frequency usage of certain vocabulary in textbooks does not match their usage in real communication, such as give, see, have, make, take, etc. Similarly, Ute Rmer (2005) drew similar conclusions by comparing the grammatical and semantic usage of vocabulary in multiple versions of German English textbooks with relevant corpus information from a large general corpus. It was found that these textbooks did not present the most commonly used words and collocation of grammatical words in the native language

context.

Bowles (2000) further explored the dimension of semantic frequency, which belongs to vocabulary depth, based on vocabulary breadth analysis. They used CCED, a dictionary based on semantic frequency for semantic arrangement, as a reference corpus to compare the semantic information presented by some textbook vocabulary with it. It was found that the semantics presented of some vocabulary in the textbook were not commonly used semantics in the corpus. For example, for some multi-semantic vocabulary, the semantics commonly used in the corpus were not or were less presented in the textbook, and instead, the very few used semantics were more presented. And this may be due to the lack of clear display of the semantic frequency of vocabulary in the teaching syllabus, resulting in the deviation of vocabulary semantics from the focus of the textbook. Therefore, Bowles suggested that in the process of textbook compilation and vocabulary teaching, a large amount of empirical data provided by corpora (such as vocabulary frequency, semantic frequency, collocation frequency, etc.) should be used to actively verify the scientific and authentic nature of vocabulary cataloging or explanation. However, in this survey, Bowles only selected the first level textbook in each edition and was limited by manual data processing, which is somewhat subjective and failed to fully examine the lexical semantic presentation throughout the entire series of textbooks.

Xie Jiacheng (2008) conducted a deep analysis of vocabulary knowledge in corresponding Chinese high school English textbooks based on data obtained from the breadth analysis, covering multiple aspects such as lexical semantics, grammatical collocation, and word collocation. The first step is to sample high-frequency vocabulary in the textbook, extract several high-frequency basic words that appear in the textbook, and then use corpus software to retrieve the vocabulary information presented in the textbook. The retrieved information is compared and analyzed with commonly used semantics, grammatical collocations, and word collocations in a large general corpus. The survey found that there are few or even no commonly used semantics and collocations of some high-frequency basic words in textbooks, especially multi part of speech vocabulary, which is particularly evident. Similar to Xie Jiacheng (2008), He Anping (2009) also adopted a

corpus research method for in-depth research, but it was not based on breadth analysis. Instead, he directly selected three different parts of speech words: verb "come", noun "way", and grammatical word "that" from Thornbury's list of the 100 most commonly used basic English vocabulary to investigate their performance in vocabulary semantics, grammatical collocation, and word collocation in the 2007 edition of the People's Education Press High School English textbook. The research results showed that the textbook presented a typical collocation paradigm of "come" well; In contrast, although "way" frequently appears in textbooks, it is more commonly used in the instruction language of practice part, lacking specialized activities and exercises to reflect and summarize the usage form and context of this most basic word; The number of times "that" is used as an antecedent of a clause is significantly higher than the number of times it is used as a demonstrative pronoun, which reflects the increased difficulty of learning high school English vocabulary and grammar.

Ma Li (2018) used the Range corpus software to examine the vocabulary of the Foreign Language Research Edition high school English textbook. Research has found that this textbook only presents a certain part of speech for words with multiple parts of speech, while ignoring other parts of speech. In addition, a large number of basic vocabulary has been presented in textbooks, but nearly 10% of vocabulary has only been presented once. Li Yahong (2020) examined the vocabulary presentation in the compulsory high school English textbooks of the New Teacher's Press. Research has found that some vocabulary in textbooks did not present the commonly used basic meanings and collocation patterns in dictionaries, and textbooks tended to present a certain part of speech or meaning of words with multiple parts of speech.

The above corpus-based vocabulary research has revealed the common drawback of textbook vocabulary, which is that the most commonly used semantics and collocations of high-frequency basic words in textbooks are less presented, and textbooks always tend to present a certain part of speech or meaning of words with multiple parts of speech.

### **3.3 Research on Corpus based Vocabulary Collocation in English Textbooks**

Context is an essential factor in vocabulary learning,

and the collocation and use of vocabulary to some extent depend on context. Therefore, vocabulary collocation is an important part of deep vocabulary knowledge. Due to the consideration of context, this part of the research usually adopts a research approach from overall to individual cases, selecting a common word, usually a verb, summarizing its usage in the context through corpus tools, and comparing it with relevant corpora of the mother tongue. He Anping and Liang Jianli (2009) compared Chinese high school English textbooks with foreign CCEC textbooks, and the research results showed that the presentation of basic vocabulary, high-frequency verb vocabulary, and grammatical collocation in domestic textbooks was basically consistent with foreign textbooks. Xie Jiacheng (2010) conducted a corpus-based survey on the in-depth knowledge of the verb "do" in two sets of domestic and two sets of foreign high school English textbooks. Research has found that two sets of English textbooks abroad present multiple virtualization usage of this word, while Chinese textbooks do not. Tang Jieyi (2015) used a corpus to study the in-depth knowledge of the vocabulary "take" in college English textbooks. She found that the vocabulary, grammatical collocations, and synonyms related to "take" presented in the textbooks were sufficient, showing diverse paradigms and collocations. Li Lin and Li Chengxin (2021) used Lancsbox corpus software to study the vocabulary presentation in business English textbooks. A study found that 80 high-frequency professional vocabulary words have a frequency of over 1000 times, showing certain characteristics in vocabulary distribution and collocation. Xia Jing (2021) used the Range corpus to study the presentation of vocabulary in the New Education Press high school English textbook. The results indicated that most textbooks only present a common meaning and collocation for vocabulary with multiple parts of speech, without presenting other meanings or paradigms.

Through the above research, it can be found that the approach to in-depth research on textbook vocabulary at home and abroad is basically the same. They all conduct in-depth research by comparing the similarities and differences in grammar, semantics, paradigms, and collocations between textbook vocabulary and large-scale general corpus vocabulary. This demonstrates the important position of corpora in in-depth vocabulary research. The compilation of textbooks, including its themes, functions,

structures, and tasks remains the main driving forces for this field of research. However, in the process of optimizing them, corpus tools and a large amount of empirical resources should be fully utilized to make the language of the textbooks as natural and authentic as possible.

#### IV. SUMMARY

In summary, the use of corpus tools has gradually become an important means for domestic and foreign scholars to evaluate textbook vocabulary. Based on corpora, domestic and foreign scholars have conducted detailed research on the breadth and depth of English textbook vocabulary in different countries, age groups, and versions. Now, the above research is summarized:

In terms of research subjects, scholars at home and abroad have covered English textbooks for different stages from elementary school to university (Norberg&Nordlund 2018; Ma Li 2018; Kim&Lee 2017), specifically in primary school English textbooks, middle school English textbooks, high school People's Education Press, foreign research edition, Oxford edition of English textbooks, new concepts, business English, college comprehensive English, and new perspectives of college English. However, expert He Anping (2015) used the theory of dynamic systems and used the verb "make" as an example to investigate and analyze English textbooks in various stages from elementary school to university in China, showing a trend of "one-stop" vocabulary development.

In terms of research methods, most domestic and foreign scholars have adopted corpus based methods, using different corpus retrieval software such as AntConc, Range, Lanksbox, WordSmith, etc. (He Anping, 2015; Xia Jing, 2021; Li Lin, Li Chengxin, 2021; Song Xiaozhou, 2016). By closely relying on the vocabulary information of large-scale general corpora through corpus tools, sufficient and reasonable corpus basis is provided for research results. Meanwhile, domestic textbook vocabulary knowledge research also focuses on reference to curriculum standard vocabulary, reflecting the rigor and timeliness of textbook research.

From the perspective of research methods, these studies are mainly divided into two categories: comprehensive analysis of vocabulary knowledge in a certain textbook and comparative analysis of vocabulary

knowledge in different textbooks.

In terms of research content, scholars both domestically and internationally have mainly focused on the breadth and depth of vocabulary in textbooks. However, it can be found that domestic scholars are enthusiastic about combining the breadth and depth of vocabulary in textbooks for research. In terms of vocabulary breadth in textbooks, domestic and foreign scholars mainly focus on the number of vocabulary in textbooks, coverage of vocabulary with other reference corpora, and word frequency; In terms of vocabulary depth, research mainly focuses on the vocabulary typical collocations, vocabulary difficulty, vocabulary distribution, vocabulary complexity, and presentation of typical semantics and paradigms in textbook vocabulary. In addition, their focus is not only on the overall vocabulary, but also on certain words and parts of speech, such as the study of a certain word such as "make", "take", "do" (He Anping, 2015; Tang Jieyi, 2015; Xie Jiacheng, 2010), and the study of the presentation of verbs and adjectives (Li Xiaoyu, 2018).

#### V. IMPLICATION

Based on literature review, there are still some shortcomings in the research of vocabulary in textbooks both domestically and internationally:

From the perspective of research objects, due to being in the early stage of using new textbooks, there is still a small amount of research on vocabulary knowledge in the revised curriculum standards in China, and the vocabulary knowledge of multiple versions of English new textbooks still needs to be further studied. From a research perspective, existing literature often focuses on the breadth or depth of vocabulary knowledge, or on certain aspects of depth. The comparative dimension is not comprehensive, making it difficult to provide comprehensive and feasible suggestions for English teachers in China.

From the perspective of research framework, there are few researchers in existing literature who have systematically organized the research dimensions of vocabulary in English textbooks, and vocabulary knowledge research has not yet formed a scientifically reasonable research framework.

Subsequent scholars may be able to expand the research on corpus based textbook vocabulary from the



above aspects.

### REFERENCES

- [1] Bowles M. A review of the legal content and its treatment in the registry approved level one EFL textbooks used in Japanese public lower secondary schools [D] Unpublished MA Thesis, 2000
- [2] Basaran, B. Word Frequency and Vocabulary for the Language Demands of Textbooks [J] *International Journal of Curriculum and Instruction*, 2022 (1): 1029-1051
- [3] Cunningsworth, A. *Choosing Your Coursebook* [M] Oxford: Heineman, 1995
- [4] Coniam, D. A Legal View of the Language Needed by Primary Students in Hong Kong: An Examination of Word Frequency in English Language Primary Materials [J] *CUHK Journal of Primary Education*, 2004 (5): 75-84
- [5] Koprowski, M. Investigating the Usefulness of Legal Phrases in Contemporary Courses [J] *Elt Journal*, 2005 (59): 322-332
- [6] Kim, J.&Lee, D. J. A Corpus Based Study on the Use of Vocabulary in High School English IA · II Textbooks, College Scholastic Ability Tests, and EBS Materials [J] *Linguistics Science*, 2017b (4): 51-74
- [7] Kim, J., Je Young, L.&Kim, J. R. A Synchronous Lexical Analysis of the South and North Korean English Textbooks [J] *Language Sciences*, 2017a (2) 25-56
- [8] Ljung, M. *A Study of TEFL Vocabulary* [M] Stockholm: Almqvist&Wiksell, 1990
- [9] Lee Y. An Analysis of English Textbooks for University Students Using a Corpus Tool and Wordlists [J], *Humanities Studies*, 2008 (2), 125-149
- [10] Moser, J. Evaluating Arabic Textbooks: A Corpus Based Lexical Frequency Study [J] *International Journal of Applied Linguistics*, 2020 (2): 248-263
- [11] Mek, R.&Mesut, G. N. A Corpus Based Research: The Vocabulary Content and Parts of Speech of A1 Level Textbooks Used in Teaching Turkish as a Foreign Language [J] *Journal of Language and Linguistic Studies*, 2021 (1), 236-267
- [12] Nation I. S.P. *Learning Vocabulary in Another Language* [M] New York: Cambridge University Press, 2001
- [13] Nation, I. S. P. Materials for Teaching Vocabulary [A] In B. Tomlinson (ed.) *Developing Materials for Language Teaching* [C] London Continuum, 2003
- [14] Norberg, C.,&Nordlund, M. A Corpus based Study of Lexis in L2 English Textbooks [J] *Linguistics, Education*, 2018 (3): 463-473
- [15] Nakayama, S. Vocabulary in Japanese EFL Textbooks: A Bidirectional Coverage Analysis [C]// The IAFOR International Conference on Education, Hawaii 2022:157-169
- [16] Rahmat, Y. N.&Coxhead, A. Investigating Vocabulary Coverage and Load in an Indian EFL Textbook Series [J] *Indian Journal of Applied Linguistics*, 2021 (3): 804-814
- [17] Sinclair, J.&Carter, R. *Trust the Test: Language, Corpus and Discovery* [M] London: Routledge, 2004
- [18] Sinclair, J. M.&Renouf, A. A Lexical Syllabus for Language Learning [A] In Carter&McCarthy (eds.) *Vocabulary and Language Teaching* [C] London: Longman, 1988
- [19] Ute Romer *Progressives, Patterns, Pedagogy: A Corpus Driven Approach to English Progressive Form, Functions Contexts, and Didactics* [M] Amsterdam: John Benjamin, 2005
- [20] Wilkins, D. A. *Linguistics in Language Teaching* [M] London: Edward Arnold. 1972
- [21] Yang, L.&Coxhead, A. A Corpus Based Study of Vocabulary in the New Concept English Textbook Series [J] *RELC Journal*, 2020 (1), 1-15
- [22] Chen Anni, Guo Aiping. A Study on the Difficulty of College English Textbooks Based on Coh Metrix: A Case Study of the Third Edition of New Horizon College English [J]. *Education Theory and Practice*, 2019, 39 (06): 46-48. (陈安妮,郭爱萍.基于 Coh-Metrix 的大学英语教材难易度研究——以《新视野大学英语》第三版为例[J].教育理论与实践,2019,39(06):46-48.)
- [23] Chen Xiaoxiao. Empirical Study on the Distribution Characteristics of Vocabulary in College English Textbooks: A Comparative Analysis between New Horizons and Brown Corpus [J]. *Electronic Teaching of Foreign Languages*, 2011 (02): 9-14. (陈潇潇.大学英语教材课文词汇分布特征实证研究——《新视野》与 Brown 语料库的对比分析[J].外语电化教学,2011(02):9-14.)
- [24] He Anping. Corpus Assisted Analysis of Basic English Textbooks [J]. *Curriculum, Textbooks, Teaching Methods*, 2007 (03): 44-49. (何安平.语料库辅助的基础英语教材分析[J].课程.教材.教法,2007(03):44-49.)
- [25] He Anping. Survey on the Breadth and Depth of Vocabulary in basic English Textbooks [J]. *English Teacher*, 2009, (06): 3-6. (何安平.基础英语教材词汇的广度和深度调查[J].英

- 语教师,2009,(06):3-6.)
- [26] He Anping. A Study on the Development of Vocabulary Knowledge in English Textbooks: A Dynamic Systems Theory Perspective [J]. *Foreign Language Teaching and Research*, 2015,47 (06): 898-908+961. (何安平.英语系列教材的词汇知识发展研究:动态系统论视角[J].外语教学与研究,2015,47(06):898-908+961.)
- [27] Huang Kun. *A Corpus based Study on Vocabulary in High School English Textbooks*[D]. East China Normal University, 2019. (黄堃.基于语料库的高中英语教材词汇研究[D].华东师范大学,2019.)
- [28] Jiang Beilei. *A Corpus based Survey on the Breadth and Depth of Verbs in High School English Textbooks*[D]. Yanbian University, 2023. (姜蓓蕾.基于语料库的高中英语教材中动词的广度和深度调查研究[D].延边大学,2023.)
- [29] Li Jun. On the Relationship between the Depth and Breadth of Vocabulary and Reading Comprehension [J]. *Foreign Language Teaching*, 2003 (02): 21-24. (李俊.论词汇的深度和广度与阅读理解的关系[J].外语教学,2003(02):21-24.)
- [30] Wang Hao, Liu Yongbing. Research on Vocabulary Teaching Activities in Middle School English Textbooks [J]. *Curriculum, Textbooks, Teaching Methods*, 2016,36 (10): 75-80. (王昊,刘永兵.中学英语教材的词汇教学活动研究[J].课程.教材.教法,2016,36(10):75-80.)
- [31] Liang Jianli, He Anping. A Corpus based Study on the Breadth and Depth of Vocabulary in the New High School English Textbook [J]. *Journal of the School of Foreign Languages, Shandong Normal University (Basic English Education)*, 2009,11 (02): 9-16. (梁健丽,何安平.基于语料库的《新高中英语》教材词汇的广度和深度研究[J].山东师范大学外国语学院学报(基础英语教育),2009,11(02):9-16.)
- [32] Li Lin, Li Chengxin. A Corpus based Study on Common Business English Professional Vocabulary [J]. *Journal of Foreign Languages*, 2021, (04): 64-69. (李琳,李称鑫.基于语料库的常用商务英语专业词汇研究[J].外语学刊,2021,(04):64-69.)
- [33] Li Xiuhong, Ding Ge. Frequency Analysis and Application of Meaningful Words in Reading Discourse under the Theory of Systems Science [J]. *Journal of Systems Science*, 2023 (01): 130-136. (李秀红,丁鸽.系统科学理论下阅读语篇实义词频分析与应用[J].系统科学学报,2023(01):130-136.)
- [34] Li Yahong. *A Corpus based Study on Vocabulary in New Education Press High School English Textbooks*[D]. Shanxi Normal University, 2021 (李亚虹.基于语料库的新人教版高中英语教材词汇研究[D].山西师范大学,2021.)
- [35] Ma Guanghui. *English Vocabulary Teaching and Research*[M]. Beijing: Foreign Language Teaching and Research Press, 2016. (马广惠.英语词汇教学与研究[M].北京:外语教学与研究出版社,2016.)
- [36] Ma Li. *A Corpus based Study on Vocabulary in Foreign Language Research Edition High School English Textbooks*[D]. Southwest University, 2019. (马丽.基于语料库的外研版高中英语教材词汇研究[D].西南大学,2019.)
- [37] Li Xiaoyu. *Corpus based Vocabulary Analysis of Middle School English Textbooks* [D]. Central China Normal University, 2020. (黎小雨.基于语料库的初中英语教材词汇分析[D].华中师范大学,2020.)
- [38] Song Xiaozhou. A Corpus based Vocabulary Study of English Professional Textbooks [J]. *Education Review*, 2016, (08): 123-126+130. (宋晓舟.基于语料库的英语专业教材词汇研究[J].教育评论,2016,(08):123-126+130.)
- [39] Tang Jieyi. A Corpus based Study on Depth of Vocabulary Knowledge in English Textbooks for Higher Education Institutions [J]. *Higher Education Exploration*, 2015, (01): 81-86. (唐洁仪.基于语料库的高等院校英语教材词汇深度知识研究[J].高教探索,2015,(01):81-86.)
- [40] Tang Meihua, Liang Maocheng. Analysis of the Differences in Vocabulary Complexity in College English Textbooks [J]. *Frontiers of Foreign Language Education Research*, 2021, 4 (01): 61-68+90. (唐美华,梁茂成.大学英语教材词汇复杂度级差分析[J].外语教育研究前沿,2021,4(01):61-68+90.)
- [41] Wang Xiaona. *A Corpus based Comparative Study of Vocabulary in Shanghai Education Oxford Edition and People's Education Edition Junior High School English Textbooks*[D]. Shaanxi Normal University, 2018. (王筱娜.基于语料库的沪教牛津版和人教版初中英语教材词汇对比研究[D].陕西师范大学,2018.)
- [42] Wang Xinyu. *Research on the Breadth and Depth of Vocabulary Knowledge in the New High School English Textbook of People's Education Press based on Corpus*[D]. East China Normal University, 2023. (王新瑜.基于语料库的人教版高中英语新教材词汇知识广度和深度研究[D].华东师范大学,2023.)
- [43] Xie Jiacheng. Investigation of Vocabulary Corpus in Middle School English Textbooks [J]. *Foreign Language Teaching in Primary and Secondary Schools (Middle School Chapter)*, 2008,31 (01): 4-8. (谢家成.中学英语教材词汇语料库调

- 查[J].中小学外语教学(中学篇),2008(01):4-8.)
- [44] Xie Jiacheng, He Anping. Corpus Research and Motivation Exploration of Appendix Vocabulary in Middle School English Textbooks [J]. *Foreign Language Teaching in Primary and Secondary Schools* (Middle School Chapter), 2008,31 (09): 1-5. (谢家成,何安平.中学英语教材附录词汇表的语料库研究与理据探究[J].中小学外语教学(中学篇),2008,31(09):1-5.)
- [45] Xie Jiacheng. A Comparative Study of Delexicalized Verb in English Textbooks Based on Corpus: Taking the Delexicalized Verb "do" as an Example [J]. *Foreign Language Teaching Theory and Practice*, 2010 (03): 13-17. (谢家成.基于语料库的英语教材虚化动词对比研究——以虚化动词“do”为例[J].外语教学理论与实践,2010(03):13-17.)
- [46] Xie Jiacheng. Analysis of Vocabulary Corpus in Middle School English Textbooks [J]. *Foreign Language Teaching Theory and Practice*, 2010 (01): 55-61. (谢家成.中学英语教材词汇语料库分析[J].外语教学理论与实践,2010(01):55-61.)
- [47] Xia Jing. *A Corpus based Study on English Textbook Vocabulary*[D]. Tianjin Normal University, 2021. (夏静.基于语料库的英语教材词汇研究[D].天津师范大学,2021.)
- [48] Zhang Yanlin, He Anping. Corpus based High School English Vocabulary Learning [J]. *Foreign Language Teaching in Primary and Secondary Schools* (Middle School Chapter), 2006,29 (08): 7-12. (张彦琳,何安平.基于语料库的高中英语词汇学习[J].中小学外语教学(中学篇),2006,29(08):7-12.)
- [49] Zhang Wei, Ma Guanghui. Vocabulary Analysis of English (New Goals) Textbook [J]. *Foreign Language Teaching in Primary and Secondary Schools* (Middle School Chapter), 2007,30 (01): 9-13. (张伟,马广惠.《英语(新目标)》教材词汇分析[J].中小学外语教学(中学篇),2007,30(01):9-13.)
- [50] Zhao Yong, Zheng Shutang. Focus on Core Vocabulary in College English Textbooks [J]. *Foreign Language and Teaching*, 2003 (06): 21-24. (赵勇,郑树棠.大学英语教材中的核心词汇的关注[J].外语与外语教学,2003(06):21-24.)
- [51] Zhou Jialin, Li Qingshen. A Comparative Study of Vocabulary in High School English Textbooks [J]. *English Teachers*, 2013,13 (01): 66-70. (周加林,李庆燊.高中英语教材词汇对比研究[J].英语教师,2013,13(01):66-70.)
- [52] Ministry of Education of the People's Republic of China. *Curriculum Standards for Ordinary High School English* (2017 Edition, Revised in 2020) [M]. Beijing: People's Education Press, 2020. (中华人民共和国教育部.普通高中英语课程标准(2017年版2020年修订)[M].北京:人民教育出版社,2020.)