Sino-US English Teaching, March 2025, Vol. 22, No. 3, 113-117
doi:10.17265/1539-8072/2025.03.006

Corpus and Knowledge Graph-Assisted Integrated English Teaching

YU Weiwei
Zhejiang International Studies University, Hangzhou, China

With the continuous advancement of information technology, corpora and knowledge graphs (KGs) have become indispensable tools in modern language learning. This study explores how the integration of corpora and KGs in integrated English teaching can enhance students' abilities in vocabulary acquisition, grammar understanding, and discourse analysis. Through a comprehensive literature review, it elaborates on the theoretical foundations and practical values of these two technological tools in English instruction. The study designs a teaching model based on corpora and KGs and analyzes its specific applications in vocabulary, grammar, and discourse teaching within the *Integrated English* course. Additionally, the article discusses the challenges that may arise during implementation and proposes corresponding solutions. Finally, it envisions future research directions and application prospects.

Keywords: corpus, knowledge graph, integrated english teaching, teaching model, language proficiency, educational innovation

Introduction

With the rapid development of information technology, the field of education is undergoing profound transformations. In English teaching, traditional methods are increasingly unable to meet the diverse learning needs of students. As emerging technological tools, corpora and knowledge graphs (KGs) provide new approaches and methodologies for English instruction. Corpora offer extensive authentic language data, assisting students in better understanding real-world language usage. KGs clearly depict relationships between concepts, promoting systematic learning among students.

This study adopts multiple research methods, including literature review and case studies, focusing on the practical applications, advantages, and challenges of corpora and KGs in English teaching. It aims to explore how to organically integrate these two technologies to construct a new integrated English teaching model. By analyzing their respective characteristics and strengths, the study designs concrete instructional schemes and evaluates their effectiveness in English education. The significance of this research lies in providing new perspectives for English teaching reform, promoting the deep integration of information technology with language education, and ultimately enhancing the quality and efficiency of English instruction.

YU Weiwei, Ph.D., associate professor, Department of Language Schience, School of English Studies, Zhejiang International Studies University, Hangzhou, China.

Literature Review

Applications of Corpora in English Teaching

A corpus is a large-scale text collection compiled for linguistic research, providing empirical data for teaching. Corpus linguistics emerged in the 1980s, with Huang Renjie and Yang Huizhong pioneering the Computer Corpus of Science and Technology English. By the mid-1990s, corpora were integrated into foreign language teaching through data-driven learning (DDL) (Zhen, 2005).

Corpora are applied in the following aspects.

Language analysis. For instance, extracting high-frequency vocabulary, phrases, and grammar (Biber et al., 1999); analyzing register differences (Hyland, 2006); and comparing English variants (Carter & McCarthy, 2006).

Textbook development. For example, providing authentic language materials (Cui & Ye, 2012) and designing vocabulary/grammar lessons based on frequency (Nation, 2001).

Teaching practice. For instance, DDL fosters learner autonomy (Johns, 1991); corpus tools (e.g., Corpus of Contemporary American English [COCA]) enhance vocabulary and grammar learning (Wang, Davies, & Liu, 2008).

Learner research. For instance, analyzing errors and first language (L1) transfer using learner corpora (Granger, 2002; Altenberg & Granger, 2001).

Applications of KGs in English Teaching

KGs represent knowledge visually, aiding in vocabulary, reading, writing, and personalized learning. It has been applied in the following aspects.

Vocabulary teaching. For example, polysemy instruction (Wang, Davies, & Liu, 2008) and instructional effectiveness and collocations (Liu, Zhang, & Chen, 2018).

Reading comprehension. For instance, clarifying concepts (Chen, Jia, & Xiang, 2019) and supplementing background knowledge (Li, Sun, & Wang, 2020).

Writing assistance. For instance, in content analysis and revision suggestions (Huang, Li, & Wang, 2021) and optimizing structure (Wang, Li, & Huang, 2021).

Personalized learning. For instance, in recommending tailored learning paths (Sun, Zhang, & Li, 2020) and integrating resources, AI-enhanced KGs improve teaching efficiency (Hou, Li, & Wang, 2024).

Instructional Design

This study takes *Integrated English (II)* as an example, using the textbook *Integrated Course (3rd Edition, Volume 2)*, edited by He Zhaoxiong and published by Shanghai Foreign Language Education Press in 2022. The aim is to investigate the characteristics and advantages of a teaching model integrating corpora and KGs.

This course, a foundational and core subject for first-year English majors, aims to cultivate students' fundamental language skills and enhance their comprehensive ability to use English, laying a solid foundation for future specialized studies. Upon completing the course, students should master basic English language knowledge (phonetics, grammar, vocabulary, and discourse structures); understand, analyze, and evaluate fundamental linguistic features and stylistic variations across genres; gain insights into the historical traditions and cultural characteristics of English-speaking countries; recognize intercultural differences; and proficiently apply listening, speaking, reading, writing, and translation skills. They should also gradually develop critical thinking and problem-solving abilities using language knowledge.

Additionally, by incorporating bilingual readings from *Xi Jinping: The Governance of China* and *Understanding Contemporary China: A Reading Course*, students are expected to articulate Chinese stories in English and cultivate educational sentiments.

This study designs a corpus and knowledge graph-assisted integrated English teaching model. The model is student-centered, emphasizes autonomous and inquiry-based learning, and fully utilizes the technical advantages of corpora and KGs to enhance instructional effectiveness.

In designing the teaching model, we first established the course objectives. There are three primary objectives.

Objective 1: Understand and master basic English knowledge (phonetics, grammar, vocabulary, and discourse structure) with the help of corpora and KGs; analyze and evaluate linguistic features and stylistic expressions across genres.

Objective 2: Deepen understanding of historical traditions and cultural traits of English-speaking countries, recognize intercultural differences via corpus and knowledge graph exploration, and proficiently apply basic language skills (listening, speaking, reading, writing, and translation).

Objective 3: Develop critical thinking, problem-solving awareness, and lifelong learning skills based on corpus and knowledge graph usage, and be capable of interpreting *The Governance of China* and *Understanding Contemporary China* as well as articulating Chinese stories in English.

Subsequently, specific teaching content and activities were designed according to the objectives. Taking Unit 3: Age and Generation Gap (Reading I) as an example, we illustrate how to combine corpus analysis of vocabulary frequency and collocational patterns with knowledge graph construction of semantic relationships.

Evaluation is an essential component of this instructional model. A diversified evaluation system is adopted, integrating formative and summative assessments, emphasizing process evaluation and student self-assessment. By using corpora and KGs, teachers can more objectively assess student progress and promptly adjust instructional strategies.

Specific Applications

In vocabulary teaching, the combined use of corpora and KGs can significantly enhance instructional effectiveness. Teachers can utilize online corpora, such as British National Corpus (BNC), COCA, CQPWeb, and Skell to analyze the target vocabulary's frequency, collocational patterns, and register characteristics, thereby designing targeted vocabulary exercises.

For example, using the BNC to compare the differences in adjective collocations between "agony" and "pain", it was found that "agony" often collocates with words like "final", "absolute", and "foremost", while "pain" typically pairs with adjectives, such as "abdominal", "severe", "real", and "great".

Using COCA to examine the stylistic distribution and diachronic trends of the word "plead", it was discovered that "plead" appears predominantly in informal registers, such as spoken language, television/media, news, and novels. Its synonyms include "beg", "declare", "defend", "petition", and "assert". The frequency of "plead" initially declined and then rose again since the 1990s.

Using CQPWeb to search for the prefix "mega*", it was found that "mega-" as a prefix combines with adjectives, nouns, and others, indicating largeness or intensity.

Using Skell to analyze collocations for "feverishly", both in terms of adjectives, verbs it modifies, and adverbs that modify it, it was found that "feverishly" tends to collocate with action verbs and adjectives related

to emotions and moods.

At the same time, KGs are utilized to visually map the semantic networks of vocabulary, helping students establish connections between vocabulary and text themes, deepening their understanding and memory.

For instance, when teaching the word “plead,” the teacher can present its semantic associations with words like “guilty”, “sentence”, and “prosecutor”, helping students build a structured knowledge network of relevant vocabulary and discourse.

The beginning of the knowledge graph covers themes, such as “worship youth”, “commercial ads”, and “daily activities”; transitional elements include “contradict premise” and “un-American to say”; the conclusion involves keywords like “automatically” and “parents’ example”. Rhetorical devices used in the text are also analyzed, such as: Repetition (e.g., “young means”), Metaphor (e.g., “sign with evil”, “slaves to style”), Parallelism (e.g., repeated use of “no longer”), and Rhetorical question **(**e.g., “aging experience ... surprise”).

By combining corpus analysis with knowledge graph visualization, students not only gain a deeper linguistic understanding, but also improve their discourse interpretation skills and critical thinking abilities.

Challenges and Solutions in Corpus and Knowledge Graph-Assisted English Teaching

Although corpora and knowledge graphs have demonstrated tremendous potential in English teaching, their practical application still faces several challenges.

First, technical barriers pose a significant issue. To address this, institutions should offer appropriate training and support to enhance the technical skills of both instructors and learners.

Second, the development and quality control of teaching resources also present challenges. Building high-quality corpora and knowledge graphs demands considerable time and effort. It is recommended to establish collaborative teacher networks to share teaching resources and improve resource utilization efficiency. Furthermore, standards for resource development should be formulated to ensure quality and applicability.

Third, balancing technological integration with traditional teaching methods is another major challenge. Overreliance on technology might neglect the humanistic aspects of language teaching and the importance of interpersonal interaction. Therefore, teachers should design instructional activities thoughtfully, aligning technological tools with students’ characteristics and learning objectives to achieve an organic integration of technology and traditional methods.

Lastly, the establishment of appropriate assessment systems is a problem that must be addressed. Traditional evaluation methods may not fully capture students’ learning outcomes under this innovative model. It is suggested to develop diversified evaluation tools, such as electronic portfolios and online assessments, to comprehensively assess students’ progress and competence development. Corpora and knowledge graphs present extensive prospects in English teaching.

Conclusions

This study explores the application of corpora and KGs in integrated English teaching, designs a new instructional model, and analyzes its specific uses in vocabulary, grammar, reading, and writing instruction. The research findings demonstrate that this technology-assisted teaching model can effectively enhance students’ language proficiency and learning engagement. However, challenges still exist in areas, such as technical thresholds, resource development, methodological integration, and assessment systems.

Future research could further explore how to optimize the integration of corpora and KGs in English teaching.

Possible directions include developing more intelligent auxiliary teaching systems, investigating how students with different learning styles adapt to this model, and extending the application of this model to other language teaching contexts to explore its broader educational potential.

In general, corpora and KGs bring both new opportunities and challenges to English instruction. By making effective use of these technological tools, we can create more efficient, engaging, and personalized learning environments, thereby promoting the advancement of English teaching to a higher level.

References

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of *make* in native and non-native student writing. *Applied Linguistics, 22*(2), 173-194.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman: Pearson Education.

Boulton, A. (2009). Data-driven learning: On paper, in practice. *Corpora and Language Teaching, 3*(3), 17-52.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.

Chen, X., Jia, S., & Xiang, Y. (2019). A review: Knowledge reasoning over knowledge graphs. *Information Fusion, 46*, 1-12.

Cui, Y., & Ye, G. (2012). Application of corpora in college English teaching. *China Higher Education, 1*, 62-64.

Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). The Netherlands: John Benjamins Publishing.

Hou, W., Li, N., & Wang, L. (2024). Integrating AI-based immersive teaching resources into college English: A knowledge graph perspective. *Modern English, 13*, 50-52.

Huang, X., Li, J., & Wang, H. (2021). A knowledge graph-based approach for improving English writing skills. *Journal of Educational Technology & Society, 24*(2), 123-135.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London: Routledge.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal, 4*, 1-16.

Li, J., Sun, Y., & Wang, H. (2020). Enhancing reading comprehension with knowledge graphs: A case study in English education. *Computers & Education, 150*, 103841.

Liu, Y., Zhang, W., & Chen, Y. (2018). Knowledge graph construction and application in language learning. *Journal of Language Teaching and Research, 9*(6), 1234-1242.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Sun, Y., Zhang, W., & Li, J. (2020). Personalized learning path recommendation based on knowledge graph. *Educational*

Wang, H., Li, J., & Huang, X. (2021). A knowledge graph-based approach for optimizing English writing structure. *Journal of Computer Assisted Learning, 37*(3), 789-801.

Wang, X., Davies, M., & Liu, G. (2008). The Corpus of Contemporary American English (COCA)—A valuable platform for English teaching and research. *Foreign Language World, 5*, 27-33.

Zhen, F. (2005). Data-driven foreign language learning: Concepts, methods, and technologies. *Foreign Language World, 4*, 19-21.