ISSN 2522-6398 Vol. 9, Issue 4: 119-124, DOI: 10.25236/FER.2026.090418

# An Empirical Study on the Effect of Corpus-Driven Instruction on Academic English Writing Production

**Jiahui Liu**[a,*]

*School of Foreign Languages, Ludong University, Yantai, China*
[a]*489424595@qq.com*
[*]*Corresponding author*

***Abstract:*** *With the increasingly consolidated status of English as the universal language for international academic communication, the competence of Academic English Writing has become a key focus and difficulty in higher foreign language teaching. This paper adopts an empirical research method to explore the impact of Corpus-driven Teaching on learners' production effect of Academic English Writing. The research shows that the corpus-based data-driven learning model can significantly improve the linguistic accuracy, lexical complexity and diversity, as well as textual coherence of students' writing output, while enhancing learners' autonomous learning ability and writing motivation. This study provides empirical support and practical guidance for the application of corpus technology in the teaching of Academic English Writing.*

***Keywords:*** *Corpus-driven Teaching; Academic English Writing; Production Effect; Empirical Study*

## 1. Introduction

As the core carrier of international academic communication, Academic English Writing is an important indicator to measure learners' comprehensive English application ability. It features unique characteristics in terms of vocabulary, sentence patterns and textual structure [1]. The traditional teaching of Academic English Writing mostly adopts a product-oriented approach, which focuses on the imitation of model essays and grammatical correctness but neglects the cultivation of learners' language application ability in real academic contexts. This traditional teaching model has many limitations such as single teaching materials, long feedback cycle and low student participation. With the rapid development of corpus linguistics, Corpus-driven Teaching has provided new ideas for the teaching of Academic English Writing. A corpus is a large electronic text database built by collecting naturally occurring continuous language use texts through scientific sampling methods in accordance with specific linguistic principles, which has the characteristics of authenticity, representativeness and scale, and can provide a large amount of real and natural language data for language teaching. The corpus-based data-driven learning model enables learners to observe, analyze and summarize the laws of real language use like language researchers, thus making more appropriate linguistic choices in academic writing [2]. Through teaching experiments, this study systematically explores the impact of Corpus-driven Teaching on the production effect of Academic English Writing, specifically analyzes its role in improving linguistic accuracy, lexical complexity, textual coherence and enhancing learners' autonomous learning ability, and integrates students' linguistic diversity into the classroom [3], so as to provide empirical evidence and practical reference for the reform of Academic English Writing teaching.

## 2. Theoretical Foundations of Corpus-Driven Teaching

As an emerging language teaching method, Corpus-driven Teaching is mainly based on the theories of corpus linguistics and data-driven learning.

### *2.1 Corpus Linguistics Theory*

Corpus linguistics theory provides the fundamental philosophical and methodological foundation for Corpus-driven Teaching. This theory holds that the understanding of the essence, structure and usage rules of language should be based on the systematic observation and description of large-scale

ISSN 2522-6398 Vol. 9, Issue 4: 119-124, DOI: 10.25236/FER.2026.090418

real language materials, rather than relying on linguists' intuition or prescriptive rules defined in advance, marking a profound shift in the paradigm of language research from prescriptivism to descriptivism. Traditional language teaching is often built on simplified and standardized language rules, while corpus linguistics reveals that natural language presents a high degree of variability, probability and context-dependence in practical use.

The revolutionary enlightenment of this theory for teaching is that the best source of linguistic knowledge is the empirical evidence of language's actual use. Corpus-based research can objectively reveal the typical collocation patterns of vocabulary, the actual usage frequency of grammatical structures, the stylistic features of different registers and the prevalence of multi-word units. Teaching should take the empirically tested linguistic patterns extracted from real corpora as its content, helping students master the most commonly used, typical and context-appropriate linguistic forms. This shifts teaching from imparting abstract rules to cultivating students' insight and sensitivity to real language use, providing a solid and objective empirical basis for teaching.

### *2.2 Data-Driven Learning Theory*

First clearly proposed by Tim Johns in the early 1990s, data-driven learning theory is the direct guiding principle and core operational model of Corpus-driven Teaching in classroom practice. This theory positions learners as language explorers or researchers and the corpus as an ore deposit of raw data to be explored. Instead of passively accepting teachers' explanations or textbook rules, students take the initiative to observe corpus retrieval results, raise questions, form hypotheses, verify their inferences through further corpus observation, and finally generalize the rules of language use.

This theory has completely reconstructed the relationship between teaching and learning. The teacher's role has transformed from an authoritative disseminator of knowledge to a designer of learning tasks, a guide in the exploration process and an assistant in technical use. The student's role has changed from a passive receiver of information to an active constructor of cognition. Data-driven learning not only imparts specific linguistic knowledge, but more importantly, cultivates students' metacognitive strategies and autonomous learning ability to solve linguistic problems through empirical means, enabling them to independently cope with diverse linguistic challenges in future academic activities.

## 3. Core Characteristics of Corpus-Driven Teaching

Centered on real language materials and supported by a technical cooperation mechanism, the corpus-driven teaching model provides an innovative solution to practical challenges such as disciplinary barriers and differences in language standards in the interdisciplinary teaching of Academic English Writing [4]. As a methodological innovation, the core characteristics of Corpus-driven Teaching profoundly reflect the paradigm shift of applied linguistics from prescriptivism to descriptivism and from teacher-centeredness to student-centeredness. It does not simply regard the corpus as supplementary materials, but takes it as the cornerstone of curriculum design and the main environment for students' learning.

### *3.1 The Essence of Teaching Materials is Massive Linguistic Data in Real Contexts*

Corpus-driven Teaching directly introduces unprocessed raw linguistic data from real academic communication scenarios, including journal papers, academic monographs and conference abstracts, into the classroom. These corpora form a large and objective database of linguistic facts, exposing students to the full picture of Academic English in practical use, including high-frequency vocabulary, typical collocations, common sentence structures and rhetorical conventions of specific disciplines, thus enabling students to acquire a vivid ability of linguistic generalization beyond the example sentences in textbooks.

### *3.2 The Core of the Learning Process is Data-Driven Induction and Discovery*

Under the data-driven model, students' role has transformed from passive knowledge receivers to active language researchers. Learning no longer starts with the explanation of grammatical rules, but with exploratory tasks designed by teachers. By observing the frequency information, contextual co-occurrence and collocation networks provided by the corpus, students independently put forward

ISSN 2522-6398 Vol. 9, Issue 4: 119-124, DOI: 10.25236/FER.2026.090418

hypotheses, verify patterns and generalize rules. This discovery-based learning deepens students' understanding of language, cultivates their sensitivity and critical awareness of language use, and enables them to understand the communicative motivations behind the rules.

### *3.3 The Extension of Teaching Objectives is the Cultivation of Metacognitive and Autonomous Learning Abilities*

In the repeated process of retrieval, observation, comparison and induction, students gradually master a set of methods and tools for exploring language, and learn to seek empirical evidence through the corpus when encountering linguistic questions, rather than merely relying on teachers or dictionaries. This ability enables students to continuously use corpus resources to support their future academic writing after the course, realizing lifelong learning. Essentially, it cultivates students' linguistic metacognitive ability, that is, the ability to monitor, evaluate and regulate their own language learning process, making them independent and confident language users.

## 4. Research Design of the Empirical Study on Corpus-Driven Teaching

To systematically explore the impact of Corpus-driven Teaching on the production effect of Academic English Writing, this study designs a rigorous empirical research scheme. A mixed research method combining quantitative and qualitative analysis is adopted to ensure the comprehensiveness and reliability of the research results.

### *4.1 Research Questions*

To systematically explore the impact of Corpus-driven Teaching on the production effect of Academic English Writing, this study designs a rigorous empirical research scheme. A mixed research method combining quantitative and qualitative analysis is adopted to ensure the comprehensiveness and reliability of the research results.

### *4.2 Research Subjects*

This study takes two natural classes of English majors in a university as the research subjects, with a total of 82 sophomore students. Among them, the experimental group with 41 students received Corpus-driven Teaching, while the control group with 41 students adopted the traditional writing teaching method. All students had intermediate English proficiency and were taking the Academic English Reading and Writing course. There were no significant differences between the two groups in terms of college entrance English scores and CET-4 scores, ensuring the reliability of the experiment.

### *4.3 Experimental Design*

The experiment lasted for one semester (16 weeks). In the first 7 weeks, both groups received the same Academic English reading training. In the subsequent 9 weeks, the experimental group received corpus-driven writing teaching, while the control group maintained traditional teaching. Traditional teaching included such links as model essay analysis, grammar explanation and teacher correction; Corpus-driven Teaching involved activities such as training on the use of corpus retrieval tools, corpus observation and analysis, and independent retrieval and revision.

The specific teaching process of the experimental group included three stages: the preparation stage (Weeks 1-2), where teachers introduced basic corpus concepts and retrieval methods, and students conducted preliminary exercises; the practice stage (Weeks 3-8), where students wrote abstracts for designated papers and made independent revisions using the corpus; the evaluation stage (Week 9), where teachers provided feedback on students' final results. The corpus resources used included the Corpus of Contemporary American English (COCA) and a small academic abstract corpus, the latter of which was constructed by the researcher by extracting the abstract part from the Chinese-English academic paper corpus.

### *4.4 Data Collection and Analysis*

The study collected a variety of data, including pre-test and post-test samples of students' compositions, questionnaires and interview materials. The composition samples were used to analyze

ISSN 2522-6398 Vol. 9, Issue 4: 119-124, DOI: 10.25236/FER.2026.090418

the changes in linguistic accuracy, lexical complexity and textual coherence. Linguistic accuracy was evaluated through error rate analysis, including the statistics of grammatical, lexical and syntactic errors; lexical complexity was measured by the type-token ratio and the proportion of academic vocabulary; textual coherence was analyzed using the cohesion theory framework of Halliday and Hasan [5].

A variety of tools and methods were adopted for data analysis. Wordsmith Tools 7.0 was used for text feature extraction, and SPSS 25.0 was used for statistical tests, including descriptive statistics and inferential statistics. The questionnaire and interview materials were coded and theme-extracted using content analysis to understand students' subjective perceptions of corpus teaching.

This comprehensive data collection and analysis method ensures that the research results are not only statistically significant, but also reflect students' real learning experience, thus comprehensively evaluating the effect of Corpus-driven Teaching.

## 5. Results and Discussion of the Empirical Study

Through the systematic analysis of experimental data, it is found that Corpus-driven Teaching has multiple positive impacts on the production effect of Academic English Writing. The research results are reported in detail and discussed from four dimensions as follows.

### *5.1 Improvement of Linguistic Accuracy*

Quantitative analysis shows that after receiving Corpus-driven Teaching, students in the experimental group had a significant reduction in linguistic errors in their writing. The results of the Wilcoxon signed-rank test indicated a significant difference between the number of errors in the post-test and that in the pre-test. Specifically, the error rate of number agreement decreased from 11.0% to 3.5%, the error rate of article usage from 6.9% to 2.1%, and the error rates of tense and spelling decreased by 3.5% and 3.1% respectively. This finding confirms the effectiveness of the corpus in improving linguistic accuracy.

The improvement of linguistic accuracy is mainly attributed to the immediate comparative feedback mechanism provided by the corpus. By comparing their own writing with the standard expressions in the corpus, students can independently identify and correct errors. For example, when expressing "realize one's ambition", many students initially used "achieve the ambition", but found the more idiomatic expression "fulfill one's ambition" by retrieving the COCA corpus, thus making corresponding revisions. This data-driven discovery learning can deepen students' understanding and internalization of language rules more effectively than traditional teacher error correction.

### *5.2 Enhancement of Lexical Complexity and Diversity*

Lexical complexity analysis showed a significant increase in lexical variation in the compositions of the experimental group. The type-token ratio of the post-test compositions rose from 45.2 to 52.7, the standardized type-token ratio from 38.5 to 44.3, the average word length from 4.5 to 5.2, and the average sentence length from 13.4 words to 15.8 words. These data indicate that Corpus-driven Teaching helps students expand their lexical resources and improve the richness and accuracy of language expression.

Notably, students not only increased their vocabulary, but also learned more academic vocabulary and professional expressions. By observing high-frequency vocabulary and collocations in the academic corpus, students consciously used more formal academic terms in their writing. For example, when expressing "an important reason", students began to use "underlying factor" instead of the simple "important reason"; when expressing "more and more", they used "increasingly" instead of "more and more". This improvement at the lexical level significantly enhanced the academic nature of writing.

### *5.3 Improvement of Textual Coherence*

Textual analysis showed that students in the experimental group made remarkable progress in the use of cohesive devices. After the teaching, the frequency of logical connectives used in students' compositions increased from 12.4 times per 1000 words to 18.7 times per 1000 words, anaphoric cohesion from 8.6 times per 1000 words to 11.2 times per 1000 words, lexical repetition from 15.3

ISSN 2522-6398 Vol. 9, Issue 4: 119-124, DOI: 10.25236/FER.2026.090418

times per 1000 words to 21.5 times per 1000 words, and the frequency of synonym substitution had the largest increase, rising from 5.2 times per 1000 words to 9.8 times per 1000 words.

Further analysis found that students not only increased the frequency of using cohesive devices, but also improved the appropriateness and diversity of their use. In the expression of logical relations, students learned to use a more diverse range of connectives besides "but" and "so", and could express relations such as transition, cause and effect, and progression more accurately. In terms of anaphoric cohesion, students could use pronouns and demonstratives more appropriately to avoid repetition and ambiguity. In lexical cohesion, students learned to achieve semantic coherence through synonyms, hyponyms and other means, making the theme of the article more prominent and the logic clearer.

### *5.4 Changes in Learners' Attitudes and Motivation*

The results of questionnaires and interviews showed that the vast majority of students held a positive attitude towards Corpus-driven Teaching. 98% of the students in the experimental group believed that this teaching method had aroused their strong interest in English writing and enhanced their writing confidence. Students generally reflected that the sense of discovery and accomplishment brought by corpus retrieval was incomparable with traditional teaching. One student stated in the interview: "Discovering language rules through my own exploration is more interesting than passively accepting teachers' explanations, and the memory is also more profound."

The improvement of learning motivation is also reflected in students' extracurricular learning time. Students in the experimental group spent an average of 2.5 more hours per week on autonomous learning of English writing than those in the control group, and were more willing to revise their compositions repeatedly and try new expression methods. The stimulation of this intrinsic motivation provides impetus for continuous learning and is also an important factor in the improvement of writing ability.

It is worth noting that the effect of Corpus-driven Teaching presents certain dynamic change characteristics. The study found that when teachers pointed out the insufficient use of a certain vocabulary in classroom feedback, students would significantly increase the frequency of using that vocabulary in the next assignment, but the frequency would gradually drop to a reasonable level over time. This phenomenon of "over-correction-balance" reflects the dynamic development process of students' language system, indicating that the effect of Corpus-driven Teaching is not a simple linear growth, but a process of dynamic adjustment.

## 6. Pedagogical Implications and Challenges

The Academic English Writing course is the main position for cultivating students' Academic English Writing competence [6]. Based on the results of this empirical study, Corpus-driven Teaching shows significant advantages in the teaching of Academic English Writing, while also facing some challenges.

### *6.1 Pedagogical Implications*

The application of Corpus-driven Teaching should focus on hierarchical implementation and gradual progression. For beginners, small and targeted corpora, such as the academic abstract corpus, can be provided first to reduce technical barriers and cognitive load. With the improvement of ability, large general corpora such as COCA and BNC can be gradually introduced to expand the scope and diversity of language input. In the teaching process, teachers should reasonably design the difficulty of tasks to ensure that corpus retrieval tasks are in line with students' "zone of proximal development" and achieve comprehensible input in the form of i+1.

The successful implementation of Corpus-driven Teaching requires the transformation of teachers' roles. Teachers should shift from knowledge disseminators to learning guides and resource providers. Specifically, teachers need to screen appropriate corpus resources in advance, design targeted retrieval tasks, guide students to interpret corpus data, and help students transform retrieval results into writing practice. In addition, teachers should provide formative assessment, focusing on students' exploration process and reflective ability, rather than only valuing writing results.

ISSN 2522-6398 Vol. 9, Issue 4: 119-124, DOI: 10.25236/FER.2026.090418

### *6.2 Challenges*

Despite its obvious advantages, Corpus-driven Teaching still faces multiple challenges in practice. First is the problem of technical barriers. The use of corpus tools has certain technical requirements for both teachers and students, which may cause technology anxiety. The solution is to develop a more user-friendly interface and provide detailed usage tutorials and technical support. Second is the insufficient teacher development. Many teachers lack professional training in corpus linguistics and find it difficult to effectively integrate the corpus into teaching. It is suggested to strengthen teacher training, share successful teaching cases and establish a teacher learning community. Third is the limitation of teaching resources. The construction of high-quality specialized corpora requires a lot of time and energy. It is suggested that universities cooperate to build a resource sharing platform to improve the efficiency of resource utilization.

## 7. Conclusions

The research shows that compared with the traditional teaching method, Corpus-driven Teaching can significantly improve learners' linguistic accuracy, lexical complexity and diversity, as well as textual coherence. These achievements are mainly attributed to the real language input provided by the corpus, data-driven discovery learning and opportunities for independent exploration, which jointly promote learners' understanding and application of the characteristics of Academic English Writing. Educators should fully recognize the teaching value of the corpus, systematically integrate it into curriculum design, and maximize its teaching benefits through scientific and reasonable task design and technical support. With the continuous development of corpus technology and the deepening of teaching practice, Corpus-driven Teaching is expected to play a greater role in the field of Academic English Writing, promoting the development of foreign language teaching towards a more scientific, personalized and efficient direction.

## References

*[1] H. Wang. A Corpus-driven Approach to Teaching Academic Writing for College Students: Taking Abstract Writing as an Example[J]. Foreign Language Research, 2020, 43(01): 49-55.*
*[2] Vasilopoulos E. Transformative plurilingualism pedagogies in English academic writing: instructor perceptions[J]. Language and Education, 2026, 40(2): 449-469.*
*[3] Wang C. A Discussion on Chinese Non-English Majors Use Corpus to Improve Academic Words Use in Academic Writing[J]. Scientific Journal Of Humanities and Social Sciences, 2025, 7(6): 172-183.*
*[4] Xinyi S. Corpus-driven: Research on the Construction of Interdisciplinary Teaching Model of Academic English Writing[J]. Frontiers in Educational Research, 2025, 8(11): 110-114.*
*[5] F. Bai, L. T. Lai. A Review of Genre Theory in Functional Linguistics[J]. Journal of Jiangxi Normal University(Philosophy and Social Sciences Edition), 2021, 54(02): 140-144.*
*[6] K. S. Lai, L. J. Mo, S. Lu, et al. Application of Flipped Classroom Teaching Model in Academic English Writing Based on Corpus[J]. Journal of Pingxiang University, 2025, 42(01): 113-116.*