Research Article

Design of English Teaching Corpus Resource-Sharing System Based on XML Technology

Li Bing

School of Foreign Languages and Literature, Shan Dong University, Jinan, Shandong 250100, China

Correspondence should be addressed to Li Bing; kevinone@tu.edu.vn

Received 25 January 2022; Revised 16 March 2022; Accepted 21 March 2022; Published 30 April 2022

Academic Editor: Naeem Jan

Copyright © 2022 Li Bing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is well known that vocabulary is the cornerstone of language learning. Therefore, vocabulary teaching is very important in English language teaching. However, time-consuming and inefficient vocabulary education has always been a long-standing issue. Teachers utilize traditional teaching methods, which are monotonous, so students lose interest and desire in studying, resulting in a drop in academic performance and even resistance to English courses, creating a vicious circle. Therefore, improving the efficiency of vocabulary teaching has become a hot topic in the field of English education today. Corpus technology has begun to infiltrate the realm of English education as the times and technology have progressed. Because of its vast capacity and real corpus, the corpus is gradually being used in English classes. Therefore, this research will explore the application of corpus-based XML technology data-driven teaching in junior high school English vocabulary teaching. Data-driven vocabulary teaching based on corpus resources has been shown to considerably improve students’ vocabulary learning effects, but its promotion requires the assistance of education departments as well as the development of instructors’ professional theoretical understanding. In order to solve the problem of data processing and data sharing of heterogeneous data in the process of digital campus construction, this study aims at the campus network application system, using data exchange technology and XML’s good extensibility, self-description, separation of form and content, etc. It realizes data sharing and sharing in heterogeneous databases.

1. Introduction

With the rapid development of modern educational technology, corpora with big data concepts and technical characteristics have played an important role in the field of language teaching and research. Corpus linguistics has grown into an emerging school of interdisciplinary interaction between modern education, linguistics, and computer science. With the characteristics of large corpus reserves, authentic sources, representativeness, and easy query and analysis, the corpus has become an indispensable resource and tool for language research and language teaching. Therefore, scientific and rational use of corpus resources in college English teaching is not only an inevitable trend of modern educational technology development but also an endogenous demand for college English teaching reform.

The formula language is essentially a multiword language unit, which widely exhibits in Japanese and written expressions in the form of fixed collocations. The use of procedural language can shorten the time for language processing and processing in the mind, so using procedural language in the actual pragmatic environment can make the expression more accurate and idiomatic and make the communication more fluent [1–4]. At present, procedural language has become an important content in English teaching at all levels. Since procedural languages usually show a high co-occurrence frequency in corpora, it has natural advantages to use the corpus to carry out procedural language teaching. This study mainly analyzes the types of English corpora and explores their retrieval skills [5, 6]. On this basis, it discusses the application of corpus resources in English programming language teaching and provides a reference for English teachers to carry out programming language teaching.

Corpus-based foreign language teaching in China started relatively late. In 2004, some scholars constructed the idea of corpus translation teaching through the methods of corpus index co-occurrence dynamic context, text equivalent
probability analysis, translation style quantification, and multiversion comparative research. In the same year, some scholars demonstrated the application value of parallel corpora in translation teaching. In 2008, some scholars conducted a preliminary analysis on the research status of corpus-based English translation textbooks and translation teaching [7–9]. It should be pointed out that the above research mainly focuses on translation teaching, and most of them are based on self-built small corpora. There are certain limitations in terms of authority, corpus richness, and representativeness, so it is difficult to achieve expectations. In recent years, with the development of network and big data technology, the types of corpora have become more abundant, the scale has also grown rapidly, and its application in the field of language teaching and research has been more extensive and in depth [10, 11].

The corpus can be divided into heterogeneous, homogeneous, systematic, and specialized according to different content and attributes, but the division of these four types is too rough and not suitable for teaching. The study tends to divide the types of corpora into multiple dimensions from the perspective of teaching, so as to facilitate teachers to select teaching resources. For example, the corpus can be divided into Japanese corpus, written language corpus, native speaker corpus, learner corpus, monolingual corpus, multilingual corpus, general corpus, special corpus, etc. It should be noted that this multidimensional division is overlapping, and a corpus can belong to multiple types at the same time [12, 13].

Corpus retrieval techniques can be divided into confirmatory retrieval and exploratory retrieval. Among them, the purpose of the confirmatory search is to verify whether a certain multiword collocation is commonly used and authentic, that is, to verify whether the multiword collocation is a formula language. The purpose of exploratory retrieval is to discover new English lexical collocations or sentence structures, that is, to explore what collocations contain certain words and whether there are formulas in them. In procedural language teaching, the two methods are often used together [14–17]. A procedural language is a common phenomenon of multiword collocation in language expression. The acquisition of procedural language is conducive to improving pragmatic ability and making language expression more fluent and authentic. Not only the modern corpus system with the characteristics of big data can provide rich corpus resources but also its flexible and convenient statistical analysis function is beneficial to the teaching of procedural language [18, 19]. Using the exploratory retrieval and verification retrieval skills of the corpus can not only reduce the pressure of teachers to find real corpus when preparing lessons but also arouse learners’ interest and enthusiasm in learning with detailed and vivid corpus. The corpus resources and retrieval skills introduced in this study have certain reference values for English programming language teaching [20–22]. Considering that Chinese English learners are accustomed to checking procedural languages in Chinese, research on procedural language teaching based on an English-Chinese bilingual parallel corpus can be carried out in the future.

What digital campus needs is a brand-new open resource sharing environment so that it has rich information and excellent service capabilities. Therefore, it is imperative to solve the problem of heterogeneous data exchange and information sharing between campus network application systems [23, 24]. At present, the campus information management system is mainly developed in a browser/server C Brower/Server B/S structure or a client/server (Client/Server C/S) mode. The back-end database uses relational databases, such as SQL Server, ORACLE, and Sybase. Based on this campus network application system supported by a relational database, an English teaching corpus resource-sharing system based on XML technology is designed. The digital campus data exchange platform is based on XML, uses XSL/XSLT to define user views, constructs legal XML documents with XML Schena or Document Type Definition (DTD), and realizes the shared data transformation of heterogeneous data. The system is controlled by the data exchange platform management center, and the IML manager, message center console, database manager, and security mechanism are responsible for the division of labor. Each heterogeneous network has its own database system and data format, but these network function nodes can use intranet or Internet and XML-based digital campus data exchange intermediary functions to realize data conversion and sharing of heterogeneous networks. The entire system communicates in a standardized XML data format, and IML filters can process and adjust various data formats and support various protocols, such as e-mail, HTTP, and FTP [25–27].

Data exchange and sharing of diverse databases is a problem to be tackled during the creation of a digital campus. By defining and using the new XML data format language ML, this study constructs heterogeneous data into standard XML data and uses a unified data exchange format to participate in data exchange. Construct a general ML data format, process different types of data streams inside and outside the system platform, extract the main part of the relevant data in each heterogeneous database, and exchange and add data according to the actual exchange requirements, so as to realize the heterogeneous database in the digital campus [28].

The research is organized as follows. The several key issues in anticipating resource sharing are presented in Section 2. Section 3 analyzes the corpus and teaching sharing. The experimental design and analysis are discussed in Section 4. Finally, in Section 4, the research work is concluded.

2. Several Key Issues in Anticipating Resource Sharing

The data exchange center is an overall structure, and the neglect of any detailed planning may lead to problems. Therefore, in order to avoid delays or rework due to lack of consideration during the construction of the data exchange center, it is necessary to understand the following key issues.

(1) Specification of data standards: first, develop information standards and then formulate unified data
standards to establish reliable data centers. This stage means establishing the basic information coding standards for the school [29, 30].

(2) Design of the central database: after the data standard is determined, the database structure of the shared data center is designed, which is the key to building the data center. The shared database structure of the data center must be unified according to the resource type stipulated in the educational resource specification of the Ministry of Education, and various program interfaces should be reserved so that the functions of the resource center can be extended at any time.

(3) Data synchronization problem: the information exchange between the business system and the shared data center database is divided into two parts, uploading the business system data to the data center and reading business information from the data center. For data synchronization scenarios, a well-designed solution should ensure real-time, accurate information flow.

(4) Handling of abnormal data: in data exchange, there is often a mismatch of data types, such as the simplest form of a date. Actually, in SQL Server 2012, there will be two completely different data representation procedures; the first is small date-time and the second is date-time, but with Oracle 10i, there is only one way to represent dates. Although there are significant differences in data representation patterns, both database management systems can be used for everyday use. The query result in SQL Server 2012 is “2015-6-9 00:00:00,” but the data retrieved by Oracle 10i is “9-6-15,” so the data retrieved in the database are “2015-6-9.” Oracle’s default date expression pattern is dd-mm-yy, where “dd” is the day, “mm” is the month, and “YY” is the year. Usually, since there is no fixed format for dates in Oracle, functions can be used to resolve the database. This is a flaw. A specific time is entered in the TODATE function and finally specified according to the data format; usually, the characters of the XML document are changed according to the specific format. Also, empty data are different; if the database has empty values, the data are actually empty. XML documents are the same concept and can usually be displayed using element attributes. In the database, the empty element character is represented as 0, which means the character length is 0, but you should make sure to remove the empty value when mapping the database or processing the XML document; otherwise, the file is wrong. In fact, in the process of converting the database to XML document format, if the data in the database is empty, then it is indistinguishable when converting the data to the XML document. In parsing the database in the XML document when the value is set, if the analysis result is empty, the character will be defaulted to DB Null by the program.

In order to realize the sharing and exchange of heterogeneous data, it is necessary to convert the data from different data sources into a consistent format and provide a consistent global schema for the local schema of each heterogeneous data source.

3. Corpus and Teaching Sharing

With the continuous development of information technology, the role of corpus technology in English teaching has become more and more important. The real and large amount of language data in the corpus not only provides rich learning resources for language learners but also provides great convenience for various researches. Different kinds of corpora provide great support for the development of language teaching. So far, many general or specialized corpora have been built abroad, such as Brown, LOB, Cobuild, BNC, COCA, and other corpora. With the development of the times, these corpora have become more and more mature, which not only provide teachers with rich teaching resources for foreign language teaching but also play a great auxiliary role for teachers in guiding pragmatic errors and phrase collocation errors. Inspired by corpus research, many researchers put forward similar language teaching ideas during the same period. The more representative ones include Antoinette Renouf, Chris Tribble, Dave Willis, John Sinclair, Tim Johns, and Michael Lewis. Among them, some focus on theoretical fore-shadowing and some focus more on teaching practice, and there are not a few who organically combine theory and teaching practice. In terms of the acquisition and development of corpus-based teaching resources, some software tools such as AntConc, Wordsmith, and Range have been developed for language teaching. With the assistance of this software, it is greatly convenient for teachers to use the corpus for corpus utilization and processing, so as to better design teaching activities. To sum up, foreign research on the application of corpus in foreign language teaching is quite rich, and the development of corpus resources and software has also been very advanced. The organic combination of theory and software can provide a great convenience for teachers to carry out teaching activities efficiently.

In the past ten years, corpus linguistics has developed rapidly in teaching and vocabulary. The research of corpus linguistics in teaching mainly includes the theoretical research of corpus-assisted foreign language teaching and the application of corpus resources in teaching practice. Li Wenzhong and Wei Jianzhong believed that the organic combination of corpus indexing and multimedia course-ware development can give full play to the exploratory nature of DDL and the pertinence of classroom teaching. The sharing of corpus resources is important for communicative teaching and personalization. Learning compensates for the absence of pragmatic expertise and linguistic intuition among foreign language teachers in our country, as well as fixing the authenticity problem of teaching materials, and provides new ways. At the same time, the application of corpus has also proposed the innovation and
transformation of foreign language teaching ideas and methods. Xu Kuihua and Zhang Weiping pointed out that the use of corpus in teaching emphasizes the student-centered, high degree of flexibility and strong pertinence and can provide appropriate and real-time teaching according to changes in students, needs, and goals. However, at the same time, they also believe that the full application of corpus linguistics in foreign language teaching is not only a challenge to the traditional rational teaching model but also a challenge to its own development. Liang Maocheng, Li Wenzhong, and Xu Jiajin believed that the corpus’s attention to language, phrases, stylistic differences, quantitative analysis, and different analysis ideas such as collocation, class connection, semantic tendency, and semantic prosody can be applied to foreign language teaching, such as vocabulary explanation, writing commentary, practice, and preparation of test questions. In terms of specific application, He Anping pointed out that the original intention of constructing large-scale corpora is for language investigation and research, not specifically for language teaching. Due to the variety of domains or the complex annotation system, it is difficult for corpus resources to directly enter classroom teaching. Therefore, corpus resources and technologies need to be processed, including two aspects, one is the processing of corpora as teaching resources and the other is corpus processing. Processing is a teaching method. The former includes how to make the corpus into the language teaching syllabus, the compilation of teaching materials, and the design of teaching activities; the latter focuses on the implementation of technology and teaching effect evaluation of corpus-assisted classroom teaching. Liang Hongmei and He Anping, after reviewing the 9th International Symposium on Teaching and Corpus, wrote a review and pointed out that, in order to make corpus into daily teaching, first of all, the concept must change, that is, from researchers selling corpus products to teachers to researchers and teachers. Teachers jointly study how to use the corpus to solve specific teaching problems, and these teaching problems can come from both teaching reflections based on corpus investigations, or from learners’ language development characteristics and common language mistakes. Research on effect evaluation and efforts to develop teaching multimodal corpus to provide the material basis and background support is carried out. Therefore, in order to make corpus practically enter daily teaching and play its due role, we must first start from teacher education.

In terms of empirical research, many researchers have conducted research on educated people at different levels. Sun Jing applied the method of corpus and XML to the teaching of English in high school and proved that the new technology and learning method are helpful for high school students to learn vocabulary, especially in the learning of in-depth vocabulary knowledge and the long-term memory of vocabulary. Zhou Jingjings discussed the specific methods of applying corpus to college English teaching in private colleges and universities and tried to apply the corpus data-driven learning model to the teaching of deep knowledge of college English vocabulary. Through data analysis, it was found that the corpus data-driven learning model is helping college students learn vocabulary collocation and class the connection and lexical context meaning have produced an obvious effect and cultivated their autonomous learning ability. However, at the same time, it is found in the experiment that this method is not suitable for some students, so it is still necessary to combine the corpus-based vocabulary teaching method with the traditional vocabulary teaching method, so as to enhance the strengths and avoid weaknesses. To sum up, corpus linguistics has made great progress in assisting foreign language teaching in China, and research scholars represented by The Anping have more and more research results in this regard. However, it is not difficult to find that corpus linguistics is still mostly used in language research and is rarely used in actual teaching.

Different from the traditional teaching view, constructivism theory believes that learning is the process of individuals constructing their own knowledge, and students should not passively accept knowledge but should actively learn. Learning involves two aspects of construction. One is to encode new information on the basis of the original experience system to construct one’s own understanding; the other is to change the original knowledge due to the entry of new experience. Therefore, from the viewpoint of constructivism, the teacher’s teaching is to impart knowledge on the surface, but it is actually to promote the students to construct their own knowledge. Constructivist scholars believe that learners should actively participate in the construction of knowledge, and the learner’s own experience will affect the understanding of new information.

In teaching reform, constructivism proposes to let students learn through problem-solving. According to this point of view, teachers should set up different groups in the classroom, and each group should complete the tasks and problems assigned by the teacher as a unit. In this process, students’ thinking activities are stimulated, and they have a deep understanding of the knowledge structure. Constructivism also advocates situational teaching. Situational teaching means that teachers should provide a real situation in the classroom that is similar to real life. In this situation, teachers guide students to explore and solve problems, which can not only greatly stimulate students’ enthusiasm and initiative in learning but also cultivate students’ ability to discover and solve problems. Constructivism also values cooperative learning. Constructivism believes that everyone constructs an understanding of things based on their own experience, so different people have different understandings of things, so there is no single correct understanding of things. Cooperative learning can help students see the understanding of things beyond their own experience and can cultivate their ability to look at problems comprehensively and form a more comprehensive understanding of things. In the use of corpus for English vocabulary teaching, teachers set target vocabulary and encourage students to make bold assumptions and positive thinking based on a large number of real corpus resource materials provided in groups. In group cooperative learning, experience collision, build a new
knowledge system, and acquire target vocabulary faster and more efficiently.

4. Experiments’ Design and Analysis

The corpus resources selected for this study are mainly from the COCA corpus, which is the largest balanced English corpus in the world. The corpus collected by the corpus includes five aspects: oral language, novels, magazines, newspapers, and academics. In December 2015, the COCA corpus was augmented with corpora from July 2012 to December 2015, reaching a total vocabulary size of 5.2 billion. The corpus has a wide range and sufficient quantity. The COCA corpus has five display methods, namely, List, Chart, Collocates, Compare, and KWIC o List. As the name implies, it is a list display, which can query the expression of different forms of a single word in a sentence. The display interface of List is shown in Figures, showing the year, classification, source, and specific context of the corpus collection, respectively. The Chart interface mainly displays the word frequency statistics charts of the searched words in different categories and different time periods. The Collocates interface mainly displays the collocation of the searched word and other words. You can select the position of the collocated word from the keyword according to the needs of the searcher. The Compare interface allows you to compare the usages and phrase collocations of the searched word with another word. KWIC (Key Word in Context) is the key interface used in this study. After inputting the target word, the contextual co-current Concordance is displayed as formulas (1)–(3):

\[
\text{AVG}_{\text{context}} = \frac{\sum_{i=1}^{n} C_i}{n},
\]

\[
C_i = \text{mean}(C_1, C_2, C_3, \ldots, C_m),
\]

\[
\text{mean}(m) = \frac{(m^n + m^b + m^i)}{3}.
\]

In order to understand the teachers’ understanding and views on the corpus-based vocabulary teaching method, an interview outline is specially prepared. The interview outline is drawn from the teachers’ vocabulary teaching methods, the current confusions encountered in vocabulary teaching, and the understanding and emotional attitude of corpus-based vocabulary teaching methods. The interviewees are five teachers. The personal information of the teachers is as follows. In order to protect the personal privacy of the teachers, the names are omitted and represented by letters. In the design of this study, teachers have two main tasks in the teaching process, material preparers, and student facilitators. From the perspective of material preparers, teachers should first select key vocabulary according to the teaching content of this lesson. The determination of keywords can depend on the teacher’s own personal experience or can be selected from the COCA corpus according to word frequency items:

\[
P_{\text{student}} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_{Fi}},
\]

\[
X_i = \max(X_{Fj1}, X_{Fj2}, X_{Fj3}, \ldots, X_{Fjn}),
\]

\[
X_{Fi} = \sum_{j=1}^{n} X_{FPj}.
\]

The second step is to search the corpus of keywords from the COCA corpus. However, it does not mean that it can be directly applied to the classroom for teaching because the number of corpus screened by the corpus is huge and the difficulty is different. Directly giving the materials to students will affect students’ understanding of sentences because some of the corpora are too difficult. Understand so that the next steps cannot be carried out. Therefore, teachers should process the corpus from the corpus and the current corpus according to the actual learning level of the students. Finally, the teacher can choose whether to put it on the PPT or print it as a handout according to the teaching content and the specific actual situation as:

\[
G_{English} = \frac{\sum_{i=1}^{n} G_j * X_j}{\sum_{j=1}^{n} X_{Fj}}.
\]

From the perspective of the student’s guide, the corpus-based vocabulary teaching method encourages students to observe independently, put forward hypotheses, and finally verify the hypotheses and summarize them. In this process, teachers are not the main body of teaching activities, but the guides who help students find and solve problems. In the teaching design, the author abandoned the traditional “3P” teaching steps, namely, present, practice, produce, and chose the “OHI” teaching steps, namely, observe, hypothesize, and induce. After completing a semester of teaching experiments, the author arranges the students in the experimental class to complete the pre-made questionnaire and informs the students before completing the questionnaire that the questionnaire does not affect their grades and can only be filled out according to the actual situation of the individual. To ensure the scientificity and accuracy of the questionnaire data, we can see (6)–(8):

\[
T&S_i = \left( \frac{t_1}{k} \frac{t_2}{k} \cdots \frac{t_l}{k} \right),
\]

\[
t_i = \frac{s_i}{\sum_{i=1}^{n} s_i},
\]

\[
\text{Teacheri} = \left( \frac{t_1}{k} \frac{t_2}{k} \cdots \frac{t_l}{k} \right).
\]

Figure 1 shows that, when compared to general vocabulary teaching method, 36 students enjoy the corpus-based vocabulary way of teaching, accounting for 80% of the total students, and 5 students choose the corpus-based vocabulary teaching method, which properly accounted for 11% of the total student population, and 4 classmates feel no
sense and cannot tell, accounting for 8.9% of the total number of students. Therefore, on the whole, most of the students in the experimental class have a positive attitude towards the new corpus-based vocabulary teaching method.

As shown in Figure 2, it can be found that, in the classroom atmosphere, compared with the general vocabulary teaching method, 39 students feel that the corpus-based vocabulary teaching method is more interesting, reaching 86%, and the corpus-based vocabulary teaching method is more boring. There are 6 classmates, reaching 13%, and no classmates who have no feeling about the corpus-based vocabulary teaching method choose this option. Therefore, most of the student’s evaluation of the classroom atmosphere of the corpus-based vocabulary teaching method is relatively positive, and they feel that the teaching method is more interesting than the traditional vocabulary teaching method.

As shown in Figure 3, it can be found that, after the corpus-based vocabulary teaching method, 32 students feel more like learning vocabulary, reaching 71%, 3 students feel more disgusted with learning English vocabulary, accounting for 6.7% of the whole class, and 10 students felt unsure of their views on vocabulary learning, accounting for 22% of the class. Therefore, from the above data, it can be found that most of the students have improved their attitude towards vocabulary learning after using the corpus-based vocabulary teaching method. However, it can also be seen that some students hold a neutral view, indicating that vocabulary learning is still an obstacle for some students.

As shown in Figure 4, it can be found that, after the corpus-based vocabulary teaching method, there are 35 students who feel that their vocabulary mastery has improved, accounting for 77% of the whole class. There are 5 students, each accounting for 11% of the class. Therefore, judging from the data in the above figure, most students feel that this vocabulary teaching method can effectively improve their vocabulary mastery.

As shown in Figure 5, it can be found that, after the corpus-based vocabulary teaching method, there are 40 students who feel that their autonomous learning ability has improved, accounting for 88 of the whole class, and only 2 students who choose to decline, accounting for 88%. There are 4 students in the class, and there are 3 students who choose to be unclear, accounting for 6.7 of the class. As a result, based on the data in the graph above, the majority of students believe that, by using this vocabulary teaching approach, they will be able to increase their self-learning ability in terms of learning strategies.
According to Figure 6, it can be found that there are 13 teachers who feel that the language materials prepared by the teacher are difficult. Students accounted for 28 of the class, 4 students who felt the difficulty was low, accounting for 8.9 of the class, and 28 students who chose moderate difficulty, accounting for 62 of the class. Therefore, judging from the data in Figure 6, most students feel that the language materials processed by teachers are in line with their actual learning level, but some students still feel that the processed language materials are too difficult.

As shown in Figure 7, it can be found that there are 5 students who hope to use the general vocabulary teaching method to teach vocabulary lessons in the future, accounting for 11 of the whole class, and 9 students choose the corpus-based vocabulary teaching method as the main vocabulary lesson. The teaching form accounted for 20 of the class, and 31 students chose to combine the two teaching methods, accounting for 68 of the class. Therefore, judging from the data in Figure 7, most students feel that the language materials processed by teachers are in line with their actual learning level, but some students still feel that the processed language materials are too difficult.

As shown in Figure 7, it can be found that there are 5 students who hope to use the general vocabulary teaching method to teach vocabulary lessons in the future, accounting for 11 of the whole class, and 9 students choose the corpus-based vocabulary teaching method as the main vocabulary lesson. The teaching form accounted for 20 of the class, and 31 students chose to combine the two teaching methods, accounting for 68 of the class. Therefore, judging from the data in Figure 7, most students feel that the language materials processed by teachers are in line with their actual learning level, but some students still feel that the processed language materials are too difficult.

According to Figure 6, it can be found that there are 13 teachers who feel that the language materials prepared by the teacher are difficult. Students accounted for 28 of the class, 4 students who felt the difficulty was low, accounting for 8.9 of the class, and 28 students who chose moderate difficulty, accounting for 62 of the class. Therefore, judging from the data in Figure 6, most students feel that the language materials processed by teachers are in line with their actual learning level, but some students still feel that the processed language materials are too difficult.

As shown in Figure 7, it can be found that there are 5 students who hope to use the general vocabulary teaching method to teach vocabulary lessons in the future, accounting for 11 of the whole class, and 9 students choose the corpus-based vocabulary teaching method as the main vocabulary lesson. The teaching form accounted for 20 of the class, and 31 students chose to combine the two teaching methods, accounting for 68 of the class. Therefore, judging from the data in Figure 7, most students feel that the language materials processed by teachers are in line with their actual learning level, but some students still feel that the processed language materials are too difficult.

According to Figure 6, it can be found that there are 13 teachers who feel that the language materials prepared by the teacher are difficult. Students accounted for 28 of the class, 4 students who felt the difficulty was low, accounting for 8.9 of the class, and 28 students who chose moderate difficulty, accounting for 62 of the class. Therefore, judging from the data in Figure 6, most students feel that the language materials processed by teachers are in line with their actual learning level, but some students still feel that the processed language materials are too difficult.

As shown in Figure 7, it can be found that there are 5 students who hope to use the general vocabulary teaching method to teach vocabulary lessons in the future, accounting for 11 of the whole class, and 9 students choose the corpus-based vocabulary teaching method as the main vocabulary lesson. The teaching form accounted for 20 of the class, and 31 students chose to combine the two teaching methods, accounting for 68 of the class. Therefore, judging from the data in Figure 7, most students feel that the language materials processed by teachers are in line with their actual learning level, but some students still feel that the processed language materials are too difficult.

5. Conclusion

This paper studies the XML file format. On the basis of understanding its data format, it is successfully embedded into the English corpus system through XML technology, and it is applied in practice. However, the combined development of technology and system is a new direction of geographic information technology and information security research, and the attack methods of attackers are constantly changing, and there are still many practical problems to be solved in this field. The research scope of this study is relatively one-sided, and its own research level needs to be improved. It is necessary to further expand the geospatial data and study hard on the digital watermarking algorithm. It is believed that the research on geographic information data can also carry out research innovations in the following aspects: quantify the effect of digital watermarking technology according to the standard evaluation method and make a set of insensibility, error range, robustness evaluation, etc. System is a complete evaluation. For embedding and extracting the results of watermarking technology, there is a clear assessment criterion. On the one hand, digital watermarking technology has a clear research direction, and on the other hand, it can provide a reference range for relevant researchers to improve the performance of the algorithm. Improving the copyright security of digital products by embedding watermarks can also realize real-time monitoring of its circulation process and further strengthen the security protection work through real-time updating of the web interface. The cooperation between digital watermarking technology and the enterprise
government will provide a great convenience in terms of regional division, jurisdiction, product safety protection, etc., and the traceability of data will greatly reduce the misappropriation of violators. Designing a watermarking algorithm for a certain format of data can effectively improve its security protection. At present, only a part of the format in the geospatial data has been completed. It is hoped that, in the future, there will be opportunities to contact related projects again and conduct more comprehensive technical research. The functions of the designed application system are not perfect, and only the test version has been implemented. If the system can be popularized and used, the functions can still be expanded and applied to various platforms.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that he has no conflicts of interest.

References


