Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 14, No. 2, 2024

A Study on the Application of Online Corpora in Grammar Teaching

Tianhua Xu

Guangzhou College of Commerce, Guangzhou 511300, China

Abstract: Traditional grammar teaching in higher education often disconnects from real-life contexts or is conducted in imagined, artificial settings. Compared to traditional teaching methods, utilizing online corpora, which contain vast amounts of authentic language, offers significant advantages for teaching. This study focuses on grammar, using Larsen-Freeman's three-dimensional grammar teaching theory as the conceptual basis for instructional processing, and discusses the author's attempts to apply corpora in grammar teaching. The significance of this teaching innovation lies in enhancing students' grammatical competence and exploring the feasibility of integrating corpus concepts and technologies into instruction.

Keywords: Online corpora; Grammar teaching; Business English.

1. Introduction

The rapid development of information technology has had a profound impact on the field of education, particularly in foreign language education. The deep integration of foreign language education with information technology, along with the adoption of a blended teaching model that combines online and offline courses, provides more possibilities and opportunities for teachers and students. The "Compulsory Education English Curriculum Standards (2022 Edition)" emphasizes the importance of modern information technology in English teaching and encourages teachers to fully utilize digital technology for innovative online teaching. Innovations in teaching models not only meet the personalized learning needs of students but also promote the balanced development of compulsory education. English grammar courses often follow traditional models that disconnect from real contexts or are taught in artificial settings, failing to effectively help students construct a grammar system. Traditional grammar teaching views grammar as a knowledge system— a series of rules that learners must master, a set of static rules that emphasize accurate expression according to these rules. Modern linguists emphasize that grammar is not merely a set of static rules but a flexible tool that adapts to changing language use contexts. They argue that grammar is a dynamic and evolving capability. "Correct" grammar books do not take into account cognitive linguists' understanding of how language, cognition, and context shape meaning. Applied linguist Larsen-Freeman, based on her research in semantics and pragmatics, proposed three dimensions of language: form, meaning, and use (2024). She emphasizes that grammar should be viewed as a dynamic system in teaching, rather than as static knowledge points, and that methods for teaching grammar should never rely solely on rote memorization.

Research indicates that the application of corpora in teaching can enhance learners' motivation and improve learning outcomes (Hu Kaibao, 2024). The corpus, as a treasure trove of authentic language, is widely recognized (He Anping et al., 2022). Implementing corpus-based teaching methods is grounded in Kennedy's assertion that "the language used most frequently in language projects and processes deserves more time invested in teaching" (Kennedy, 2010, p. 281). The online corpus used in this study is based on a wealth of authentic language examples available on the internet, creating a cognitive network environment for grammar that explores the form, function, and meaning of grammatical rules in real contexts. With the support of online corpora, the three-dimensional grammar teaching theory can be better implemented in practical instruction, providing students with a more inspiring and effective language learning experience. This research is based on Larsen-Freeman's three-dimensional grammar teaching theory and explores the application of corpora in grammar teaching. The aim of this study is to enhance students' grammatical competence and explore the feasibility of integrating online corpus concepts and technologies into teaching, thereby bringing new perspectives and methods to grammar instruction.

2. Corpus and Teaching

Corpora play a significant role in advancing the learning and teaching of second languages/foreign languages. They provide learners with greater exposure to authentic language and offer important information about the distribution and central positioning of various language features in different contexts. Additionally, research on corpus-based teaching has found that engagement with corpora facilitates the development of learners' deeper cognitive functions, including inductive data-driven learning, learning by discovery, problem-solving abilities, the development of analytical skills, and independent language learning strategies.

Despite the significance of corpora in language research and language learning and teaching, "corpus literacy" (Mukherjee, 2006; O'Sullivan, 2007) remains lacking. Currently, research and practice incorporating corpora into the learning and teaching of second languages/foreign languages are still quite rare. McEnery et al. (1997) discussed how to implement corpora in English teaching at secondary schools in the UK. However, empirical studies on the application of corpora in teaching, both domestically and internationally, are extremely scarce. Therefore, there is an urgent and necessary need for empirical research on the involvement of corpora in foreign language teaching, especially in English instruction at the undergraduate and graduate levels in higher education.

3. Methodology

The primary objective of this study is to construct a corpus of business English derived from corporate websites and to explore its applications in business English teaching. The specific research methods involve several key steps:

3.1. 3.1 Data Collection

The data for this research was collaboratively gathered by faculty and students from the School of Foreign Languages at our university. The selected corpus consists of materials sourced from the official English websites of Fortune Global 500 companies. These companies hold significant influence in the international market, and their website content is both authentic and professional, providing an accurate reflection of the practical use of business English. Initially, web scraping techniques were employed to collect textual data from these websites, covering various aspects such as company profiles, product information, industry news, and job openings. This research also employed the online enTenTen corpus, which is introduced in section 3.2.

3.2. Corpus Processing

The collected textual data underwent a cleaning and preprocessing phase, which involved removing HTML tags, eliminating redundant information, and performing part-of-speech tagging. During this stage, natural language processing tools such as NLTK (Natural Language Toolkit) and SpaCy were utilized to ensure the corpus's high quality and usability. The texts were classified and filtered based on industry, function (such as product descriptions, company profiles, and customer service), and linguistic characteristics. Linguistic annotations, including word frequency statistics, phrase extraction, and collocation analysis, were applied to the selected texts to provide rich data resources for subsequent teaching.

This study employed corpus software Sketch Engine to analysize. The processed data will undergo tokenization, stop word removal, and word frequency calculations, followed by the creation of an index for future retrieval and analysis. The corpus will integrate various query functions to facilitate easy access for both teachers and students. Additionally, a user-friendly online platform will be developed to enable teachers and students to access and utilize corpus information, allowing for more flexible teaching methods.

The participants in this study were undergraduate students from three classes in the English Department and Business English Department at Guangzhou College of Business. The corpus used was the TenTen corpus family from Stanford University, specifically the enTenTen English corpus, which contains 3.3 billion tokens and 2.8 billion word forms. This corpus was constructed using Stanford NLP (Natural Language Processing) tools. The study used the fourth-generation corpus analysis software Sketch Engine, which is powerful and widely applied across various disciplines (Baker and Collins, 2023).

Word Sketch grammar (WSG) is a set of rules defining the grammatical relations (=columns/categories) in a Word Sketch. In other words, WSG tells Sketch Engine which words should be regarded as collocations of the search word and also what type of collocation they are. WSG defines the criteria using POS tags, distance between words, and other criteria. The criteria are written using CQL. WSG is language dependent, the same WSG cannot be shared across languages. Typically corpora in the same language use the same WSG, but exceptions exist. Users can write their own WSG to match their specific needs. Corpora in unsupported languages can make use of a universal WSG which provides only basic statistics of words surrounding the keywords ignoring the grammar of the language. Uses can write their own Sketch Engine grammars and include specific grammatical relations of their choice. The instructional content of this teaching innovation covered all aspects of the syllabus, but instead of traditional grammar teaching methods, it employed a corpus-assisted three-dimensional teaching approach. The specific steps are as follows:

a) Teaching the Basics of Corpora and Operational Methods: Introduce the course content and objectives, including the fundamental concepts of corpora and corpus linguistics. Teach students simple corpus search methods, gradually introducing more advanced features.

b) Group Students: Organize students into groups in preparation for subsequent collaborative learning and activities.

c) Design a Corpus-Assisted Course Plan Integrated with the Syllabus: Based on the syllabus, design teacher-led corpus searches and group/individual activities until students master the knowledge points. Assess students' understanding of each knowledge point, including writing, speaking, and testing, and provide timely feedback and guidance.

d) Main Data Collection: The primary data for this study includes the following: (i) student work, including corpus search assignments, grammar exercises, written reports on their corpus data analysis and findings, and reflections on vocabulary and grammar corpus studies; (ii) teachers' teaching logs, lesson plans, example teaching activities, reflection logs, records of teacher discussion sessions, and discussions between the author and teachers; and (iii) post-class surveys from both students and teachers.

3.3. Teaching Applications

a) To validate the practicality of the corpus, this study will assess its effectiveness through actual teaching experiments. The implementation details have been extensively discussed in the August 2024 issue of the journal "Language and Literature Teaching and Research," focusing on the following aspects:

b) Course Design: Based on the data from the corpus, we will design relevant English courses that cover common communicative scenarios and language usage patterns, including both written communication and spoken interaction.

c) Case Analysis: By analyzing examples from the corpus, students will be assisted in understanding and applying English language skills. Further research will collect learning data from both experimental and control groups for comparative analysis, focusing on key indicators such as vocabulary growth, improvements in listening and speaking skills, and comprehension of language use in real communicative contexts. By the end of a semester of teaching practice, it is expected that the

performance of students in the experimental group will significantly exceed that of the control group in these areas.

d) Assessment Criteria: Targeted assessment criteria based on the corpus will be constructed to effectively evaluate students' performance in business English writing and oral expression.

e) Interactive Learning: Students will be encouraged to utilize the corpus for group collaborative projects, enhancing their application skills and teamwork through the study of real case scenarios.

3.4. Other Applications

In addition to the primary focus on grammar teaching, the corpus can be utilized in various other applications to enhance business English education:

Firstly, assessment preparation: The corpus can serve as a resource for developing practice materials and mock assessments, allowing students to familiarize themselves with the types of language and terminology commonly used in business contexts. This preparation can be essential for students who are preparing for standardized tests or professional certifications in business English.

Secondly, curriculum development: Insights gained from analyzing the corpus can inform curriculum design across levels. Educators can tailor their syllabi to include relevant vocabulary, phrases, and structures identified within the corporate texts, ensuring that the instruction aligns with current business communication practices.

Thirdly, research and analysis: Teachers and students can engage in research projects that involve analyzing trends in business language usage based on the corpus data. This can help them develop critical thinking and analytical skills while deepening their understanding of industry-specific language.

Fourthly, professional development: Teachers can leverage the corpus in workshops aimed at improving their own teaching practices. By examining real-world examples from the corpus, teachers can explore innovative ways to integrate authentic materials into their classrooms, enhancing the relevance and applicability of their instruction.

Fifthly, digital Literacy skills: In navigating and utilizing the corpus, students will also develop essential digital literacy skills. They will learn how to effectively search, retrieve, and analyze online information, preparing them for the increasingly digital workplace.

Last but not least, cross-cultural communication: The corpus provides a rich resource for exploring cross-cultural differences in business communication. Students can analyze how language varies across regions and industries, fostering their understanding of global business practices and enhancing their intercultural competence.

The integration of the online corpus into business English teaching extends far beyond grammar instruction, offering a wealth of opportunities for improving language proficiency and preparing students for real-world business scenarios.

4. Discussion and Results

Over the course of a semester, students learned various grammar topics, including articles, nouns, auxiliary verbs, modal verbs, negation, subjunctive mood, voice, and clauses. After the grammar topics were explained, students searched the corpus for examples to deepen their understanding or independently discovered patterns within the corpus to comprehend grammatical rules. Students not only identified frequently occurring grammatical patterns but also developed their deductive and inductive reasoning skills. They were able to complete individual or group assignments using the corpus, and teachers discovered many issues students encountered during learning (such as topics that most interested students or areas where students were particularly solid), allowing for timely guidance and feedback. As the course progressed, students learned to formulate more complex queries and used these for writing essays. The vast number of examples generated by the concordance function could be applied in test questions. Surveys indicated that all 132 participating students had a positive attitude towards this approach. This innovative teaching model not only helped students consolidate their grammar knowledge but also enhanced their self-directed learning skills and proficiency in corpus technology.

Through a semester of practice, the author made the following observations:

a) Suitability of Certain Grammar Points: Some grammar points are more suitable for this teaching method than others. For example, student feedback on the subjunctive mood was less positive compared to other grammar topics. Through observations and post-class interviews, the author suggested possible reasons: (i) the use of the subjunctive mood is subjective and complex; (ii) searches in the corpus often did not yield the expected suitable examples; and (iii) there was a lack of timely communication and feedback between teachers and students.

b) Differentiated Improvement Across Grammar Aspects: In grammatical instruction using online corpora, students exhibited varying degrees of improvement in the form, meaning, and use of grammatical items. Notably, there was significant progress in understanding grammatical forms, while comprehension of meanings and practical application were relatively weak. This disparity may stem from multiple factors and warrants further analysis and exploration.

c) Challenges of Information Overload: Students often encountered situations of information overload or struggled to extract effective information accurately. How to effectively apply the rich findings of corpus linguistics to second language teaching is a question worth exploring. Through reflection, the author believes that students could be trained to utilize online corpora for self-directed learning and to enhance their language skills.

Acknowledgment

Research project: Guangzhou College of Commerce University-level Research project: Research on the Construction and Application of online Corporate Websites Corpus. (No. 2020XJYB030)

References

[1] Baker, P. and Collins, L. (2023). Creating and analysing a multimodal corpus of news texts with Google Cloud Vision's automatic image tagger. Applied Corpus Linguistics, 3(1), p.100043.

[2] Kennedy, G. (2010). An introduction to corpus linguistics transferred to digital print on demand. London: Longman.

[3] Larsen-Freeman, D., Wen, Z., and Mohebbi, H. (2024). The Past and the Future of Language Learning and Teaching: An Interview with Diane Larsen-Freeman. Language Teaching Research Quarterly, 39, pp.7–17.

[4] Li, Y., He, A., and Huang, L. (2022). A Multi-Factor Statistical Method Shift in Learner Corpus Research. Corpus Linguistics, 9(2), pp.1-13+166.

[5] Hu, K. and Gao, L. (2024). Development of Foreign Language Disciplines in the Context of Large Language Models: Issues and Prospects. Foreign Language World, (2), pp.7–12.