

A Review and Prospect of Corpus-Based English Teaching Research in Chinese Middle Schools

Lingyu Yang

China West Normal University, Nanchong 637001, China
Email: 1286775647@qq.com

How to cite this paper: Yang, L. Y. (2025). A Review and Prospect of Corpus-Based English Teaching Research in Chinese Middle Schools. *Education and Social Work*, 2(2), 61-68. ISSN Print: 3079-515X; ISSN Online: 3079-5168.
<https://doi.org/10.63313/ESW.9077>
Published: 2025-07-10

Copyright © 2025 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Abstract

This paper conducts a literature study on papers related to corpus-based English teaching published in *Foreign Language Teaching in Schools* (Middle School Edition) and *English Teaching & Research Notes* over the 20-year period from 2003 to 2023. The results show that: (1) In terms of overall trends, such research began in 2005, grew rapidly and peaked in 2010, then dropped sharply. Over the next decade, it experienced multiple sudden rises and falls, with zero papers published in 2020. However, growth resumed in 2022, and continued growth is expected in 2023. (2) Regarding research objects, studies on senior high school education are significantly more numerous than those on junior high school, and are comparable to studies covering comprehensive school stages. (3) In terms of research methods, empirical quantitative research dominates. (4) In research content, studies focus heavily on the role of corpora in comprehensive English teaching and vocabulary teaching, with fewer studies on speaking, discourse, and translation. These findings provide insights for corpus-based teaching research in middle schools.

Keywords

Corpus; English Teaching and Research; 20-Year Review

1. Introduction

A corpus is a database of large amounts of language information used in real contexts, designed for research and application. It is a "repository" that stores such language materials on a large scale according to specific purposes and methods (Hou Jing, 2019). Corpora are applied in dictionary compilation, language teaching, and traditional linguistic research, serving as the foundational resource for corpus linguistics. Corpus linguistics emerged in the 1960s and flourished in the 1980s. Despite initial marginalization by the rationalist-based Chomskyan school, it gradually gained recognition with the construction of large-scale Chinese and English corpora.

Corpus linguistics research in China emerged in the late 20th century (He Anping,

1997), going through three stages: initial exploration, scale expansion, and mature application. The initial stage focused on language data collection, sorting, and annotation; the development stage emphasized expanding corpus scale and diversifying language materials; the mature stage featured in-depth data processing and expanded applications, such as language analysis and dictionary compilation. Several English teaching corpora targeting Chinese middle school learners have since emerged, bringing new perspectives to middle school English teaching in China. This study aims to review and analyze the effectiveness and limitations of corpus applications in middle school English teaching in the era of big data, using quantitative and qualitative methods to synthesize relevant literature.

English Teaching & Research Notes and Foreign Language Teaching in Schools (Middle School Edition) are core journals in China's basic education foreign language field. Their content closely aligns with the practical needs of middle school foreign language teaching, covering both theoretical discussions and practical applications, thus demonstrating high utility. With primary readers being middle school English teachers and researchers, these journals effectively reflect the latest progress and status of corpus-based English teaching research in Chinese middle schools. This paper surveys articles on corpus-based English teaching published in these two journals from 2003 to 2023, analyzing the status and trends over the two decades to promote further development in this field.

2. Research Methods

The author searched for "corpus" in article titles in English Teaching & Research Notes and Foreign Language Teaching in Schools (Middle School Edition) from 2003 to 2023, and selected 48 articles related to corpus-based middle school English teaching research based on titles, keywords, and abstracts. Descriptive statistical methods were employed.

These 48 articles were analyzed through reading, categorization, and synthesis to review corpus-based middle school English teaching research in China over the past 20 years (2003-2023). Comparisons of the number of papers and publication years yielded overall distribution and phased statistical results (see Figures 1 and 2).

3. Status of Corpus-Based English Teaching Research

3.1. Research Trends

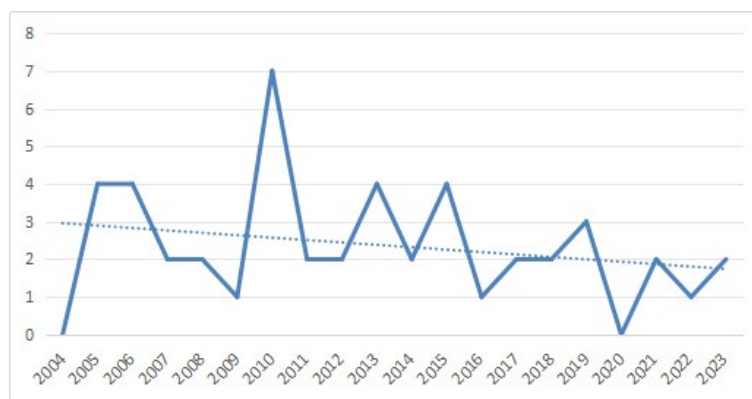


Figure 1: Annual Distribution of Core Papers on Corpus-Based Middle School English Teaching Research (2004-2023)

Figure 1 shows that corpus-based middle school English teaching papers first appeared in 2005 (Huang Bing, 2005; Wang Longyin & He Anping, 2005; Li Jianping, 2005; Liu Xingbing, 2005). Subsequent growth of 2 or more papers year-on-year occurred in 2010 (7 papers), 2013 (4), 2015 (4), and 2021 (4), with 2010 marking a peak of 7 papers. Over the 20 years, the overall trend evolved from nonexistence to rapid growth in 2010, followed by a sharp decline in 2011. A subsequent rise from 2016 to 2019 was again followed by a drop, with growth resuming in 2022 and expected to continue in 2023. Phased statistics by five-year intervals (see Figure 2) show 13, 16, 11, and 8 papers in the four stages, respectively, indicating the second stage as the boom period, with repeated sudden fluctuations reflecting instability in this field of research.

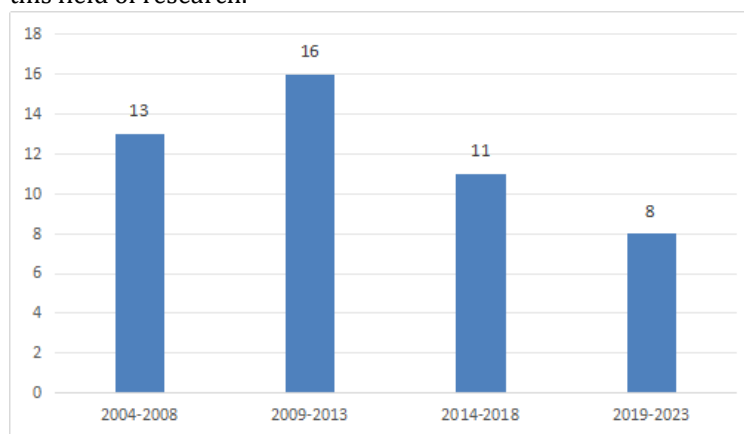


Figure 2: Phased Distribution of Core Papers on Corpus-Based Middle School English Teaching Research (2004-2023)

3.2. Research Stages

This study categorizes research stages into junior high, senior high, and comprehensive (covering the entire middle school period). Results are shown in Figure 3.

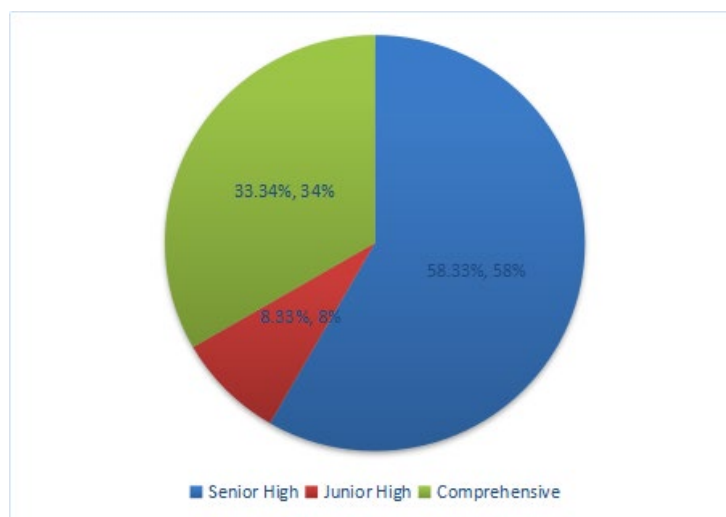


Figure 3: Distribution of Corpus-Based Middle School English Teaching Research by School Stage (2004-2023)

Statistically, senior high school research significantly outnumbers that on junior high school and is comparable to comprehensive-stage studies. Growing attention to senior high school in recent years suggests a need for more focus on junior high school research.

3.3. Research Methods

Research methods are divided into empirical and non-empirical (Wen Qiufang, 2001). Empirical research, based on facts and data, uses rigorous methods and statistical analysis to test hypotheses, aiming to discover objective laws with replicability and verifiability. It is data-driven, quantitative, and methodologically rigorous. Non-empirical research emphasizes theoretical discussion, subjective interpretation, and speculative analysis (e.g., through case studies) to explore phenomena, focusing on subjectivity, interpretability, and understanding to reveal intrinsic logic.

Of the 48 papers, empirical research dominates with 34 papers (70.83%), while non-empirical research accounts for 14 papers (29.17%). This dominance reflects both the empirical nature of corpus linguistics and a tendency toward quantification over diversity in middle school applications. Figure 4 shows empirical research flourished from 2011 to 2015, with subsequent fluctuations but consistent numerical superiority. Non-empirical research remained stable but limited in quantity.

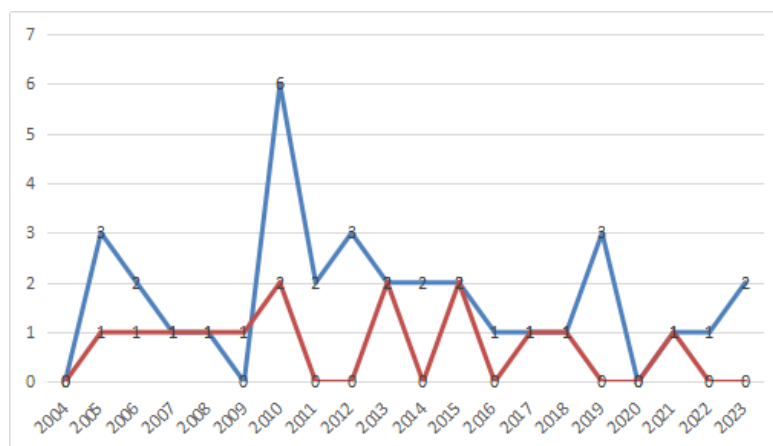


Figure 4: Research Methods in Corpus-Based English Teaching Research (2004-2023)

3.4. Research Content

Research content is inherently linked to methodology (Wen Qiufang, 2001). Drawing on Biber (2000), Kennedy (2000), and Aijmer (2009), as adjusted for China's context, this study classifies corpus linguistics research into 11 categories, including vocabulary, grammar, and translation.

Statistical analysis reveals significant disparities in research content over 20 years. Vocabulary teaching research dominates with 16 papers, followed by corpus applications in teaching, writing, and grammar. Discourse, reading, and college entrance exam proposition research are scarce. Detailed trends are as follows:

1. **Vocabulary:** Vocabulary teaching is the most studied area, primarily focusing on senior high school. Research includes textbook vocabulary analysis, thematic/high-frequency vocabulary teaching, and teachers' awareness of lexical chunks. Textbook vocabulary studies examine the breadth and depth of middle school English textbook vocabulary and identify common issues in appendices (Xie Jiacheng, 2008). Studies on senior high students' vocabulary use include Cao Rongsheng & Zhang Yufang (2012) on lexical phrases in writing and Wu Qingling (2012) on teaching/reviewing high-frequency vocabulary. Vocabulary's central role in reading, writing, and speaking explains its prominence.
2. **Corpus Construction:** Researchers aim to address teaching challenges through personalized corpora. Zhang Aixiong (2006) discusses how small personalized foreign language teaching corpora aid research and teaching via multifunctional deep processing. Jiang Fabing (2010) highlights the value of small senior high English corpora for teaching and research. Most studies focus on teacher support, e.g., Xia Jianqing (2010) built a corpus to provide teachers with multi-aspect resources.
3. **Teaching Applications:** Five papers explore corpus potential in teaching. Xie

Jiacheng (2006) examines online corpora's role in classroom teaching and autonomous learning. Gui Shichun et al. (2010) predict corpus-assisted foreign language teaching as an inevitable trend. Wu Jun (2010) introduces corpus-based reference books, and Yang Suxiang (2015) demonstrates CQPweb's application in phonetics, vocabulary, and grammar teaching.

4. Discourse Analysis: Research focuses on classroom discourse and teacher-student interaction. Wang Longyin & He Anping (2005) investigate middle school English classroom discourse to promote corpus-supported discourse analysis. Wang Yingna & Liu Yongbing (2011) compare questioning/responses in key and non-key senior high schools to enhance effective output.
5. Writing: Growing emphasis on writing has spurred corpus-based research. Bi Shaoqin (2007) uses schemata theory and corpora to support pre-writing, writing, and revision. Chen Liyun & Li Chunyan (2018) explore corpus-based writing feedback models for advice letters. Ying Jiqiao et al. (2019) show self-built narrative corpora improve logical coherence and linguistic accuracy in continuation writing.
6. College Entrance Exams: Studies use self-built corpora to explore exam applications. Deng Xiaoyan & Zhang Hongqiang (2010) apply personalized corpora to teaching and proposition. Pan Mingwei & Xu Xinxing (2010) use learner and native speaker corpora to assess the content validity of Shanghai's college entrance exam grammar items, noting the need for deeper exploration.

Research on discourse, speaking, and translation is scarce. Speaking's marginalization in exams reduces focus on oral corpora. Discourse research is limited by teaching priorities on vocabulary/grammar and its inherent complexity (context, genre, coherence). Translation receives less attention due to minimal emphasis in middle school curricula. These gaps warrant more research.

4. Problems in Domestic Corpus Teaching Research

While achievements exist, several issues persist:

4.1. Inadequate Corpus Construction and Application

China's corpora (e.g., Chinese English Learner Corpus, English-Chinese Parallel Corpus) in scale are smaller than foreign counterparts (e.g., British National Corpus, Corpus of Contemporary American English). Specialized middle school English teaching corpora are limited, with content often exceeding students' proficiency or misaligning with syllabi, restricting teacher/researcher choices.

Many teachers lack understanding of corpus functions and application skills, hindering effective use. Technical barriers (data extraction, annotation) further limit utilization.

4.2. Insufficient Integration with Other Methods

Teachers rarely combine corpora with traditional methods (grammar-translation, communicative approach). Interdisciplinary collaboration with education, psychology, and sociology is lacking, restricting innovation.

4.3 Disconnect from Teaching Practice

Most corpora serve university research rather than middle school classrooms. Inadequate teaching-oriented retrieval tools and syllabus misalignment increase teachers' workload, despite nearly ten large learner corpora existing in China.

5. Conclusion

This study analyzes 48 core journal papers (2004-2023) to assess progress and gaps in corpus-based middle school English teaching research. Key recommendations for future research include:

1. Enhance corpus quality/diversity with real, comprehensive, and up-to-date materials. Avoid duplication, overcome regional barriers, and develop national integrated corpora.
2. Bridge research-practice gaps by aligning corpora with teaching needs, collaborating with frontline teachers, and promoting best practices via workshops.
3. Explore integration with task-based/situational teaching to leverage synergies.
4. Draw on cognitive science, neuroscience, IT, and AI to develop "corpus+" models, enhancing teaching efficiency and acquisition mechanisms.

Limitations include excluding other journals, monographs, and theses, affecting generalizability. However, this study summarizes trends and provides insights. Advances in technology suggest growing adoption of corpus-based methods in Chinese English teaching.

References

- [1] Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press, Cambridge.
- [2] Granger, S., Huang, J., & Petch-Tyson, S. (Eds.). (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. John Benjamins Publishing Company, Amsterdam.
- [3] Johansson, S., & Stenström, A.-B. (Eds.). (1991). *English Computer Corpora: Selected Papers and Research Guide*. Mouton, Berlin.
- [4] Ying Jiqiao, Su Yindan, & Zhang Chunqing. (2022). Teaching Practice of Using Self-built Small Narrative Corpora to Assist Continuation Writing. *English Teaching & Research Notes*, (8), 32-37.
- [5] Chen Liyun, & Li Chunyan. (2018). Exploring Corpus-based Teaching Models for Senior High School English Writing Feedback Classes: A Case Study of Advice Letters. *English Teaching & Research Notes*, (11), 64-70.
- [6] Chen Liyun. (2010). Schema Theory, Corpora, and Senior High School English Writing Teaching. *Foreign Language Teaching in Schools (Middle School Edition)*, 33(8), 41-45.
- [7] Jiang Qiaoling, & Zhang Chunqing. (2021). Exploring Corpus-assisted Thematic Vocabulary Teaching Methods in Senior High Schools. *Foreign Language Teaching in Schools (Middle School Edition)*, 44(10), 41-47.

- [8] Xie Jiacheng, & He Anping. (2008). A Corpus-based Study and Rationale Exploration of English Vocabulary Lists in Middle School Textbooks. *Foreign Language Teaching in Schools (Middle School Edition)*, 31(9), 1-5.
- [9] Wu Jun. (2010). Corpus-based Reference Books and Middle School English Teaching. *Foreign Language Teaching in Schools (Middle School Edition)*, 33(2), 19-22.
- [10] Yang Suxiang. (2015). CQPweb Online Corpus Retrieval Platform and Its Application in Foreign Language Teaching. *Foreign Language Teaching in Schools (Middle School Edition)*, 38(7), 1-8.
- [11] Wang Yingna, & Liu Yongbing. (2011). A Corpus-based Comparative Study of Classroom Questioning in Senior High Schools. *Foreign Language Teaching in Schools (Middle School Edition)*, 34(5), 38-44.
- [12] Jiang Jingyang, & Liu Shunjia. (2015). A Quantitative Analysis of Senior High School Students' Vocabulary Application Ability in Writing. *Foreign Language Teaching in Schools (Middle School Edition)*, 38(3), 29-35.
- [13] Xie Jiacheng. (2008). A Corpus-based Investigation of Vocabulary in Middle School English Textbooks. *Foreign Language Teaching in Schools (Middle School Edition)*, 31(1), 4-8.
- [14] Bi Shaoqin. (2023). A Comparative Study of Writing Sections in the New and Old Versions of Senior High School English Textbooks (People's Education Press). *English Teaching & Research Notes*, (4), 77-81.
- [15] Chen Zexuan, & He Guangkeng. (2007). A Comparison of Classroom Questioning by Senior High School English Teachers Before and After the Curriculum Reform: A Corpus-based Survey. *Foreign Language Teaching in Schools (Middle School Edition)*, 30(7), 7-12.
- [16] Zhou Yun. (2013). An Introduction to the Corpus of Contemporary American English. *Foreign Language Teaching in Schools (Middle School Edition)*, 36(10), 1-5.
- [17] Wen Qiufang. (2001). *Research Methods and Thesis Writing in Applied Linguistics*. Foreign Language Teaching and Research Press, Beijing.
- [18] Xu Manfei. (2004). Reflections on Deepening Research Based on English Learners' Spoken Corpora. *Journal of South China Normal University (Social Science Edition)*, (4), 143-145.
- [19] Song Hongbo, & Wang Xueli. (2013). A Review of Domestic Corpus Linguistics Research in the Past Decade. *Shandong Foreign Language Teaching*, 34(3), 41-47.
- [20] Wang Lifei, & Zhou Dandan. (2004). A 12-Year Review and Current Status of Spoken English Research in China. *Foreign Language World*, (6), 7-14.
- [21] He Anping. (1997). On Corpus Research. *Journal of Foreign Languages (Shanghai International Studies University)*, (5), 22-27.
- [22] Chen Lu, & Wei Han. (2005). The Role of Spoken English Corpora in Oral English Teaching. *Computer-Assisted Foreign Language Education*, (3), 23-26.