Lexical Cohesion and Corpus Construction in Primary and Secondary School English Teaching

Haiming Zhou

School of Foreign Language, Yancheng Normal University, Yancheng, China

Qin Liu

School of Foreign Language, Yancheng Normal University, Yancheng, China

Abstract

English learning in primary and secondary schools is the critical period of language acquisition, which is very important for developing students' language ability. Among them, vocabulary teaching cohesion is the basis of English learning, which affects students' learning at different stages to a great extent. In the process of English learning, if the teaching vocabulary cohesion is not reasonable, it will not only reduce students' learning efficiency, but also affect students' ability training in listening, speaking, reading and writing skills. This paper studies the lexical cohesion in primary and secondary school English teaching and constructs the corresponding lexical corpus.

Keywords: Lexical Cohesion, Corpus, Primary and Secondary School, English Teaching

1. Introduction

In the basic stage of English teaching in primary and secondary schools, textbooks are the core part of English curriculum resources and important content and means for students to learn and teachers to teach. As an important cornerstone of English textbooks (Meli, 2020; Benstead, 2021; Acp, 2021), vocabulary is the basic language knowledge that primary and secondary school students must master. No matter listening, speaking, reading and writing, they cannot do without the support of vocabulary. The English curriculum standards for primary and secondary schools make specific requirements for vocabulary learning at different stages and include multi-level vocabulary lists as teaching reference. The classification of vocabulary can meet the cognitive ability of learners of different ages, accord with the gradual development of English language learning ability of students in compulsory education stage, and ensure the

integrity, continuity and gradual progress of English curriculum.

English teaching materials are the main content of English vocabulary teaching. The cohesion of vocabulary in English teaching materials for primary and secondary schools is the basic condition to make the cohesion of English vocabulary teaching in primary and secondary schools have real practical significance (Aino, 2020). The fluency of vocabulary connection between different teaching stages will greatly affect the integrity, continuity and gradual progress of English courses in primary and secondary schools, as well as the learning effect of students. If the vocabulary of primary school and middle school is not connected properly, it will be reflected in students' learning to a certain extent, which makes students' vocabulary foundation is not solid and hinders the improvement of English learning ability (Weicht, 2020; Purwanti, 2020; Larsen, 2021; Hogan, 2020). However, in actual teaching, even under the guidance of the same curriculum, English learning from primary school to junior high school cannot completely meet the requirements of integrity, continuity and gradualism due to the use of different textbooks.

This paper will study and analyze the vocabulary cohesion of different versions of English textbooks by constructing a vocabulary corpus, and put forward vocabulary lists for different English learning stages according to the research content, aiming to judge the rationality of textbook selection in primary and secondary English teaching and put forward reasonable suggestions for textbook selection. At the same time, the research of this paper will help teachers to make a more reasonable arrangement of vocabulary teaching in English textbooks and ensure the cohesion of vocabulary teaching in different grades.

2. Vocabulary Cohesion

The main task of language learning is to acquire vocabulary and gradually master it in use. Textbook vocabulary refers to the sum of all the words and fixed phrases covered in the textbook. Vocabulary is the basic unit and material of language, which needs to be paid close attention to in second language learning (Ralli, 2021; Parashchuk, 2021; Kirigin, 2021). It greatly affects the quality of classroom teaching, if you do not learn enough vocabulary, it is difficult to accurately understand the text, more can not learn the sublimation of the thought of the article. In this paper, the teaching material mainly refers to vocabulary teaching materials in the textbooks of reading and listening text covers all the words, according to the text, vocabulary teaching outline as foundation, according to the requirements of teaching outline for different stage of learners will be from the perspective of the database will be teaching in vocabulary and comparing with the syllabus vocabulary, to understand the teaching material in the vocabulary difficulty, And understand the analysis of the vocabulary coverage in the textbook, discuss whether the current primary school textbooks are reasonable.

Cohesion meant to make it easy for people to know whether the words were a whole or just a collection of separate language sentences when they saw or heard chapters longer than one sentence in length. Cohesion can be divided into grammatical cohesion and lexical cohesion. Lexical cohesion is the anaphoric cohesion established by the structure of words, including lexical repetition and lexical co-occurrence, which mainly refers to the mutual relationship of words in the discourse (Rodrigues, 2021; Butler, 2021; Giomi, 2021).

This paper discusses the lexical cohesion of primary and secondary school students, which

extends from the internal relations of texts to the lexical cohesion of textbooks. That is to say, in order to achieve the teaching objectives of vocabulary, the vocabulary of the two stages of primary and secondary school students should connect and "transition" naturally, and become a whole that can be easily understood and learned step by step, under the premise of unified teaching standards. So that students can smoothly realize the continuation of English learning from primary school to junior high school.

3. Lexical Corpus

The construction of lexical corpus should mainly consider five issues, namely, the overall design and planning of lexical corpus, the collection of words, the coding system of lexical corpus, the processing level of lexical corpus and the processing tools of lexical corpus.

3.1 Design of Lexical Corpus

First of all, it is necessary to have an overall design and planning for the constructed corpus, including the application objectives, scale, implementation steps, quality of the corpus, internal consistency, scalability, software and hardware compatibility, etc. Among them, the determination of corpus scale plays a key role in the application value of corpus. The function of a corpus depends on the distribution and size of the stored corpus (Moser, 2020). Because the distribution of corpus directly affects the scope of application of statistical results, and the size of database determines the reliability of statistical data. However, bigger is not always better. A corpus with tens of millions of words can obtain a highly reliable word frequency standard and meet most research needs based on word frequency. However, a corpus with a scale of hundreds of millions of words is not more important for research than one with a scale of tens of millions of words. With the rapid development of Internet technology, information in all fields increases exponentially, and the speed of searching and collecting corpus has been greatly improved by using advanced computer programs. In view of the huge amount of Internet text data, the scale of the corpus in this study should be more than ten million words, and the specific scale depends on the time span of corpus collection.

3.2 Vocabulary Collection

Corpus collection should mainly consider the way of obtaining corpus, the data format of corpus file, the size of corpus file, the selection criteria of corpus and the proportion of various texts in corpus. Among them, it is particularly critical to determine the selection criteria of corpus, because it is directly related to the representativeness and scientificity of the constructed corpus. The seven principles of corpus collection include authenticity, reliability, scientificity, representativeness, authoritativeness, distribution and fluidity. It should be noted that the representativeness of corpus sampling is relative and largely depends on the research purpose.

3.3 Coding of Lexical Corpus

When the corpus reaches a certain scale and develops to a certain stage, it will involve the problems of resource sharing and data file exchange in different computer environments. Therefore, for the convenience of use, the annotation methods of corpus files in corpus should follow uniform norms. Standard Generalized Mark-up Language (SGML) is the most ideal text markup format in corpus. At present, the representative Corpus Encoding system based on SGML is CES (Corpus Encoding Standard). CES is a specialized corpus coding standard

co-sponsored by EAGLES (Language Engineering Standards Advisory Committee), MULTEXT (Multilingual Text Tools and Corpus Project) and other organizations of The European Computational Linguistics community.

3.4 Processing Levels of Lexical Corpus

The processing level of corpus should be determined according to different usage requirements and specific application tasks. The processing levels of corpus mainly include part-of-speech tagging, syntactic tagging, semantic tagging, discourse coreference tagging and so on. With the development of corpus technology and computer technology, the level of corpus annotation develops from coarse-grained to fine-grained, so as to better serve all kinds of language research and natural language processing tasks.

3.5 Processing Tools of Lexical Corpus

Building a corpus is a time-consuming and laborious project, so it is necessary to use software tools as a tool for corpus processing. In the process of building the library, the frequently used tools include text collator, word restoration tool, word segmentation and part-of-speech tagging tool, syntax and semantic analyzer and so on. With the development of computer technology, there are natural language processing toolkits that combine the above tools. At present, NLTK (Natural Language Toolkit) and Stanford CoreNLP are widely used. The former, developed by Steven Bird and Edward Loper of the University of Pennsylvania, is a set of natural language processing tools based on the Python programming language. It can be used for cleaning, encoding, syntactic and semantic analysis of natural language. The latter is a set of natural language processing tools developed by the natural language processing team led by Christopher Manning of Stanford University. This tool can be used to restore word form, part-of-speech tagging and syntactic structure tagging.

4. Analysis of Lexical Cohesion in Textbooks

This paper analyzes the total amount of vocabulary in primary school textbooks, the presentation of vocabulary in primary school textbooks and the coverage of vocabulary in primary school textbooks respectively.

4.1 Vocabulary Characteristics of Primary School Textbooks

Through the text analysis data of primary school textbook vocabulary, we can see that the total amount of textbook vocabulary is large, and it occupies a large proportion in the first-level vocabulary and has a high coverage rate of basic vocabulary. In general, it covers most of the basic vocabulary, there are many basic words, although there are some difficult words, but its proportion is maintained within a certain range. On the other hand, words have a higher repetition rate. Repetition is one of the important means of vocabulary learning, and two characteristics should be considered, namely the number of repetitions and repetition interval. The repetition of the same words in primary school textbooks reduces the difficulty of learning vocabulary to a certain extent, but it is also necessary to pay attention to the frequency of repetition. Excessive repetition of words is also not conducive to vocabulary learning.

Through the comparison with the second-level vocabulary in the syllabus, it is found that if the textbook uses most of the extra-standard vocabulary, it will increase the difficulty of students' learning, thus affecting their ability training in listening, speaking, reading and writing skills, and also increasing the cultivation of emotional attitude.

4.2 Vocabulary Characteristics of Middle School Textbooks

In order to meet the characteristics of students' learning stage, the vocabularies in the first grade textbooks should occupy a large proportion in the first level vocabulary, and the smallest proportion in the third level vocabulary. In other words, there is more use of basic words and less use of more difficult words in the tertiary vocabulary. In this way, it is beneficial for students to establish confidence and interest in English learning in grade one, and to understand and use language to express topics. At the same time, the word family in the textbook should be more abundant, because it reflects the change and derivation of vocabulary. In order to ensure that the teaching materials contain rich word families, teachers are required to summarize the changes and derivational relations based on this word with students in addition to the basic words in the teaching word list. For example, the teaching word learn is required to teach its noun and verb forms. Learn, learn, learn, learn, learn, and learn. Therefore, students are required to have certain associative memory in vocabulary learning.

4.3 Vocabulary Cohesion in Primary and Middle School Textbooks

Vocabulary repetition in textbooks helps students to deepen their memory, but the high repetition rate makes some students relax their learning tension, which is not conducive to the improvement of students' higher level. The vocabulary in primary school textbooks should have a certain coverage rate in junior high school textbooks. In terms of the overall coverage rate, the words in primary school textbooks are generally the be verb, personal pronoun, demonstrative pronoun, and commonly used prepositions that repeat the most frequently in the textbooks of grade one, which ensures that after entering grade one from primary school, we should not only learn new words, but also have certain repeating words.

5. Conclusion

Textbook vocabulary is a basic and important part of English teaching. If the vocabulary of the textbook is too large, the classroom and learning burden will be increased for a long time. It is easy for students to blur the focus of learning and reduce the learning efficiency. At the same time, the teaching burden of the teacher is too heavy, so that the book teaching task cannot be completed in the given time. Therefore, from the perspective of textbook compilation, attention should be paid to the connection between different stages in textbook compilation.

The disconnection and lack of cohesion in the content of textbooks will greatly affect the learning development of learners. Therefore, according to the requirements of the teaching syllabus, the textbook should be selective to choose vocabulary, so that teachers can do a good job in the reasonable connection of vocabulary teaching. Therefore, it is necessary for the teaching system of each stage to communicate and understand each other, and confirm the teaching content and key points of each stage according to the outline and the law of cognitive development of students at each stage, so as to avoid too much unnecessary repetition of vocabulary learning, so that the knowledge structure of students can reach the spiral rise.

Reference

Acp, A., Gchr, A., Htb, A. et al. (2021). Strengthening the teaching of the narrative genre: story and fable in primary school children in the Department of Magdalena – Colombia. A commitment to the use of ICT Games and Bayesian Logistic Regression. *Procedia Computer Science*.

Aino, A., Kirsikka, K., Aino, A. et al. (2020). Learning and Teaching Experiences with a Persuasive Social Robot in Primary School–Findings and Implications from a 4-Month Field Study.

Benstead, H. (2021). Primary initial teacher trainees' perceptions of and perspectives on the role of the teaching assistant in English primary contexts. *Support for Learning*. https://doi.org/10.1111/1467-9604.12384

Butler, C. S., & Simon-Vandenbergen, A. M. (2021). Social and Physical Distance/Distancing: A Corpus-Based Analysis of Recent Changes in Usage. *Corpus Pragmatics*. https://doi.org/10.1007/s41701-021-00107-2

Giomi, R. (2021). The Place of Interpersonal Lexemes in Linguistic Theory, with Special Reference to Functional Discourse Grammar. *Corpus Pragmatics*. https://doi.org/10.1007/s41701-020-00094-w

Hogan, J. A., Valverde-Barrantes, O. J., Ding, Q. et al. (2020). Morphological variation of fine root systems and leaves in primary and secondary tropical forests of Hainan Island, China. *Annals of Forest Science, 77*(3). https://doi.org/10.1007/s13595-020-00977-7

Kirigin, T. B., & Perak, B. (2021). Lexical Sense Labeling and Sentiment Potential Analysis Using Corpus-Based Dependency Graph.

Larsen, S. A., Little, C. W., & Coventry, W. L. (2021). Exploring the Associations Between Delayed School Entry and Achievement in Primary and Secondary School. *Child Development, 92*(2). https://doi.org/10.1111/cdev.13440

Meli, B M. (2020). Qualities of an Ideal Effective Teaching for Primary and Secondary School Teachers in Tanzania.

Moser, J. (2020). Evaluating Arabic textbooks: A corpus-based lexical frequency study. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.12321

Parashchuk, V., Yarova, L., & Parashchuk, S. (2021). Automated Complexity Assessment of English Informational Texts for EFL Pre-service Teachers and Translators. *Arab World English Journal*. https://doi.org/10.24093/awej/call7.11

Purwanti, E., & Setiyadi, H. (2020). Online education program policies toward teacher teaching ability and learning process in primary school.

Ralli, A. M., Kazali, E., Kanellou, M. et al. (2021). Oral Language and Story Retelling During Preschool and Primary School Years: Developmental Patterns and Interrelationships. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-021-09758-3

Rodrigues, H., Nyberg, E., & Coheur, L. (2021). Towards the benchmarking of question generation: introducing the Monserrate corpus. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-021-09545-5

Weicht, R., Ivanova, I., & Gikopoulou. O. (2020). The CRADLE teaching methodology: developing foreign language and entrepreneurial skills in primary school pupils. *Entrepreneurship Education, 3*(3). https://doi.org/10.1007/s41959-020-00036-2

Copyright Disclaimer