Education Reform and Development, 2025, Volume 7, Issue 4
https://ojs.bbwpublisher.com/index.php/ERD
Online ISSN: 2652-5372
Print ISSN: 2652-5364

Research on Strategies for Improving College Students' English Writing Skills Based on Internet Online Corpora

Hang Li*

Chongqing Jiaotong University, Chongqing 400007, China

**Author to whom correspondence should be addressed.*

Abstract: This paper studies strategies for improving college students' English writing literacy based on online corpora on the Internet. Factors related to learners of English writing corpora include corpus sources, years of English learning, educational level, and major. Task factors include text types, writing time limits, and the use of reference books. The corpus consists of three subcorpora: General English, Business English, and Academic English, with a storage capacity of 3.256 million words. Its main uses include setting writing standards for science and engineering college students, constructing autonomous learning platforms, studying the characteristics of interlanguage phrases, conducting diachronic studies of interlanguage, and translation studies.

Keywords: Corpus; College students; English writing; Interlanguage

Online publication: April 28, 2025

1. Development background

The "College English Teaching Syllabus" (2020 Edition) (hereinafter referred to as the "Syllabus") points out that the goal of college English teaching is to cultivate students' English application abilities, and it proposes three levels of teaching objectives from the five aspects of English listening, speaking, reading, writing, and translation. The third-level goal for "written expression ability" requires: students can appropriately use writing skills. They can express their personal views in written English more freely; they can write expository and argumentative texts on a wide range of social and cultural topics with a certain depth of thought, clear expression, rich content, clear structure, and strong logic [1].

The "Syllabus" also points out that in the era of informatization and intelligence, multimedia technology and big data, virtual reality and artificial intelligence technology have become important means of foreign language education teaching. How to improve students' English writing ability has always been a focus in domestic English teaching, and the emergence of online corpora has provided a new path for English writing, especially in the use of vocabulary and the appropriateness and diversity of sentence expression [2].

A learner corpus refers to "an electronic collection of real learner language texts collected according to certain design specifications to improve foreign language teaching." Over the past 30 years, corpus linguistics, with a large amount of real texts and powerful retrieval and statistical tools, has opened up new ways for language description and analysis [3]. Since the early 1990s, the rapid construction and in-depth development of learner corpora have greatly promoted research on learner language output based on corpora and have continuously changed the face of traditional second language acquisition and foreign language teaching research.

Due to cultural differences between China and the West and the negative transfer of the mother tongue, college students generally find it "very difficult" to write a decent English composition. In the college English four and six-level exams, the writing part scores are far below the passing line, which shows that the current English writing level of college students is low, and the writing situation is not optimistic [4]. The reasons are mainly that teachers and students do not pay enough attention to English writing courses, and there is too much listening and reading training in domestic college English teaching, with insufficient class hours for English writing instruction. Secondly, English writing teaching methods are outdated, and many teachers do not use the vast teaching resources on the Internet platform, English corpora, and other digital age teaching tools. Finally, students lack interest in English writing; English is a foreign language for college students, and they have not established a good English thinking habit and learning interest, and they lack autonomous learning ability.

2. The role of English writing corpora

The study has independently constructed standard model text corpora and student composition corpora using online corpus resources on the Internet. Through four aspects of vocabulary, vocabulary precision, phrase combination, and syntactic construction, the study has deeply analyzed the characteristics of college students' English writing [5]. At the same time, the study has explored how to use corpora to compare the differences between student works and standard models, and accordingly carried out multi-angle quantitative analysis and qualitative evaluation of college students' English compositions. Based on this, the study has summarized the common problems in college students' English writing and put forward targeted suggestions for English writing teaching to improve students' writing ability.

2.1. Corpus structural design

In building the corpus, the study follows the design criteria aimed at ensuring the accuracy and credibility of subsequent research, with the ultimate vision of "helping teachers optimize writing teaching and improve students' writing skills" [6]. The design of the corpus focuses on three core elements. The first is the determination of the corpus style, which guides the subsequent corpus collection and structural design, and the study clearly positions the corpus as a collection of written texts. Second, the text selection process of the corpus is crucial, and the study carefully retrieves and organizes online resources on the Internet to select the latest, high-quality, and widely representative text materials to ensure the timeliness, comparability, and integration of the corpus. Finally, the data comparison and analysis functions of the corpus are indispensable, and the study deeply analyze the actual needs of teachers and students in English writing teaching, and accordingly clarify the type, scale, structure, and content of the corpus, and successfully establish a small-scale student composition corpus and standard model text corpus suitable for teaching, providing strong support for teaching [7].

2.2. Corpus data collection and organization

College students' real written texts constitute the data foundation of the student composition corpus, which includes at least 1000 works. The standard model text corpus widely collects a variety of resources such as excellent English writing at home and abroad, college English four and six-level exam full-score models, and original English writing materials published on the network platform, as a reference benchmark, and its sample size naturally exceeds the student composition corpus [8]. The total sample size of the standard corpus integrated on the Internet platform is sufficient to meet this demand.

In the process of building the corpus, the study has adopted a variety of collection methods, such as online retrieval, voice entry, and scanning recognition technology. After collection, the study has carefully organized and preprocessed the texts, including correcting non-standard punctuation marks and adjusting paragraph formats, to ensure the accuracy and timeliness of subsequent vocabulary analysis, collocation statistics, and retrieval results [9].

2.3. Corpus data statistics and analysis

The study uses corpus resources to compare the student composition corpus with the standard model text corpus, and deeply analyze the current situation and characteristics of college students' English writing from four aspects: vocabulary, vocabulary precision, phrase usage, and syntactic construction. The specific analysis includes the following aspects:

(1) The similarities and differences in the number of high-frequency words and their usage frequency between the two corpora;

(2) The sensitivity and precision of students' vocabulary selection in similar contexts;

(3) Preferences and characteristics of phrase combinations in English writing practice, as well as the frequency of idiomatic usage;

(4) The frequency and accuracy of complex syntactic structures in English writing;

(5) An exploration of the correlation between the above four dimensions of data and the overall quality of student compositions.

3. Design principles

3.1. Overall planning

In foreign language teaching research, the specific effectiveness of learner corpora is deeply influenced by a series of controllable variables [10]. These variables can be roughly divided into two categories: one is related to the learners themselves, covering learning environment, mother tongue background, and foreign language proficiency; the other is related to task execution, including task time limits, execution environment, and available reference materials.

3.2. Text types

The distinction of text types fundamentally answers the purpose of learning a language. As mentioned by previous studies, general English forms the foundation, and English for specific purposes (covering business English and academic English) is "an important direction for promoting our country's college English teaching to a higher level" [11]. In addition, these three types of texts have significant differences in communication intentions, theme content, discourse structure, vocabulary, and grammar [12].

4. Application Prospects

To meet this challenge, the study will select examples from domestic and foreign English teaching materials based on preset occupational scenarios to construct a model composition corpus, to assess the degree of fit or deviation of student compositions from model standards [13]. For tasks such as company and product introductions, the study will integrate text information from the official websites of well-known European and American companies with the model composition corpus to form an EOP reference corpus, that is, a business English writing corpus.

5. Conclusion

Collecting real writing texts from college students and online English resources to expand and establish a college English writing corpus can not only provide rich materials for college students' writing but also provide practical reference value for English writing teaching corpus research. At the same time, the capacity of the corpus can be increased with the continuous use in writing teaching, and its functions can be continuously improved based on the feedback from teachers and students, enhancing the quality and timeliness of the data in the corpus [14]. The expansion and application of the online corpus is the biggest innovation of this project.

Combining the actual situation of college English writing teaching, using the comparative research method of the corpus, combining teaching content with corpus tools, understanding the problems that students are prone to in writing through data collection and analysis, and improving teaching methods and teaching effects [15]. Students can cultivate autonomous learning ability and improve their writing literacy through comparative learning with the corpus.

In summary, the establishment and development of the English writing corpus for Chinese science and engineering students are of great significance for an in-depth understanding of the English writing ability of Chinese science and engineering college students. This corpus is not only a valuable resource but also is expected to play a core role in the field of educational technology.

Funding

2024 annual project of the 14th Five-Year Plan for Education Science in Chongqing Municipality, "Research on Improving College Students' English Writing Literacy Based on Internet Online Corpus" (Project No.: 105); The 12th China Foreign Language Education Fund project, "Construction and Research of a Multidimensional Evaluation Database for College Students' English Reading and Writing Skills Supported by Information Technology" (Project No.: ZGWYJYJJ12A060); Education and Teaching Reform Project of Chongqing Foreign Language Society, "Construction and Research of a Multidimensional Evaluation Corpus for College Students' English Reading and Writing Skills in the Context of 'Internet+'" (Project No.: 06)

Disclosure statement

The author declares no conflict of interest.

References

[1] Al-Kharabsheh A, Hamadeh N, 2017, Shifts of Cohesion and Coherence in the Translation of Political Speeches.

Advances in Language and Literary Studies, 5: 26–28.
[2] Behnam B, Yaghchi MA, 2019, The Impact of Formal Instruction of References and Conjunctions on Reading Comprehension of Iranian ESP Students. Procedia – Social and Behavioral Sciences, 9: 62–66.
[3] Boas F, 1940, Race, Language and Culture. Macmillan, 1940: 1–237.
[4] Chanyoo N, 2016, A Corpus-Based Study of Connectors and Thematic Progression in the Academic Writing of Thai EFL Students. ProQuest LLC, 15: 33–36.
[5] Manning CD, Schütze H, 1999, Foundations of Statistical Natural Language Processing. MIT Press, USA.
[6] Aston G, Burnard L, 1998, The BNC Handbook. Edinburgh University Press, 1998: 1–268.
[7] Atkins S, Clear J, Ostler N, 1992, Corpus Design Criteria. Literary and Linguistic Computing, 11: 102–106.
[8] Leech G, 1992, Computers and Corpus Analysis. Computers and Written Texts, 1992: 1–246.
[9] Scott M, 2008, WordSmith Tools Version 5. Lexical Analysis Software, 2008: 1–276.
[10] Mohammed, Sadiya A, 2015, Conjunctions as Cohesive Devices in the Writings of English as Second Language Learners. Procedia – Social and Behavioral Sciences, 5: 22–26.
[11] Petersen U, 2004, Emdros – A Text Database Engine for Analyzed or Annotated Text. International Conference on Computational Linguistics, 2004: 1–253.
[12] Ravid D, Berman RA, 2010, Developing Noun Phrase Complexity at School Age: A Text-embedded Cross-linguistic Analysis. First Language, 10: 98–101.
[13] Read J, 2000, Assessing Vocabulary. Cambridge University Press, London.
[14] Richards B, 1987, Type/Token Ratios: What Do They Really Tell Us? Journal of Child Language, 11: 88–90.
[15] Vygotsky LS, 1978, Mind and Society: The Development of Higher Psychological Processes. Harvard University Press, Cambridge.