

A Review of 23-Year Corpus-based Research on Writing English Teaching in China from 2001-2023

Jiixin Lin, Yi Zuo

Sichuan University of Science & Engineering 643000, China

Abstract: This study is based on 145 core journal papers collected by China National Knowledge Infrastructure (CNKI) from 2001 to 2023, and explores the progress and shortcomings of English writing teaching research based on English writing corpora in China over the past 20 years. The results show that: (1) In terms of overall trend, this type of research has experienced rapid growth since 2001 and reached its peak from 2009 to 2016, before gradually decreasing. (2) In terms of research subjects, college students and graduate students make up the vast majority, with less involvement of middle school students. (3) In terms of research content, it mainly involves the construction, application, investigation, comparison, and review of corpora, as well as semantic rhyme, vocabulary, and chunks in English writing based on corpora. However, in providing teachers with more comprehensive training on corpus application, and combining artificial intelligence technology with corpus application to develop intelligent auxiliary writing tools, current research is very limited. The above findings have implications for the design of writing corpora and related teaching research.

Keywords: Writing English Corpus; Writing English Teaching; Review of Past Twenty Three Years.

1. Introduction

English writing is an indispensable part of the learning process for learners. Good English writing skills not only help learners better express their thoughts and viewpoints, but also contribute to their academic and career development. However, many learners face various challenges in English writing. They may lack appropriate vocabulary and sentence structures to accurately express their meaning. In addition, writing style, structure, and organization are also challenges that learners need to overcome. Therefore, it is crucial to find an effective teaching method to help learners improve their English writing ability. Corpus in English writing teaching can provide students with rich language materials, help them improve their writing ability, and promote their autonomous learning and language awareness. Corpus based English writing teaching method has been one of the teaching methods that has received much attention in recent years. A corpus is a database that contains a large number of real language samples, providing rich language resources and examples for learners to refer to.

Since the 1960s, corpus linguistics has made great progress. In language teaching research, Tim Johns (1991) proposed the case of data-driven learning, which examines the co-occurrence of semantic and grammatical structures of target vocabulary by using the index rows of corpus retrieval results, and attempts to improve the teaching effectiveness of easily confused and polysemous words. Many countries have established various corpora, categorized into different types, languages, etc., which play a crucial role in language teaching.

In China, corpus linguistics developed at the end of the 20th century, and many scholars began to study the application of corpora in the field of foreign language teaching (He Anping and Xu Manfei, 2003; Wang Lifei and Wen Qiufang, 2007; Liang Yan et al., 2004). Therefore, in recent years, there have been some English writing corpora specifically designed for Chinese English learners to meet their needs and characteristics. Based on this, this study analyzed the development status of corpus based English

writing teaching research in China over the past 20 years from several aspects, including research trends, research objects, and research content, by reviewing 145 core journal papers collected by China National Knowledge Infrastructure from 2001 to 2023. It also found some shortcomings. I hope these research results can provide reference and inspiration for current and future related research.

2. Research Status

The author utilized the advanced search function on China National Knowledge Infrastructure (CNKI) to search for topics, article titles, and more in the keyword and abstract search boxes, enter "English writing" and "corpus" respectively, and select SCI source journals, core journals, and CSSCI database to search for papers from 2001 to 2023. After careful screening and reading, a total of 145 relevant core journals were selected. By comparing the publication years and quantities of papers, the overall distribution and stage statistical results were obtained, as shown in Figure 1 and Figure 2:



Figure 1. Schematic diagram of article publication time

According to Figure 1, the core papers on corpus based English writing teaching in China included in CNKI began in 2001 (Li Jingquan and Cai Jinting, 2001), and subsequently achieved growth of more than four papers in 2011 (14 papers), 2013 (15 papers), and 2017 (8 papers) compared to the previous year. In 2013, there was an explosive growth, with 15 papers published, becoming the highest in 23 years. Although the number of publications has shown varying degrees of decline in other years, overall, the development

trend over the years has been from scratch, with a significant increase from 2011 to 2013, a gradual decrease after 2017, and another wave of growth in 2020, followed by a gradual decrease. Roughly using a 7-year interval for phased statistics (see Figure 2), the results of the three stages are 36, 77, and 32, respectively.

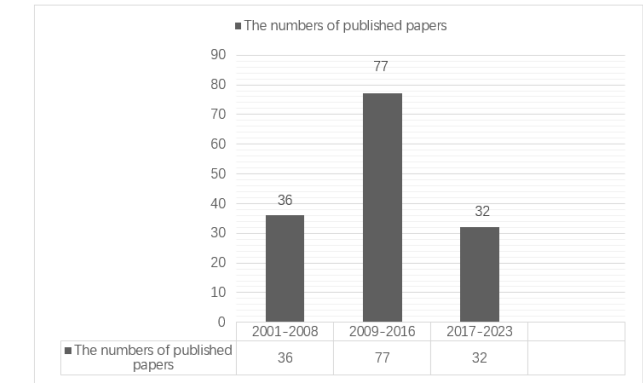


Figure 2. Periodic publication of core research papers on corpus based English writing teaching from 2001 to 2023

As shown in Figure 2, the initial stage progressed slowly, followed by explosive growth in the second stage. However, it gradually slowed down in the third stage, reaching its peak in the second stage and showing a phased downward trend. Therefore, the comprehensive trend of corpus based writing teaching research in China can be summarized as a slow and gradual increase, followed by a phased surge, and finally a decline.

3. Research Objects

The author obtained the distribution of English speaking research objects based on writing corpora in China over the past 20 years by comparing and statistically analyzing the research objects of the target paper, as shown in Figure 3.

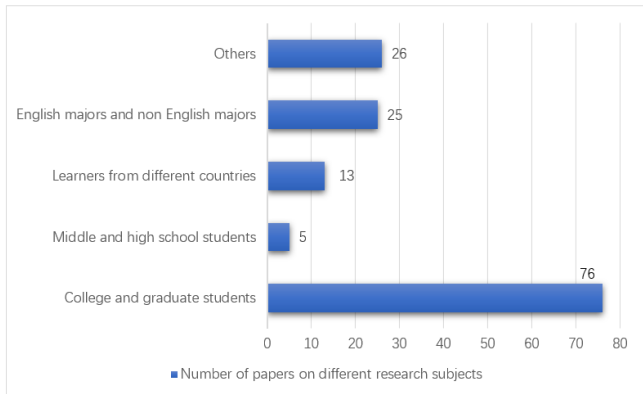


Figure 3. Distribution of research objects on corpus based writing teaching in China from 2001 to 2023

As shown in Figure 3, the research subjects involved 25 articles from English majors and non English majors, accounting for 17.2%; There are 13 papers involving learners from different countries, accounting for 9%; 5 articles involving middle and high school students, accounting for 3.4%; 76 papers involving college and graduate students, accounting for 52.4%. There are 26 papers without specific research subjects, accounting for 18%. From the above analysis, it can be concluded that in the past 23 years, the coverage of corpus-based writing teaching research in China has been quite extensive, mainly including Chinese English

learners and English learners from other countries, although the latter only accounts for 9% of the total proportion. This indicates that corpus-based writing teaching in China mainly focuses on the domestic market and provides services for English writing teaching in China.

4. Research Content

In this study, the author carefully checked and compared the research content of 145 core papers, and obtained the overall classification and statistical results of the research content, as shown in Table 1.

Table 1. Classification of Corpus based English Writing Teaching Research Content in China from 2001 to 2023

Number	research contents	Number of articles	Percentage (%)
1	semantic prosody	17	11.7%
3	vocabulary	36	25%
4	lexical chunks	8	5.5%
5	corpus construction	8	5.5%
6	Literature review	8	5.5%
7	teaching	38	26.2%
8	applications	21	14.5%
9	Corpus survey and comparison	7	4.8%
	Academic Writing		
10	other	2	1.3%

As shown in Table 1, based on the number of papers and their proportion in a total of 145 papers, the research content is as follows: there are 38 papers on teaching applications, accounting for 26.2%; There are 36 articles about vocabulary, accounting for 25%; 21 articles accounted for 14.5% of the corpus survey and comparison; There are 7 academic writing articles, accounting for 4.8%; There are 17 articles about semantic prosody, accounting for 11.7%; There are 8 articles about lexical chunks, accounting for 5.5%; There are 8 articles about corpus construction, accounting for 5.5%; There are 8 literature reviews, accounting for 5.5%; There are 2 articles about others, accounting for 1.3%.

In summary, these studies during this period have extensively covered various fields of English writing teaching, from the use of vocabulary and grammar to the standardization of academic writing, from semantic expression to corpus construction and analysis, demonstrating the fruitful achievements and research trends in corpus-based English writing teaching in China.

There are 21 studies focusing on corpus surveys and comparisons, accounting for 14.5% of the total number of papers. For example, Cai Shaolian (2006) compared traditional writing teaching and corpus based teaching methods in two classes, using questionnaire surveys and interview style surveys, proving that corpus based teaching can cultivate learners' writing interest, improve their writing level, and solve the problem of time-consuming and inefficient writing teaching; Li Ke and Zhang Yuan (2021) investigated the rhetorical ability of students in English writing and verified the constituent elements of rhetorical ability in English writing by constructing a small-scale English writing corpus; Several scholars have pointed out that English writing teaching based on corpus and other network technologies has a positive promoting effect on improving students' motivation to learn English writing (Yu Li and Guo Pingjian, 2008; Wang Yi and Xuan An, 2010; Wang Yi, 2011; Yang Xiaoqiong and Dai Yuncai, 2015). In addition, many

scholars have compared different corpora, such as Zhao Weibin's (2003) comparative study on the usage characteristics of logical connectors in native speaker corpora and Chinese English learner corpora; Lu Haiyan (2007) conducted a comparative analysis of the use of personal pronouns in Chinese university English learner corpora and native language learner corpora; Wang Lifei and Chen Gong (2008) pointed out that there are significant differences in the use of split sentences between Chinese students and native speakers; Lou Baocui (2013) compared and examined the distribution, structure, and stance expression characteristics of high-frequency reporting verbs in academic papers of undergraduate students from Chinese and British universities; Li Jianping (2013) compared the vocabulary length of composition between Chinese and American high school students by establishing English writing corpora in both countries; Li Xiaohong (2021) explores the usage characteristics of phrase frames in the introduction of doctoral dissertations in Chinese English linguistics from a functional perspective by comparing the corpus of Chinese and English doctoral dissertations; Lou Baocui and Wang Mimi (2021) compared corpora and pointed out that there are differences in the structure of lexical chunks used by Chinese and foreign scholars when expressing the same discourse function.

Other categories such as semantic prosody, lexical chunks, and academic writing account for 11.7%, 5.5%, and 4.8% of the total research, respectively. The study of semantic rhyme involves word item collocation (Zhang Zhijie, 2011; Zhang Huiping and Liu Yongbing, 2013; Liu Guobing and Zhang Xiaolian, 2021). There are 8 research papers on language chunks, such as Jin Zhiru (2011) who pointed out that strengthening the teaching of language chunks in foreign language teaching can greatly improve students' English writing ability; Lou Jie (2011) proposed that the online corpus assisted writing model can help improve the chunk application ability of middle and advanced English proficiency students in English writing; Xu Guoqin (2017) pointed out that integrating prefabricated language chunks with foreign language writing teaching can improve vocabulary richness in second language writing. In addition, many scholars have conducted research on the use of lexical chunks by Chinese and foreign learners in English writing (Liu Xiaoling and Liu Xinxin, 2009; Xu Fang, 2012; Gao Jun and Yang Lu, 2018). Seven academic writings, mainly involving the use of language corpora for teaching academic English writing to students (Chen Yeyu and Xiang Xinni, 2015), as well as the use of corpora to analyze the characteristics of learners' English academic writing (Wang Jingjing and Lv Zhongshe, 2017; Lou Baocui and Yao Wenting, 2019; Lou Baocui and Qiu Mengyao, 2020; Li Xiaohong, 2021).

In addition, 8 reviews and 8 corpus construction papers (accounting for 5.5% of the total) pointed out that there is still great room for improvement in corpus driven English writing teaching, such as the use and existing problems of metadiscourse in academic English writing by graduate students (Zuo Xiuyuan, 2023); The interface research of corpus and discourse analysis has the problem of how to interface (Xu Fang, 2015); Using corpora to provide feedback on student writing (Wang Ying and Li Zhenyang, 2012); There are still problems with the use of logical connectors (Wu Xiaochun, Yang Zhong, Du Ping, 2011). In addition, it also pointed out the significant significance of building an English oral and written language corpus for Chinese students

(Wang Lifei and Wen Qiufang, 2007).

In short, from a content perspective, corpus-based English writing teaching in China has received widespread attention and research over the 23 years from 2001 to 2023. The corpus method has positive effects in the construction of English writing scoring index models, academic writing difficulties, and the cultivation of rhetorical abilities. By comparing and analyzing student compositions, corpora can help teachers identify problems in academic English writing and provide targeted training. In addition, corpus surveys and comparative studies have also achieved some results, such as improving students' interest in writing, solving the problem of inefficient writing teaching, and conducting comparative studies on different corpora.

In addition, the study also found that factors such as vocabulary, semantic prosody, and chunks have a significant impact on English writing. By studying the vocabulary usage characteristics, chunk teaching, and item matching of Chinese college students in English writing, their writing ability and vocabulary richness can be improved. However, despite some progress, there are still some problems and challenges. For example, there is still room for improvement in corpus driven writing teaching, and further research is needed on the use of metadiscourse in academic English writing, the interface between corpora and discourse analysis, and how to use corpora to provide effective feedback for student writing. In summary, corpora have broad application prospects and important significance in writing teaching. Further in-depth research and exploration are needed in the future to continuously improve writing teaching methods, enhance students' writing ability and expression level.

5. Summary

This article reviews and analyzes 145 core journal papers published by China National Knowledge Infrastructure (CNKI) from 2001 to 2023, and analyzes the research situation, research objects, and research content of corpus-based English writing teaching in China. It reveals its development progress and shortcomings, which will also become a focus of future research. Firstly, it is necessary to break the paradigm dominated by college and graduate students, and increase efforts to research corpus based English writing teaching for primary and secondary school students. Secondly, the application of corpora requires teachers to possess certain skills and abilities. On the one hand, teachers should possess skills in corpus retrieval and analysis, as well as the ability to correctly interpret data, and improve these skills through teacher training and professional development. In this way, they can effectively utilize corpora to support student writing instruction and ensure correct application effects. On the other hand, teachers should also encourage students to maintain autonomy and creativity. Although corpora provide rich materials and examples, relying solely on templates and sentence structures may limit students' expressive abilities. Therefore, teachers should encourage students to engage in innovative and personalized writing to cultivate their unique writing style. Thirdly, the effectiveness and reliability of a corpus depend on the quantity and quality of the data used. Although there are currently a large number of corpora available for use, there is still a need for more rich and diverse corpora to support research on writing teaching, and accurate and precise annotation is needed to improve data credibility. Moreover, the content in the corpus is diverse, making it difficult for

students to find suitable example sentences and corpus. To solve this problem, artificial intelligence technology can be combined with corpus applications to develop intelligent assisted writing tools, providing personalized writing feedback and suggestions, as well as more accurate assisted writing support, to help students improve their writing quality.

References

- [1] Johns T. Should you be persuaded: Two samples of data-driven learning materials[J].English language research journal,1991 (4): 1-16.
- [2] CAI Shao-lian. A corpus-based empirical study of English writing teaching [J]. Foreign Language Teaching, 2008 (06): 61-64+68.
- [3] Dong Jihua, Dong Shuai, Wang Hao. Establishment and application of English writing Scoring index model based on text features [J]. Chinese Journal of Examinations, 2023 (02): 42-48. (in Chinese)
- [4] Gao Jun, Yang Lu. A study on the use of Lexical Blocks in Academic English for graduate students of science and engineering [J]. Journal of Foreign Languages, 2018,34 (06): 77-85. (in Chinese)
- [5] He Anping, Xu Manfei. A study of Small Words in Chinese college students' oral English [J]. Foreign Language Teaching and Research,2003(06):446-452+481. (in Chinese)
- [6] He Huaqing. Analysis of lexical errors in the Writing of non-English majors: a corpus-based study [J]. Journal of Foreign Languages,2009(03):2-9.
- [7] Jin Zhiru. The role of Chunk Theory in English Writing Teaching [J]. Continuing Education Research,2011(02):147-149. (in Chinese)
- [8] Li Jianping. A comparative study of English writing vocabulary length between Chinese and American Middle School students [J]. Contemporary Education Science,2013 (18): 46-48.
- [9] Li Jingquan, CAI Jinting. Misuse of Articles in Chinese students' English Writing: a corpus-based study [J]. Journal of Foreign Languages Institute of PLA,2001(06):58-62.
- [10] Li Ke, Zhang Yuan. The composition and investigation of English Learners' Writing Rhetoric Ability [J]. Foreign Language Circle,2021(03):55-63. (in Chinese)
- [11] Li Xiaohong. Functional features of phrase frame in academic English writing: A Comparative study based on the Chinese and English doctoral thesis Introduction Corpora [J]. Foreign Language audio-visual Teaching,2021(01):98-104+16.
- [12] Liang Yan, Feng You, Cheng Liangkun. A review of empirical research on Chinese corpora in recent ten years [J]. Journal of the PLA Institute of Foreign Languages,2004 (06):50-54.
- [13] Liang Yan, Feng You, Cheng Liangkun. A review of empirical research on Chinese corpora in recent ten years [J]. Journal of the PLA Institute of Foreign Languages,2004 (06): 50-54.
- [14] Liu Guobing, Zhang Xiailian. Study on collocation valences of academic English verbs from a corpus-driven perspective [J]. Foreign Language Audio-visual Teaching, 2021 (01):105-111+17. (in Chinese)
- [15] Liu Ping, Wu Liangping, Liu Liya. Research on the application of CQPweb in ESP writing teaching [J]. Foreign Language Field,2016(05):11-19.
- [16] Liu Xianghong, Wang Jiayi, Xia Yanping et al. A corpus-based perspective on the lexical features of engineering students' English writing [J]. Journal of Foreign Languages, 2009 (02):121-123.
- [17] Liu Xiaoling, Liu Xinxin. A corpus-based study on lexical block structure and Pragmatic function of college students' written Words [J]. Foreign Language,2009,6(02):48-53.
- [18] Lou Baocui, Qiu Mengyao. A study on the characteristics of learners' English academic Writing reports [J]. Foreign Language Studies, 2019,37(06):67-73. (in Chinese)
- [19] Lou Baocui, Wang Mimi. A comparative study on the corpus of academic English lexical Chunks for Chinese and foreign science scholars [J]. Shandong Foreign Language Teaching, 2021, 42(06):30-41.
- [20] Lou Bao-Cui, Yao Wenting. The Use of Position Adverbs in learners' Academic English Writing [J]. Journal of Henan Normal University (Philosophy and Social Sciences Edition),2019,46(03):114-120.
- [21] Lou Baocui. Application of corpora in Academic English teaching for graduate students [J]. Academic Degrees and Graduate Education,2020(07):51-56. (in Chinese)
- [22] Lou Baocui. A Comparative analysis of reporting Verbs and stance Expressions in Academic Papers of Chinese and English University students [J]. Shandong Foreign Language Teaching,2013,34(02):50-55. (in Chinese)
- [23] Lou Jie. On the influence of Web-based corpus assistance on the Use of Chunks in English Writing [J]. Foreign Language and Language Teaching,2011(06):61-64. (in Chinese)
- [24] Lu Haiyan. A quantitative and comparative study on the use of personal reference in College English Writing [J]. Journal of Henan Normal University (Philosophy and Social Sciences Edition),2007(04):157-158.
- [25] Shang Zhihui. An Exploration of Online English Writing Teaching Model -- The Application of Process Style Teaching Method in multimedia and Internet environment [J]. Journal of Distance Education,2008(03):45-48+17.
- [26] Wang Jingjing, Lu Zhongshe. A study on author self-referential language in academic English writing for science and engineering doctoral students [J]. Foreign Language Field, 2017 (02):89-96. (in Chinese)
- [27] Wang Lifei, Chen Gong. The characteristics of split sentences in College students' English Writing: A corpora-based study [J]. Foreign Language Teaching and Research,2008(05):362-367+401. (in Chinese)
- [28] Wang Lifei, Wen Qiufang. The construction and research review of "Chinese Students' English Spoken and Written Corpus" [J]. Journal of Foreign Languages,2007(01):22-28.
- [29] Wang Lifei, Zhang Yan. A corpus study on the use of high-frequency Verbs in college students' English argumentative essays [J]. Foreign Language Teaching and Research,2007 (02):110-116+160-161. (in Chinese)
- [30] Wang Yi, Xuan An. The Impact of Web-based interactive autonomous Learning Model on college students' English Writing Level and motivation [J]. Foreign language audio-visual teaching,2010(05):48-52.
- [31] Wang Yi. The impact of Blended Learning on College students' English Writing Proficiency [J]. Open Education Research, 2011, 17(02):81-87. (in Chinese)
- [32] Wang Ying, Li Zhenyang. A review of research on electronic feedback models in Foreign second Language writing [J]. Foreign Language Audio-visual Teaching,2012 (04): 11-16. (in Chinese)
- [33] Wu Xiaochun, Yang Zhong, Du Ping. A review on the use of logical connectives in Interlanguage Discourse [J]. Journal of Beijing Jiaotong University (Social Sciences),2011, 10(01): 126-130.

- [34] Xi Liyun, Zhao Changyou, Self-built College English Writing Corpus and College English Writing Teaching [J]. Academic Circles, 2013(S1):264-267. (in Chinese)
- [35] Xu Fang, A review of academic English writing Research [J]. Foreign Language Teaching and Research, 2015, 47(01):94-105+160-161.
- [36] Xu Fang, A study on the Use and development of academic Lexical Chunks for Chinese Learners [J]. Chinese Foreign Language, 2012, 9(04):51-56.
- [37] Xu Guoqin. A corpus-based innovative research on the development of vocabulary in higher Vocational English Writing [J]. Vocational Education Forum, 2017(11):76-80.
- [38] Yang Xiaoqiong, Dai Yuncui. A practical research on College English autonomous writing teaching model based on correction network [J]. Foreign Language Audio-visual Teaching, 2015(02):17-23. (in Chinese)
- [39] Yu Li, Guo Pingjian. A study on the enthusiasm of art students in English Writing under the environment of information Technology [J]. Modern Educational Technology, 2008 (08): 72-77+71.
- [40] Zhang Huiping, Liu Yongbing. English vocabulary learning and concept transfer: A case study of verb collocation, class joining and semantic rhyme [J]. Foreign Languages and Foreign Language Teaching, 2013(06):38-43.
- [41] Zhang Jidong, Liu Ping. Investigation and analysis of causative structures and corresponding lexical phenomena in Chinese College Students' English Writing [J]. Foreign Language Research, 2005(03):35-39+80.
- [42] Zhang Zhijie. Problems of Semantic rhyme in College Students' English Writing [J]. Journal of Northeast Normal University (Philosophy and Social Sciences Edition), 2011 (02): 225-226.
- [43] Zhao Weibin. Quantitative and Comparative analysis of the Use of logical connectives in Chinese Students' English Compositions [J]. Foreign Language Teaching, 2003(02):72-77. (in Chinese)
- [44] Zhu Huimin, Wang Junju. Lexical richness development in English Writing: a longitudinal study based on self-built corpora [J]. Foreign Language Circle, 2013(06):77-86.
- [45] Zuo Xiuyuan. Approaches to the cultivation of graduate students' metacutterance ability in Academic English: A review of a Corpus-based Comparative Study of metacutterance in Academic English [J]. Foreign Language audio-visual Teaching, 2023(02):102.