

# A Research Overview of Corpus-Assisted Enhancement of English Writing Proficiency

Zhong Aiguo

Yangzte University China

\*Corresponding author: [973216290@qq.com](mailto:973216290@qq.com)

**Keywords—** Corpus, English writing instruction, writing ability, research overview

Received: 05 Feb 2026;

Received in revised form: 03 Mar 2026;

Accepted: 08 Mar 2026,

Published on: 11 Mar 2026

©2026 The Author(s). Published by TheShillonga. This is an open-access article under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>)



**Abstract—** With the deep development of big data and artificial intelligence technologies, corpus research has garnered increasing attention and recognition. Initially, corpora were collections of speech materials aimed at linguistic research; now, they serve as deeply annotated knowledge resources supporting knowledge discovery and extraction. One of the most popular research areas is the integration of corpora with teaching. Studies have found that corpora, as an empirical research tool, provide rich authentic language materials for English writing instruction. This exposure helps students encounter genuine language use, improving the accuracy and authenticity of their writing. We recognize the multifaceted advantages of applying corpora in English writing instruction. These benefits include aiding students in identifying and correcting writing errors, enhancing language awareness, enriching writing resources, and fostering innovation in teaching models. The study categorizes the application of corpora into two main types: indirect and direct use. Indirect use of corpora reveals common issues in English writing and provides theoretical basis and practical guidance for improving teaching resources and methods. Direct use of corpora focuses more on practical operations, showing more significant effects in writing instruction by correcting specific writing errors, and innovating teaching models and methods, thereby directly promoting students' writing skills. Although research on corpus-assisted English writing abilities has yielded considerable results, there is still room for improvement. It is suggested that research outcomes should be closely integrated with frontline teaching practices and expanded research samples. Teachers need to incorporate corpus tools and methods into everyday teaching to achieve mutual promotion of research and teaching. Moreover, educational departments and schools should provide professional training to enhance teachers' corpus knowledge and application skills. Additionally, students' attitudes towards and proficiency in using corpora should be

*emphasized, encouraging them to actively use corpus resources for exploratory learning and self-correction. Future research could further explore the retrieval efficiency of corpora and the intelligence level of their teaching applications, combining interdisciplinary research to comprehensively understand and explore the overall role of corpora in English writing instruction, thereby better advancing the role of corpora in improving English writing skills.*

## I. INTRODUCTION

A corpus is an electronic collection of texts compiled according to specific criteria, designed to represent a language or its various varieties and styles, thereby facilitating linguistic description and analysis (Kennedy, 1998; Liang, Li, & Xu, 2011). Originating in the 1960s, corpus linguistics once ranked among the fastest-growing disciplines within linguistics. Although initially met with skepticism from the rationalist Chomskyan school, it rapidly expanded from Europe to the global stage due to its empirical research methods, focus on the social functions of language, and advantages in historical, synchronic, and sociolinguistic studies (Gui, 2011). It has gained widespread popularity across many application-oriented fields, among which corpus-based language teaching research constitutes a major branch (Shirato & Stapleton, 2007; Gilquin, 2015; Dose-Heidelmayer & Götz, 2016; Pérez-Paredes & Bueno-Alastuey, 2019; Wang, 2020). Corpora can be categorized by collection method into written and spoken corpora, and by language variety into native, foreign, and lingua franca corpora (Liang et al., 2011). Research in corpus linguistics began in China in the late 20th century (He, 1997) and developed rapidly in the early 21st century, sparking a surge of studies in foreign language teaching (He & Xu, 2003; Wang & Zhou, 2004; Wei, 2004; Liang et al., 2004).

A number of corpus-based studies focusing on English learners in China have also emerged (Gui & Yang, 2003; Dong & Chu, 2010; Liang, Li, & Xu, 2010), introducing new elements, dynamics, and methods to the challenging and often slow-progress domain of English writing instruction. However, a review of the literature reveals a relative scarcity of comprehensive summaries on corpus-assisted writing pedagogy. Most existing studies

employ quantitative methods, with qualitative research and syntheses being less common (Zhou & Wang, 2022; Ji, 2022; Hong, Wang, & Li, 2022; Huang & Wang, 2021). Based on a review of relevant literature, this study outlines the conceptual foundations of corpora and writing instruction, as well as the main approaches to using corpora in assisting English writing teaching. It aims to help researchers and educators better utilize corpora to support English writing instruction. By providing a summary, this research seeks to offer recommendations for future studies on corpus-based writing pedagogy, reduce the difficulties teachers and students face when using corpora for writing teaching and learning, enhance their understanding and motivation for using corpora, improve the efficiency of corpus utilization, and ultimately contribute to the advancement of English writing proficiency.

## II. THE CONNOTATION OF CORPUS

The term "corpus," derived from Latin meaning "body," refers in modern linguistics—particularly in corpus linguistics—to "a systematically organized collection of authentic language materials intended for research purposes" (Liang, 2016). Renowned linguist Crystal (1991) described a corpus in his work as a body of linguistic data, whether spoken or written, often in transcribed form, serving as a foundation for language description or a means of testing linguistic hypotheses.

In general, while definitions of a corpus vary among scholars, there is broad consensus regarding its essential characteristics. A corpus typically exhibits four fundamental features: (1) a large volume of language data; (2) authenticity and representativeness of the materials; (3) storage and retrieval via computer or cloud-based systems; and (4) application in linguistic investigation and analysis.

The integration of corpora into linguistics has introduced new research content and methodologies for second language acquisition. On one hand, corpora provide researchers with extensive authentic language materials, enriching research topics; on the other hand, corpus retrieval software enables systematic and efficient processing of language data, offering clear and intuitive statistical results that support innovative research approaches (Sinclair, 1991; Atkins & Clear, 1992; Gui & Yang, 2003; Yang, 2002; Sinclair, 2004; McEnery, Xiao, & Tono, 2006; Huang & Li, 2007; Yu, 2009). For instance, corpus-assisted English language teaching can be applied in phonetics, grammar, vocabulary, and reading instruction (He, 2017). Teaching materials derived from spoken or written corpora can supply learners with abundant authentic language samples in natural contexts (Warren, 2016).

With advancements in computer technology and deeper linguistic research, corpus-based methods have become a vital approach to collecting language data. Corpus linguistics relies on authentic language usage and employs statistical analysis to reveal patterns of language use, offering new pathways for linguistic studies.

As more corpora become freely accessible to learners, their application in foreign language teaching continues to expand. However, the effectiveness of corpora in instructional contexts also depends on teachers' expertise in corpus linguistics. For example, Caines and McCarthy (2016) pointed out that due to limitations in foreign language teachers' knowledge of corpus linguistics, research findings and technologies related to spoken corpora have not yet been widely or effectively integrated into language teaching.

In summary, while corpus-assisted English language teaching possesses incomparable advantages and potential, its implementation requires three conditions: (1) A greater variety of corpora should be made freely available to English teachers and learners; (2) Teachers need professional training in corpus linguistics, including how to select or construct corpora and how to design corpus-based teaching activities; (3) Students should receive training in using corpora and recognize their dual role in corpus-assisted learning—not only as learners but also as "researchers" capable of conducting in-depth

exploratory studies of English through corpora.

Owing to the significant imbalance between investment and outcomes in corpus-assisted English teaching, very few applied studies in domestic research have implemented corpora in instructional practice. Boulton (2010) accurately summarized the prevailing attitude among teachers toward corpus-assisted instruction as characterized by high difficulty, demotivation, low pedagogical relevance, and inefficiency. Nonetheless, integrating corpora with foreign language teaching holds promise for developing more targeted and efficient English teaching methods.

### III. THE ENGLISH WRITING AND CONNOTATION OF WRITING ABILITY

English writing, as a comprehensive language skill, offers a clear reflection of a learner's English proficiency. Collins and Gentner (1980) defined it as a process involving two main stages: the generation of ideas and the drafting of text. Calkins (1986) further emphasized that writing is not merely a act of recording, but a dynamic process of developing stories and thoughts, underscoring its significance as a language skill. From an information-processing perspective, Hyland (2003) proposed that writing involves organizing fragmented mental information and transforming it into structured written expression. Vicki Urquhart and Monette McIver (2005) suggested that writing enables students to establish connections among different characters, events, and ideas. Through targeted systematic thinking and questioning, it effectively hones logical reasoning and cultivates critical thinking. Thus, writing serves both as a means of communicating experiences and sharing ideas, and as a manifestation of the learner's integrated language awareness and productive language ability.

Writing ability is regarded as the capacity to integrate information, construct texts, and convey ideas through written language. The academic community has extensively discussed the components of writing ability and proposed various evaluation criteria (Wen, 2010). Artiga (1992) identified five dimensions for assessing writing competence: language use, writing conventions, content process, style, and evaluative techniques. Yoon and Hirvela (2004) broke down writing ability into the skills of

conveying information, controlling the writing process, mastering writing thinking, and evaluating writing. Writing is a process through which people express thoughts and emotions linguistically; it is also a progressive process of identifying, analyzing, and solving problems. In this sense, writing ability equates to problem-solving ability (Liu, 1999). Zhao (2008) further proposed that writing ability encompasses five aspects: basic writing skills, strategic competence, linguistic competence, discourse rhetorical competence, and sociocultural competence. Specifically, students should be able to clearly articulate their stance on familiar topics and support their viewpoints with persuasive arguments. Moreover, they need to be capable of systematically describing personal experiences and events around them, such as extracurricular activities and campus competitions, while employing rhetorical techniques to ensure clear structure and vivid narration. Students should also be able to engage in basic discussions of sociocultural content using the internet and mobile devices—for example, exploring traditional culture and festival customs.

It is evident that writing ability is a comprehensive skill essential for students learning English. Therefore, corresponding writing instruction holds significant importance. Although corpus-based research is not yet common domestically, studies have shown that corpora play an indispensable role in enhancing English writing proficiency, warranting further analysis of the feasibility of integrating corpora into writing instruction.

#### **IV. CURRENT RESEARCH ON CORPUS-ASSISTED ENHANCEMENT OF ENGLISH WRITING SKILLS**

Since the late 1920s, educators and scholars have begun to explore the use of corpora for a wide range of research purposes. From both theoretical linguistic perspectives and practical foreign language teaching viewpoints, corpora have demonstrated positive impacts and effectiveness in improving the writing skills of English learners (Li, 2024). Based on a review of existing literature related to corpus-assisted English writing development, as well as Leech's (1997) classification of corpus applications in English language education, current studies can be broadly grouped into two main approaches: (1) Indirect

use of corpora to enhance English writing skills, which primarily involves teaching resource development and language assessment; (2) Direct use of corpora to improve English writing ability, focusing on how to utilize corpora in classroom teaching and how to instruct students in using corpora independently.

This paper likewise adopts these two pathways to examine research on corpus-assisted enhancement of English writing proficiency.

##### **4.1 Research on the Indirect Use of Corpora**

Studies on the indirect use of corpora to enhance English writing skills primarily focus on analyzing errors in English writing or examining the effects of corpora on writing performance. Through such analysis, researchers aim to identify factors that hinder the improvement of writing competence and draw pedagogical implications to refine teaching resources, adjust instructional methods, or improve language testing systems. Ultimately, these studies offer recommendations to facilitate teaching activities targeted at boosting English writing proficiency.

Research on the indirect application of corpora can be categorized into two main strands: (1) Error analysis in English writing, (2) Analysis of the impact of corpora on English writing.

##### **4.1.1 Error Analysis in English Writing**

Error analysis in English writing is conducted through two main approaches: self-collected essays and the use of existing corpora.

##### **4.1.1.1 Analysis Based on Self-Collected Essays**

Most studies adopting this approach gather writing samples from secondary school students, often in large quantities, to ensure representativeness and objectivity.

These studies employ corpus-based methodologies by collecting hundreds to thousands of routine writing samples from junior and senior high school students. Through systematic analysis, they identify common errors made by foreign language learners and propose comprehensive summaries aimed at helping students reduce mistakes, enhance writing techniques, and ultimately increase their interest in English learning while improving writing proficiency (Bett, 2010; Ondrakova, 2016; Talakoob & Mansour, 2017; Shi & Yu, 2019; Zhang, 2020; Leng, 2020).

Some scholars shift their focus from secondary school

students to undergraduate or graduate students, concentrating specifically on errors in English academic writing.

These studies typically collect essays from university students to construct small-scale specialized corpora for analyzing writing errors. Through comparative and statistical examination, researchers investigate the causes of linguistic inaccuracies, reveal characteristics of adjective usage, explore lexical features in learner writing, and summarize pedagogical implications for English writing instruction (Shamsudin, 2010; Xu, 2014; Phuket, 2015; Liu & Wang, 2015; Divsar, 2017; Qian, 2021; Sun, 2018).

#### 4.1.1.2 Analysis Using Existing Corpora

While many studies utilize self-constructed corpora, the majority—particularly in China—rely on established corpora for research. A prominent example is the Chinese Learner English Corpus (CLEC), which has been extensively employed in domestic studies. Comprising over one million words collected from secondary school to university students, CLEC serves as a valuable resource for enhancing the writing proficiency of Chinese English learners across various educational stages (Gui & Yang, 2003).

These studies primarily focus on the English writing of Chinese university students, using CLEC to conduct in-depth analyses. The research identifies the causes of common errors, attributing them mainly to three factors: interlingual transfer, intralingual transfer, and influences from the learning environment. Consequently, recommendations are made for instructors to refine teaching methodologies, with particular emphasis on verb usage. In addition to enhancing comparative instruction between English and Chinese, educators are encouraged to improve students' awareness of diverse verb collocations and applications. Beyond increasing language input, it is also essential to provide more opportunities for language output, enabling learners to master and flexibly apply various lexical and grammatical forms in practical contexts, thereby improving the overall effectiveness of English writing instruction (He, 2004; Lin, 2004; Wang & Zhang, 2007; He, 2009; Zhu, 2009; Cheng, 2010; Li, 2022).

#### 4.1.2 Analysis of the Role of Corpora in English Writing

Research on the impact of corpora on English writing also adopts various approaches, particularly by examining how corpora influence teaching and assessment models. These studies aim to provide insights and references for refining instructional and evaluative frameworks, thereby enhancing the effectiveness of English writing pedagogy and increasing students' adaptability to corpus-assisted writing instruction.

Dong and Chu (2010) utilized a computer-based corpus platform to explore a data-driven approach to English writing instruction. Their findings demonstrate that corpus tools and methods effectively monitor the accuracy of language output and ensure fluency, confirming the practical value and potential of corpora in writing pedagogy. In a comparative experiment, Cai (2008) investigated differences between corpus-based teaching methods and traditional writing instruction. The results revealed that corpus-assisted teaching offers significant advantages in delivering writing theory, improving students' self-evaluation skills, and promoting autonomous learning. Additionally, Cai emphasized another key benefit: corpora not only stimulate students' interest in writing but also tangibly enhance their writing proficiency. Further expanding on this line of inquiry, Blanchard et al. (2013) examined the role of corpora in the assessment of English writing. They reported the development of a new corpus comprising texts produced by non-native English writers. This corpus demonstrated considerable utility in native language identification, grammatical error detection and correction, and automated essay scoring. The development of such a resource provides new tools and methodologies for teaching and evaluating English writing among non-native speakers. These findings are further corroborated by the research of Wang and Wu (2021), who emphasize the dual role of corpora in English teaching: corpora serve not only as instructional resources but also as pedagogical aids. They argue that applying corpora to the teaching of continuation tasks can effectively assist students in predicting discourse themes, identifying genre features, comprehending plot development, inferring meanings of unfamiliar words, and discerning characters' emotional attitudes—thereby offering new directions for pedagogical design. Du (2022) further points out that in continuation writing, corpora help students accumulate

diverse and authentic language materials, leading to richer and more accurate expression, as well as more vivid character portrayal. Similarly, Zhang (2022) contributes unique insights through a multi-corpus study of continuation tasks. By employing AntConc and a self-built corpus based on three versions of high school English textbooks, Zhang conducted keyword analyses of both source texts and graded continuation samples. The study reveals that keywords sensitively reflect lexical style and spelling errors, which to some extent indicate the quality of continuation writing.

The studies above illustrate both commonalities and variations in how different scholars indirectly utilize corpora to investigate English writing proficiency. In summary, through the indirect application of corpora, researchers have systematically identified factors hindering the development of writing ability—such as interlingual transfer, intralingual transfer, influences of the learning environment, lexical errors, and grammatical inaccuracies—while also exploring how corpora can enhance both instructional and assessment models. These improvements, whether direct or indirect, ultimately contribute to the development of students' English writing skills. Nevertheless, translating these research insights into tangible classroom benefits requires extensive practical application and adaptation by frontline teachers and researchers. Without empirical validation, the practical impact of such theoretical implications remains limited.

#### 4.2 Research on the Direct Use of Corpora

Studies on the direct application of corpora to enhance English writing skills primarily focus on utilizing corpora to identify and correct errors in writing, with the aim of uncovering factors that influence the effectiveness of such methods and proposing recommendations for improving writing proficiency. Alternatively, this line of research explores how the direct integration of corpora can transform English writing instruction and examines the potential improvements in pedagogical practices facilitated by corpus tools. Research on the direct use of corpora can be categorized into two main strands: (1) Analysis of direct corpus use for error correction in English writing, (2) Analysis of the role of direct corpus application in teaching English writing.

##### 4.2.1 Analysis of Corpus-Based Error Correction in

##### English Writing

Research indicates that the direct application of corpora for error correction in English writing yields insights across multiple dimensions, such as error types, corrective feedback, and language awareness.

Dolgova (2019) conducted a study focusing on how learners use corpus tools to correct various types of writing errors and examined the effectiveness of these tools across different error categories. The results revealed that learners' success rates in error correction were closely related to the specific types of errors they encountered. Crosthwaite, Storch, and Schweinberger (2020) further explored the role of written corrective feedback within corpus-assisted error resolution. Their research investigated how such feedback influences the efficiency with which learners employ corpus tools to address errors. The study emphasized that teachers should carefully consider whether their written corrective feedback facilitates effective interaction between learners and corpus resources when integrating these tools into writing instruction. Lou (2020) identified issues related to lexical collocation and pragmatic accuracy in the academic English writing of graduate students. The study proposed the use of concordance analysis from specialized corpora to enhance linguistic awareness and rhetorical skills in academic writing.

##### 4.2.2 Analysis of the Role of Corpus-Based Instruction in English Writing

Research on the direct application of corpora in English writing instruction has a relatively longer history and greater complexity, yielding more specific and representative conclusions.

Most studies in this category innovate English writing teaching models through corpora, thereby enhancing students' specific writing skills.

Some scholars focus particularly on abstract writing competencies.

For example, Yoon and Hirvela (2004) conducted a study on corpus use in two ESL academic writing courses. Their research specifically examined students' corpus usage behaviors and their perceptions of the advantages and disadvantages of corpora as a second language writing tool. Both qualitative and quantitative data indicated that, overall, students found the corpus approach beneficial for the development of L2 writing skills and reported

increased confidence in writing in their second language. Cotos (2014) conducted a study based on multiple corpora, investigating the implementation of activities drawing on both native-speaker and learner corpora. The research demonstrated the effectiveness of data-driven learning methods in raising learners' awareness of observed language phenomena, promoting knowledge acquirement, and improving the use of target language forms. Results showed that students exhibited improvements in the frequency, diversity, and accuracy of adverb usage, with those exposed to corpora containing their own writing making more significant progress. Li (2024) adopted a quantitative research paradigm, conducting a 10-week teaching experiment. Using SPSS 20.0 software, the study compared pre- and post-test scores between experimental and control classes, and summarized data from post-experiment questionnaires and interviews. The findings revealed that the corpus-assisted continuation task teaching model effectively enhanced students' continuation writing performance and overall competence, increased their interest in English learning, and contributed to the development of autonomous learning abilities.

Meanwhile, a greater number of scholars have focused on enhancing specific English writing skills

Many of them often utilize corpus as pedagogical methods through the implementation of innovative teaching models.

For instance, Chen (2014) implemented a 12-week "Input–Output–Consolidation" teaching model that integrated lexical chunk instruction with writing practice. By comparing pre- and post-test results, the study demonstrated significant improvements in students' use of lexical chunks and overall writing performance. Ying (2021) similarly proposed a new pedagogical model for corpus-assisted instruction, introducing a corpus-based approach for teaching practical English writing in senior high schools. This model combines data-driven learning principles with frequency-guided language teaching methods, structured around seven steps: corpus selection, output stimulation, input facilitation, independent writing, evaluation and revision, consolidation and refinement, and the development of a self-built corpus. Building on previous studies, Chen (2024) also adopted a corpus-based instructional model for English writing, employing a

student-centered data-driven learning approach known as the CCC model (Collaboration–Consolidation–Construction). The findings revealed that: (1) Corpus-assisted English teaching significantly improved the lexical richness of junior high school students' writing. (2) It notably enhanced students' awareness of vocabulary use strategies related to lexical diversity. (3) It substantially elevated the overall quality of students' writing.

Chen's study demonstrates that corpora can effectively increase lexical richness in student writing by positioning learners in the role of "researchers." This approach addresses previous limitations in corpus-assisted English teaching, such as teacher-dominated instruction, limited interactivity, and low efficiency.

Other scholars have employed existing or self-constructed corpora to support English writing instruction through practical teaching applications.

For example, Guan (2022) conducted a study using the WebCorp corpus retrieval tool to provide rich linguistic input for college English writing courses, helping students accumulate high-frequency vocabulary. Through this process, students were able to summarize and master typical word usage and collocations. During writing tasks, Guan guided students in using WebCorp to search for key terms and gather relevant language materials to enrich their writing. After drafting, students again utilized WebCorp to check and refine questionable collocations and sentence structures, thereby improving the overall quality of their compositions. In this study, both teachers and students actively engaged as users of the online corpus, collectively participating in the teaching and learning process. Han and Zong (2023) made use of multiple corpora in their research. In a writing revision course for senior high school students, they introduced two corpora—COCA (Corpus of Contemporary American English) and SkELL (Second Language English Learner Corpus)—to enhance the authenticity of students' language expressions. The study emphasized the student's role as the primary user of web-based corpora. Through independent searching and application of corpus resources, students were able to more effectively refine their writing, making their language use more aligned with native speaker norms. Liu (2023) adopted a more targeted

approach by constructing a self-built corpus for analysis. Based on error analysis, output theory, and in-depth corpus examination, the study identified five common error types among high school students in the use of non-finite verbs: errors in tense, form, confusion between infinitives and gerunds, confusion between finite and non-finite verbs, subject-verb agreement errors, and spelling mistakes. Among these, the most frequent was the confusion between finite and non-finite verbs. The study further explored the causes of these errors, attributing them to three main factors: intralingual transfer, interlingual transfer. Ultimately, the research demonstrated that instructional designs incorporating self-built corpora are highly effective in addressing non-finite verb errors in English writing. This teaching method significantly improved students' writing performance and showed markedly better outcomes compared to traditional error correction approaches. Finally, Li (2024) provided a comprehensive investigation into specific English writing competencies. Grounded in corpus linguistics, scaffolding theory, and data-driven learning theory, Li analyzed collected data and found that applying a corpus-based scaffolding approach to high school English writing instruction notably enhanced students' initiative in writing. Reflected in their written expressions, students showed significant improvement across multiple dimensions: lexical richness, collocational authenticity, sentence variety, discourse coherence, and writing speed. Both writing scores and overall quality demonstrated steady progress.

The aforementioned studies reveal both commonalities and distinctions among scholars in their direct application of corpora to enhance English writing proficiency. Overall, corpora serve as an effective tool to help students identify and correct errors in areas such as lexical collocation, pragmatic accuracy, tense, verb forms, and the use of infinitives versus gerunds. These studies underscore the importance of error typology, corrective feedback, and language awareness in improving writing competence. Furthermore, the direct integration of corpora has driven innovation in English writing pedagogy, exemplified by models such as the "Input-Output-Consolidation" framework and the CCC (Collaboration-Consolidation-Construction) corpus-assisted teaching model. By incorporating lexical chunk instruction,

data-driven learning, and frequency-based language teaching methods, these approaches have significantly enhanced students' writing skills—including lexical richness, collocational accuracy, syntactic variety, and discourse coherence. Compared to research on the indirect use of corpora, studies involving direct application demonstrate more tangible and observable improvements in writing outcomes. However, such studies remain relatively limited in number. They often involve longer implementation periods, greater operational complexity, and multiple influencing factors. Moreover, due to relatively small sample sizes, the generalizability of their findings is somewhat constrained. There is a need for further collaboration among researchers to expand sample sizes and diversity through practical application, in order to validate and enhance the representativeness of these results.

## V. CONCLUSION AND IMPLICATION

Overall, this review demonstrates that corpus-based approaches in English writing instruction encompass a wide range of study populations, including learners from secondary school to undergraduate and graduate levels, reflecting the applicability and value of corpora across various educational stages. Moreover, the research covers multiple linguistic dimensions—such as vocabulary, grammar, syntax, and discourse—highlighting the comprehensive role of corpora in enhancing writing pedagogy. It is evident that corpora play a significant role in supporting the development of English writing proficiency.

First, the application of corpora offers new perspectives and methodologies for English writing instruction. By providing access to abundant authentic language materials in real communicative contexts, corpora enable students to engage with natural language use, which is crucial for improving the accuracy and authenticity of their writing. Corpus-assisted teaching models, such as the "Input-Output-Consolidation" framework and the CCC (Collaboration-Consolidation-Construction) model, incorporate systematic instructional procedures that have been shown to significantly enhance students' lexical richness, grammatical accuracy, and overall writing quality.

Second, research on corpus-assisted writing development employs diverse methodologies, which can be broadly categorized into indirect and direct uses of corpora. Findings indicate that both approaches contribute substantially to the improvement of students' writing abilities. Specifically, corpora help learners identify and correct errors, raise metalinguistic awareness, enrich writing resources, and foster innovation in teaching models. Furthermore, corpora provide robust empirical evidence and diverse research perspectives that continue to inform and refine English writing instruction.

Studies involving the indirect use of corpora have revealed common challenges in English writing—such as interlingual transfer, intralingual transfer, environmental influences, and specific lexical or grammatical errors—and have proposed corresponding solutions. These investigations provide both theoretical foundations and practical guidance for refining teaching resources and instructional strategies.

In contrast, research on the direct application of corpora tends to focus more on practical implementation, demonstrating more pronounced effects within writing instruction. By enabling real-time error correction and supporting innovative teaching models and methods, the direct use of corpora contributes tangibly to the enhancement of students' writing skills. Findings indicate that corpus-assisted instruction effectively improves learners' lexical richness, syntactic variety, and discourse coherence, while also fostering greater autonomy and motivation in English writing.

Overall, scholars both domestically and internationally have achieved substantial results in researching the application of corpora to enhance English writing skills; however, there remains considerable room for improvement. It is essential to closely integrate research outcomes with frontline teaching practices and to expand the scope of research samples. Teachers should actively incorporate corpus tools and methods into daily instruction to foster mutual enhancement between research and pedagogy. In this process, teachers' competence in using corpora plays a crucial role. Future studies could investigate teachers' corpus literacy, while educational authorities and schools should provide targeted professional development to strengthen teachers'

knowledge and applied skills in corpus-based instruction. Simultaneously, students' attitudes toward and proficiency with corpora cannot be overlooked. Learners should be positioned as “researchers” in the corpus-assisted learning process, encouraged to actively utilize corpus resources for exploratory learning and self-correction. Moving forward, research could also focus on establishing and refining various types of corpora—especially personalized corpora tailored to different learning stages and needs—to provide more diverse teaching and learning resources for both educators and students.

Furthermore, sustained empirical research on corpus-assisted English writing instruction should be conducted, incorporating continuous feedback to optimize teaching methods and strategies in response to evolving educational demands. Building on previous studies, modern information technologies—such as artificial intelligence and big data analytics—can be leveraged to enhance the efficiency of corpus retrieval and the intelligence of pedagogical applications. Interdisciplinary research integrating knowledge from education, psychology, applied linguistics, and other related fields will contribute to a more comprehensive understanding and exploration of the multifaceted role of corpora in English writing instruction, thereby more effectively advancing the enhancement of English writing proficiency through corpora.

## REFERENCES

- [1] Atkins, S., & Clear, J. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1-16.
- [2] Artigal, J. (1992). Some considerations on why a new language is acquired by being used. *International Journal of Applied Linguistics*
- [3] Bett, J. (2010). Error analysis: The ESL learners' dilemma. *International Journal of Learning*
- [4] Blanchard, D., Tetreault, J., Higgins, D., Cahill, A., & Chodorow, M. (2013). TOEFL11: A corpus of non-native English. *ETS Research Report Series*, 2013(2), i-15.
- [5] Boulton, A. (2010). Data-driven learning: On paper, in practice. *Linguistic Insights*.
- [6] Caines, A., McCarthy, M., & O'Keeffe, A. (2016). [Chapter title unknown]. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology*

- (pp. 348-361). Routledge.
- [7] Cai, S. L. (2008). *Jiyu yuliaoku de yingyu xiezuo jiaoxue shizheng yanjiu* [An empirical study on corpus-based English writing teaching]. *Waiyu Jiaoxue* (Foreign Language Education), (6), 61-64, 68.
- [8] Calkins, C. (1986). *The art of teaching writing*. Heinemann Educational Books.
- [9] Chen, B. M. (2014). *Yuliaoku fuzhu cikuai jiaoxue zai yingyu xiezuo jiaoxue zhong de yingyong* [The application of corpus-assisted chunk teaching in English writing instruction]. *Jichu Yingyu Jiaoyu* (Basic English Education)
- [10] Chen, J. H. (2024). *Jiyu yuliaoku de chuzhongsheng yingyu xiezuo cihui fengfuxing de xingdong yanjiu* [An action research on the lexical richness of junior high school students' English writing based on a corpus] (Unpublished master's thesis). Jimei University.
- [11] Chen, L. Y., & Li, C. Y. (2018). *Yuliaoku shijiao xia de gaozhong yingyu xiezuo pingjiangke moshi tansuo* [Exploring a model for high school English writing commentary classes from a corpus perspective]. *Guangdong Jiaoyu* (Zonghe ban) (Guangdong Education (Comprehensive Edition)), (3), 53-55.
- [12] Collins, A., & Gentner, D. (1980). A framework for a cognitive theory of writing. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 51-72). Lawrence Erlbaum Associates.
- [13] Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, 26(2), 202-224.
- [14] Crosthwaite, P., Storch, N., & Schweinberger, M. (2020). Less is more? The impact of written corrective feedback on corpus-assisted L2 error resolution. *Journal of Second Language Writing*, 49, Article 100729.
- [15] Crystal, D. (1991). Stylistics profiling. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 221-238). Longman.
- [16] Divsar, H., & Heydari, R. (2017). A corpus-based study of EFL learners' errors in IELTS essay writing. *International Journal of Applied Linguistics and English Literature*, 6(3), ?-?. (Missing page numbers)
- [17] Dolgova, N., & Mueller, C. (2019). How useful are corpus tools for error correction? Insights from learner data. *Journal of English for Academic Purposes*, 39, 97-108.
- [18] Dong, H. X., & Chu, S. H. (2010). *Jiyu jisuanji wangluo yuliaoku qudong de yingyu xiezuo jiaoxue moshi yanjiu* [A study on a computer network corpus-driven English writing teaching model]. *Waiyu Dianhua Jiaoxue* (Technology Enhanced Foreign Language Education), (9), 27-32.
- [19] Dose-Heidelmeyer, S., & Götz, S. (2016). The progressive insight into spoken learner language: A corpus-based analysis of use and misuse. *International Review of Applied Linguistics in Language Teaching*, 54(3), 229-256.
- [20] Du, Y. (2022). *Yuliaoku fuzhu gaozhong yingyu duhou xuxie de yingyong shijian* [Application practice of corpus-assisted continuation writing in high school English]. *Zhongxuesheng Yingyu* (English for Middle School Students), (22), 35-36.
- [21] Gillquist, G. (2015). The use of phrasal verbs by French-speaking EFL learners: A constructional and collocational corpus-based approach. *Corpus Linguistics and Linguistic Theory*, 11(1), 51-88.
- [22] Guan, C. (2022). *Jiyu WebCorp yuliaoku de yingyu xiezuo jiaoxue moshi yanjiu* [A study on the English writing teaching model based on the WebCorp corpus]. *Haiwai Yingyu* (Overseas English), (21), 1-2.
- [23] Gui, S. C. (2011). Preface. In M. C. Liang, W. Z. Li, & J. J. Xu, *Yuliaoku yingyong jiaocheng* [A practical guide to corpus applications]. Foreign Language Teaching and Research Press.
- [24] Gui, S. C., & Yang, H. Z. (2003). *Zhongguo xuexizhe yingyu yuliaoku* [Chinese learner English corpus]. Shanghai Foreign Language Education Press.
- [25] Gui, S. C., & Yang, H. Z. (2003). *Zhongguo xuexizhe yingyu yuliaoku (CLEC)* [Chinese learner English corpus (CLEC)]. Shanghai Foreign Language Education Press.
- [26] Han, J., & Zong, H. (2023). *Zhuti yujing xia cujan cihui chuchan de gaozhong yingyu zonghe xiezuo jiaoxue* [High school integrated English writing teaching to promote vocabulary output in thematic contexts]. *Zhongxiaoxue Yingyu Jiaoxue yu Yanjiu* (English Teaching and Research in Primary and Secondary Schools), (10), ?-?. (Missing page numbers)
- [27] He, A. P. (1997). *Tan yuliaoku yanjiu* [On corpus studies]. *Waiyuoyu* (Journal of Foreign Languages), (5), 22-27.
- [28] He, A. P. (n.d.). *Yuliaoku fuzhu yingyu jiaoxue rumen (xiuding ban)* [A beginner's guide to corpus-assisted English teaching] (Revised ed.). Foreign Language Teaching and Research Press.
- [29] He, A. P., & Xu, M. F. (2003). *Zhongguo daxuesheng*

- yingyu kouyu small words de yanjiu [A study on small words in Chinese college students' spoken English]. *Waiyu Jiaoxue yu Yanjiu* (Foreign Language Teaching and Research), (6), 446-452, 481.
- [30] He, H. Q. (2009). Fei yingyu zhuan ye xuesheng xiezuo zhong de cihui cuowu fenxi—Yi xiang jiyu yuliaoku de yanjiu [Analysis of lexical errors in non-English majors' writing: A corpus-based study]. *Waiyujie* (Foreign Language World), (3), 2-9.
- [31] Hong, M. F., Wang, X. Y., & Li, P. (2022). Yuliaoku zai yingyu jiaoxue zhong de yingyong yanjiu zongshu [A review of the application of corpora in English teaching]. *Yuyan yu Wenhua Luntan* (Language and Culture Forum), (2), 265-274.
- [32] Huang, C. N., & Li, J. Z. (2007). *Yuliaoku yuyanxue* [Corpus linguistics]. The Commercial Press.
- [33] Huang, S. Q., & Wang, D. B. (2021). Guonei yuliaoku yanjiu zongshu [A review of corpus research in China]. *Xinxi Ziyuan Guanli Xuebao* (Journal of Information Resources Management), 11(3), 4-17, 87. <https://doi.org/10.13365/j.jirm.2021.03.004>
- [34] Hyland, K. (2003). *Second language writing*. Cambridge University Press.
- [35] Ji, W. T. (2022). Jin shi nian guonei jiyu yuliaoku de waiyu jiaoxue yanjiu zongshu [A review of corpus-based foreign language teaching research in China in the last decade]. *Yingyu Guangchang* (English Square), (21), 35-38.
- [36] Kedney, G. (1998). *An introduction to corpus linguistics*. Addison Wesley Longman.
- [37] Leng, H. A., Tan, K. H., & Guang, Y. L. (2020). Error types in Malaysian lower secondary school student writing: A corpus-informed analysis of subject-verb agreement and copula be. *3L: Language, Linguistics, Literature*, 26(4), 120-132.
- [38] Li, T. T. (2024). Yi yuliaoku wei zhijia de gaozhong yingyu xiezuo nengli tisheng de shizheng yanjiu [An empirical study on improving high school English writing ability with corpus scaffolding] (Unpublished master's thesis). Harbin Normal University.
- [39] Li, Y. (2022). Jiyu yuliaoku de daxue yingyu xiezuo zhuwei yizhi wuyong tanxi [A corpus-based analysis of subject-verb agreement errors in college English writing]. *Hawai Yingyu* (Overseas English), (23), 51-54.
- [40] Li, Y. D. (2024). Yuliaoku fuzhu gaozhong yingyu duhou xuxie jiaoxue de shizheng yanjiu [An empirical study on corpus-assisted continuation writing teaching in high school English] (Unpublished master's thesis). Southwest University.
- [41] Liang, M. C. (2016). *Shenme shi yuliaoku yuyanxue* [What is corpus linguistics?]. Shanghai Foreign Language Education Press.
- [42] Liang, M. C., Li, W. Z., & Xu, J. J. (2010). *Yuliaoku yingyong jiaocheng* [A practical guide to corpus applications]. Foreign Language Teaching and Research Press.
- [43] Liang, M. C., Li, W. Z., & Xu, J. J. (2011). *Yuliaoku yingyong jiaocheng* [A practical guide to corpus applications]. Foreign Language Teaching and Research Press.
- [44] Liang, Y., Feng, Y., & Cheng, L. K. (2004). Jin shi nian woguo yuliaoku shizheng yanjiu zongshu [A review of empirical corpus research in China in the last decade]. *Jiefangjun Waiguoyu Xueyuan Xuebao* (Journal of PLA University of Foreign Languages), (6), 50-54.
- [45] Lin, D. H. (2004). Zhongguo xuesheng yingyu xiezuo zhong de congju cuowu—Yi xiang jiyu yuliaoku de yanjiu [Clause errors in Chinese students' English writing: A corpus-based study]. *Jiefangjun Waiguoyu Xueyuan Xuebao* (Journal of PLA University of Foreign Languages), (3), 49-52.
- [46] Liu, B., & Wang, Y. K. (2015). Fei yingyu zhuan ye yanjiusheng yingwen xiezuo zhong xingrongci shiyong tezhen yanjiu [A study on the characteristics of adjective use in English writing by non-English major postgraduates]. *Zhongguo Waiyu* (Foreign Languages in China), 12(4), 45-53.
- [47] Liu, R. Q. (1999). *Waiyu jiaoxue yanjiu de fazhan qushi* [Trends in foreign language teaching research]. *Waiyu Jiaoxue yu Yanjiu* (Foreign Language Teaching and Research), (1), 7-12.
- [48] Liu, S. (2023). Jiyu yuliaoku de gaozhong yingyu xiezuo fei weiyu dongci jiaoxue yanjiu [A study on teaching non-finite verbs in high school English writing based on a corpus] (Unpublished master's thesis). Central China Normal University.
- [49] Lou, B. C. (2020). *Yuliaoku zai yanjiusheng xueshu yingyu jiaoxue zhong de yingyong tansuo* [Exploring the

- application of corpora in teaching academic English to postgraduates]. *Xuewei yu Yanjiusheng Jiaoyu (Academic Degrees and Graduate Education)*, (7), 51-56.
- [50] McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.
- [51] Ondrakova, J. (2016). *The issue of errors in teaching foreign language*. University of Hradec Kralove.
- [52] Pérez-Paredes, P., & Bueno-Alsteu, M. C. (2019). A corpus-driven analysis of certainty stance adverbs: Obviously, really and actually in spoken and learner English. *Journal of Pragmatics*, 140, 22-32.
- [53] Phuket, P. R. N., & Othman, N. B. (2015). Understanding EFL students' errors in writing. *Journal of Education and Practice*, 6(32), 99-107.
- [54] Qian, Y. L. (2021). *Jiyu yingyong yuliaoku de gaozhong yingyu xiezuo zhi duhou xuxie jiaoxue sheji [Teaching design for continuation writing in high school English based on applied corpus]*. *Xiaoyuan Yingyu (Campus English)*, (45), 153-154.
- [55] Shamsudin, S., & Mahady, N. R. A. (2010). Corpus linguistics based error analysis of first year Universiti Teknologi Malaysia students' writing. [Unknown Source], 1-7.
- [56] Shi, S. Q., & Yu, T. (2019). *Jiyu yuliaoku de cikuai jiaoxue dui gaozhongsheng yingyu xiezuo de yingxiang [The influence of corpus-based chunk teaching on high school students' English writing]*. *Jichu Waiyu Jiaoyu (Basic Foreign Language Education)*, (3), 3-10, 106.
- [57] Shirato, J., & Stapleton, P. (2007). Comparing English vocabulary in spoken learner corpus with an active speaker corpus: Pedagogical implications arising from a case study in Japan. *Language Teaching Research*, 11(4), 393-412.
- [58] Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- [59] Sinclair, J. M. (2004). *How to use corpora in language teaching*. John Benjamins.
- [60] Sun, Z. Y. (2018). *Yunyong yuliaoku fuzhu gaozhong yingyu xiezuo [Using corpora to assist high school English writing]*. *Yushu Wai Xuexi (Gaozhong Ban Shangxun) (Language, Math, and Foreign Learning (High School Early Edition))*, (5), 81-88.
- [61] Talakoob, F., & Mansour, K. (2017). Advanced EFL learners' knowledge of different collocation types. *International Journal of Applied Linguistics and Language Research*, 4(2), 267-276.
- [62] Urquhart, V., & McIver, M. (2005). *Teaching writing in the content areas*. Association for Supervision and Curriculum Development.
- [63] Wang, L., & Wu, J. F. (2021). *Yuliaoku fuzhu yuedu jiaoxue zai duhou xuxie zhong de yunyong [The application of corpus-assisted reading teaching in continuation writing]*. *Jiaoxue Yuekan (Zhongxueban) (Teaching Monthly (Secondary School Edition))*, (4), 19-24.
- [64] Wang, L. F., & Zhang, Y. (2007). *Daxuesheng yingyu yilunwen zhong gao pin dongci shiyong de yuliaoku yanjiu [A corpus study of high-frequency verb use in college students' English argumentative essays]*. *Waiyu Jiaoxue yu Yanjiu (Foreign Language Teaching and Research)*, (2), 110-116, 160-161.
- [65] Wang, L. F., & Zhou, D. D. (2004). *Woguo yingyu kouyu yanjiu 12 nian: Huigu yu xianzhuang [12 years of spoken English research in China: Retrospect and current situation]*. *Waiyujie (Foreign Language World)*, (6), 7-14.
- [66] Wang, Q. (2020). A corpus-based contrastive analysis of it-clefts in spoken Hong Kong English: Research from the International Corpus of English (ICE). *Australian Journal of Linguistics*, 1-27.
- [67] Warren, M. (2016). *Introduction to data-driven learning*. Routledge.
- [68] Wei, N. X. (2004). *Zhongguo xuexizhe yingyu kouyu yuliaoku chushi yanjiu [An initial study of the Chinese learner English spoken corpus]*. *Xiandai Waiyu (Modern Foreign Languages)*, (2), 140-149, 216-217.
- [69] Wen, Q. F. (2010). *Eryu xide zhongdian wenti yanjiu [Key issues in second language acquisition research]*. Foreign Language Teaching and Research Press.
- [70] Xu, M. Y. (2014). *Jiyu yuliaoku de daxue yingyu zuowen lianjie ci wu lianjie cuowu fenxi yanjiu [A corpus-based study of conjunction errors in college English compositions]* (Unpublished master's thesis). University of Electronic Science and Technology of China.
- [71] Yang, H. Z., & Wei, N. X. (2002). *Yuliaoku yuyanxue daolun [An introduction to corpus linguistics]*. Shanghai Foreign Language Education Press.
- [72] Ying, J. Q. (2021). *Yuliaoku fuzhu gaozhong yingyu yingyongwen xiezuo jiaoxue moshi tanjiu [Exploring a teaching model for corpus-assisted practical English*

- writing in high school]. *Jiaoxue Yuekan (Zhongxueban) (Teaching Monthly (Secondary School Edition))*, (4), 7-12.
- [73] Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283.
- [74] Yu, G. L. (2009). *Yuliaoku yuyanxue de yanjiu he yingyong [Research and application of corpus linguistics]*. Sichuan University Press.
- [75] Zhang, C. Q. (2022). Yuliaoku fuzhu duhou xuxie yangwen guanjianci fenxi ji dui jiaoxue de qishi [Keyword analysis of sample texts in corpus-assisted continuation writing and its implications for teaching]. *Zhongxiaoxue Shuzihua Jiaoxue (Digital Teaching in Primary and Secondary Schools)*, (6), 10-14.
- [76] Zhang, H. H. (2020). Jiyu zijian yuliaoku de gaozhongsheng yingyu xiezuo cihui dapei cuowu fenxi [An analysis of vocabulary collocation errors in high school students' English writing based on a self-built corpus]. *Haiwai Yingyu (Overseas English)*, (6), 86-87.
- [77] Zhao, J. H. (2008). Qiantan ruhe tigao yingyu xiezuo nengli [A brief discussion on how to improve English writing ability]. *Yanbian Jiaoyu Xueyuan Xuebao (Journal of Yanbian Institute of Education)*, (3), 116-118.
- [78] Zhou, W. J., & Wang, S. W. (2022). Woguo jiyu yuliaoku de yingyu kouyu jiaoxue yanjiu 20 nian zongshu (2002-2021) [A 20-year review of corpus-based spoken English teaching research in China (2002-2021)]. *Dangdai Waiyu Yanjiu (Contemporary Foreign Languages Studies)*, (4), 123-132.
- [79] Zhu, Y. (2009). Yuliaoku zai yingyu xiezuo jiaoxue zhong de yingyong [The application of corpora in English writing teaching]. *Waiyuji (Foreign Language World)*, (3), 24-31.