

# An FCE Corpus-Based Analysis of Writing Errors in Chinese English Learners: Common Types, Patterns, and Pedagogical Implications

Yuanyuan Wu<sup>1</sup>, Xinping Jiao<sup>2</sup>

<sup>1,2</sup> National University of Defense Technology, China  
e-mail: <sup>1</sup>wuyuanyuan24@nudt.edu.cn, <sup>2</sup>jiao\_1@yeah.net

Received 30 March 2025 | Received in revised form 20 April 2025 | Accepted 23 April 2025

APA Citation:

Wu, Y. & Jiao, X. (2025). An FCE Corpus-Based Analysis of Writing Errors in Chinese English Learners: Common Types, Patterns, and Pedagogical Implications. *Journal of English Language Teaching and Linguistics*, 10(1), 2025, 141-152. doi: <http://dx.doi.org/10.21462/jeltl.v10i1.1527>

## Abstract

*L2 Writing ability is one of the important criteria for measuring the comprehensive ability of English learners. Researchers have gradually deepened their research on writing ability from the study of common error patterns of second language learners. In recent years, it has also shifted to the use of natural language processing technology to achieve automatic error recognition of written texts. However, there are still few studies that conduct systematic and real investigations on the writing errors of specific second language learners, especially Chinese English learners. Therefore, this paper uses the Cambridge Learner Corpus First Certificate in English to study the test scripts produced by Chinese English learners in a real test environment. The study found that Chinese English learners are greatly affected by the negative transfer of their mother tongue, and often make mistakes in tense, subject-verb inconsistency, punctuation, articles, and prepositions. This provides real data presentation for the existing common errors in writing of Chinese English learners. Through the analysis of common error patterns, this study proposes four teaching strategies: establishing a joint classroom of English and Chinese, intensive training on common errors, cultivating awareness of mother tongue interference, and providing personalized feedback, which provides possible research directions and inspirations for existing writing teaching.*

*Keywords: FCE corpus, L2 writing, second language teaching, writing errors*

---

## 1. Introduction

Writing ability is one of the important assessment criteria for English learners in China. It includes comprehensive requirements for learners' grammar mastery, vocabulary mastery, language organization ability, expression logic, etc. Writing ability is not just a simple stacking of words, grammar, and syntax, but also requires learners to use language flexibly to convey their views. Given this, Chinese teachers and teaching syllabuses focus on the cultivation of learners' writing ability. In addition, many English level tests, such as international tests like IELTS, TOEFL, Cambridge level tests, and Chinese local tests like CET-4 and CET-6, all include the assessment of writing ability, and the writing part usually occupies a large proportion. This not only reflects the important position of writing ability in the process of English learning, but also emphasizes the irreplaceable role of writing as a comprehensive language skill in real communication and academic scenarios.

In the process of writing teaching and learning, learners' errors are a key issue that cannot be ignored. Errors in the writing can help teachers and learners identify specific problems in learning languages. By correcting these problems, learners can improve their writing skills accordingly. Since Coder (1967) proposed the concept of "error analysis", different researchers have conducted in-depth research on common errors in the second language acquisition process. Richards (1970) found that Interlingual Errors, Overgeneralization Errors, and Errors Induced by Teaching are the main sources of learners' writing errors. Dulay et al. (1982) developed *Creative Construction Error* based on the first two types of errors and provided a more detailed error classification, including omission, addition, mis selection, and mis ordering. They also used the methods of error analysis and comparative analysis hypothesis to explore the specific causes of errors. Therefore, the study of learners' writing errors has practical significance. On the one hand, teachers can deeply understand the common problems of students in the second language acquisition process and the root causes of their occurrence through the results of error analysis, so as to formulate more scientific and effective teaching strategies. On the other hand, learners can obtain feedback from the results of error analysis, thereby identifying their weaknesses in the learning process and then making targeted improvements.

As research deepens, error analysis methods are constantly combined with new technologies to improve the validity and efficiency. Among them, corpora provide researchers with data-driven systematic analysis methods by collecting and organizing large-scale real corpora as data. Biber et al. (1994) explained that corpora have a model for empirical analysis of actual usage patterns in English grammar, and showed that the early conclusions based on intuition were insufficient or even incorrect.

Despite the rich research on second language writing and the continuous progress of automated error detection, there are still few studies that use real test data to systematically examine the writing errors of specific learner groups. This also means that Chinese English learners are still underrepresented in corpus-based research. This study uses real test scripts to reveal the main error patterns of Chinese English learners and provide targeted teaching strategies to address this gap.

This paper selects Cambridge Learner Corpus First Certificate in English (CLC FCE) as a research carrier to explore the characteristics of writing errors of Chinese learners in the process of English learning. FCE has rich language data that can truly reflect learners' language performance. The FCE corpus mainly represents learners at the B2 level of English. Learners at this stage are prone to make some mistakes, but they can also reflect a certain level of language.

The FCE corpus also uses strict error annotation to finely classify learners' errors, including grammar, vocabulary, spelling, syntactic structure, and problems. Using the annotated data of these counties, researchers can quickly locate and extract target errors and improve the accuracy of analysis. However, the current analysis based on the FCE corpus focuses on improving the algorithm to enhance the ability of automatic error annotation. Yannakoudakis, et al. (2011) used supervised machine learning technology to train a scoring model using the FCE corpus as a training set. Existing research

based on the FCE corpus focuses on using the FCE corpus to build models to improve the accuracy of error recognition (Stahlberg & Kumar, 2020), developing algorithms to identify vocabulary or grammatical errors (Zhao, et al., 2019, Mesham, et al., 2023), and researching frameworks to improve the accuracy of error marking (Rei, 2017). Second, there are few literatures that conduct in-depth analysis of the error patterns of English learning writing by specific native English learners. Therefore, this paper uses the FCE corpus to explore the error types and patterns of specific Chinese English learners in writing, and hopes to provide some inspiration for existing Chinese English teaching methods.

## **2. Literature Review**

### **2.1 Corpus-Based Studies in Second Language Writing**

As a systematic collection of real language data, corpus is an important tool for analyzing language use. Researchers often apply it to the analysis of writing errors of second language (L2) learners in order to improve learners' writing ability. Granger (2004) highly praised the value of corpora in terms of representation, arguing that research should spare no effort in building and analyzing learner corpora to help verify existing hypotheses or provide support for the formulation of new hypotheses. Further, Yoon and Hirvela (2004) showed that students have a positive attitude towards corpora in second language writing, particularly for mastering word usage and improving writing confidence. Yoon's (2008) case study showed that corpora can not only solve students' immediate writing and language problems, but also improve their vocabulary and grammar awareness and language sensitivity. Students' sense of responsibility, independence, and self-confidence in writing have also been enhanced.

Through corpus analysis of the use of epistemic modality by native and non-native authors, Chen (2010) found that the textbook design at the time was inadequate, and the significantly low use rate of certain epistemic modal means in writing by non-native authors needs to be paid attention to. Tono, et al. (2014) explored the impact of corpus on different types of error correction and found that it was effective for correction of omission and addition errors, but limited for misconstruction errors. Satake (2020) reinforced that corpus is particularly suitable for correcting omission errors, and the target phrases and co-occurrence frequency information can help learners summarize patterns and make accurate corrections. More recently, Lan et al. (2022) observed notable differences in the use of noun phrases between L1 and L2 students, while Ueno and Takeuchi found that corpora have significant effects on second language learning in the short term, but the long-term effects may be limited. Taken together, all findings mentioned above suggest that corpora not only benefit for resolving immediate language problems but also foster long-term writing development (Gilquin, 2024).

### **2.1 Theoretical Basis**

Dagneaux, et al. (1998) mentioned that computer-assisted error analysis can make up for the limitations of traditional error analysis (EA) in terms of high subjectivity and lack of dynamic learning process analysis. Using the methods and tools of corpus linguistics, learners can view error instances in context and compare them with non-error forms, better reflecting the actual needs of learners and thus improving the efficiency of teaching tools. Although error annotation is time-consuming and requires a strict error labeling system, its rewards are also huge (Granger, 2003). Corpora containing error annotations can provide researchers with detailed error statistics and even automated error analysis. Corpus analysis of specific error types can help researchers view errors in context to more reliably describe learners' interlanguage and provide effective suggestions for their learning process.

Zheng and Park (2013) used error analysis, comparative analysis hypothesis, and creative construction hypothesis (CCH) to analyze the errors in the 168 essays they took. Error analysis helps understand the source of errors. CAH emphasizes the role of the mother tongue in second language acquisition and mainly explores the negative impact of mother tongue transfer. CCH shows that even without mother tongue interference, learners may make creative errors, that is, such errors do not exist in the mother tongue. Learners have innovative constructions in the process of second language acquisition. Among them, error analysis and contrastive analysis are also the main analysis methods in

Dulay, Burt, and Krashen (1982). Mohamed, et al. (2004) mainly used error analysis to classify errors and identify the causes of errors. In addition to using the three common methods, Eng et al. (2020) also used the Interlanguage Theory to compare the error types and probabilities of the two groups of students.

### **2.3 Chinese English Learners' Writing Errors**

In a study analyzing the errors made by Chinese learners in English writing, Mohamed, et al. (2004) found that the errors mainly occurred in verbs, prepositions, and spelling. The reason for the errors was overgeneralization and simplification, which is one of the manifestations of students' weak grasp of grammatical rules. Studies found that common errors in Chinese students' English writing include misuse of conjunctions, errors in sentence connection (use of symbols), omission of articles and plural forms, errors in sentence order, and incorrect forms of verbs and nouns (Zhang & Park, 2013; Zhang et al., 2022). This is consistent with the errors made by Dulay, Burt, and Krashen (1982) for global second language learners, namely language transfer, overgeneralization, and creative construction errors. Eng et al. (2020) classified the errors made by Chinese English learners in IELTS writing in a more detailed way. Common errors include: omission, misuse, misformation, repetition, structure errors, vocabulary errors, errors of cohesion/coherence, errors of expression, misordering, and sentence errors.

### **2.4 Pedagogical Implications of Error Analysis**

Granger's series of studies (2003, 2004, 2015) collectively highlight the pivotal role of error-annotated learner corpora in second language acquisition and teaching. She emphasized that the results of corpus research need to be incorporated into the design of teaching syllabi and teaching materials. For error annotation, researchers need to conduct more sophisticated processing to bridge the gap between second language acquisition and English teaching. In addition, learner corpora can play a crucial role in foreign language teaching in three key aspects: selection, description, and sequencing. By identifying language forms that learners do not use frequently or overuse, the selection of teaching content can either strengthen or weaken certain aspects of language learning. The analysis of correct and incorrect usage patterns can help compile teaching materials to provide more refined and higher-quality language descriptions. Granger's work thus bridged the gap between second language research and pedagogy, advocating for a more data-driven and learner-centered approach to writing instruction.

Heydari and Bagheri (2012) believe that errors in the learning process can help teachers locate learners' progress in target language learning, help teachers understand learners' difficulties at different teaching stages, and thus inspire teachers to adopt more effective teaching strategies. In addition, corpus data can provide a reference for the sequence of teaching content to ensure that it matches the learners' development stage. Therefore, by improving the selection, description and sequencing of teaching materials, we can provide learners with corresponding language needs, thereby improving the effectiveness and quality of language teaching. Yoon (2008) believes that corpora have significant teaching potential in academic English writing. Although they are not the best method for all learners, they are important language resources to help solve writing problems and are also important tools to promote learners' independent writing. If properly planned, corpus learning can enhance writing experience and confidence, thereby promoting the development of overall writing ability.

Corpora are potential auxiliary tools, but their wide application value in writing teaching needs further research (Tono et al., 2014). Lan et al. (2022) suggest providing L2 students with more opportunities to expand noun phrases and teaching compressed noun phrases in graduate writing courses. Lan et al. not only provide inspiration for customized grammar teaching, but also recommend the use of data-driven learning and online corpus resources to support teaching. Gilquin (2024) suggests that teaching applications should be explicitly considered when designing corpora so that learners can benefit from bridging practical data and teaching methods.

### **3. Research Method**

#### **3.1 Corpus**

The First Certificate in English (FCE) corpus is part of the Cambridge Learner Corpus, which is dedicated to studying language errors in English as a second language. FCE is an exam suitable for a wide range of English learners, corresponding to the intermediate B2 level. English learners at this stage have a certain level of English writing ability and will also make some common writing errors, which is suitable for analyzing errors in writing. In addition, the content of this corpus is scored by professional graders according to strict scoring criteria, and the data is highly accurate and representative.

Additionally, the FCE corpus includes accurate error annotations, which are based on a rigorous classification system. The specific error annotation letters are presented in the Appendix. This study selects 66 exam scripts produced by Chinese learners from the FCE corpus, based on the availability of clearly marked error annotations, consistent scoring criteria, and topic comparability across scripts. This corpus contains 1,390 sentences, 3,449 marked errors, and 8,315 tokens.

#### **3.2 Error Analysis & Contrastive Analysis Hypothesis**

This paper mainly adopts the analysis method of Error Analysis and Contrastive Analysis Hypothesis to deeply explore the common types of errors and their causes in Chinese English learners' writing, and put forward suggestions for improvement in English writing teaching. Error analysis helps to understand the source of errors (Zheng & Park, 2013), and classifies errors to identify the causes of errors (Mohamed, et al, 2004). Error analysis not only lists learners' errors, but also provides an analytical framework for further exploring the reasons behind the errors, helping to reveal the difficulties and obstacles in second language writing, and then providing targeted guidance for language teaching. In this paper, the error analysis method is applied to the writing samples of Chinese learners in the FCE corpus, focusing on presenting the types of errors that these learners often make in English writing. Through the analysis of error types, this paper reveals the difficulties that students generally face in English writing, such as grammatical problems such as tense use, articles, and subject-verb agreement, thereby providing data support and theoretical basis for the improvement of teaching methods.

CAH emphasizes the role of the mother tongue in second language acquisition (Zheng & Park, 2013), and anchors the cause of errors in the transfer of the mother tongue. This hypothesis holds that the differences between the learner's mother tongue and the target language will directly lead to errors, and the interference of the mother tongue will be reflected in the learner's second language output. In this study, CAH was used to analyze the types of errors in Chinese learners' English writing, especially the differences between the mother tongue Chinese structure and the English structure, such as word order, verb form, article use, punctuation use, etc., to help further understand why Chinese learners make more frequent mistakes in English writing, and the specific impact of mother tongue transfer on their writing ability.

The data analysis process included several steps: 1) using Python to extract and count error tags; 2) categorizing error types based on linguistic features; and 3) conducting a cross-linguistic comparison to interpret errors within the CAH framework. Each error was analyzed in its sentence context to identify patterns and sources of writing error.

By combining error analysis and contrastive analysis hypotheses, with the help of Python, this paper not only presents the common error types of Chinese English learners in the FCE corpus, but also deeply analyzes the specific manifestations of these errors, and combines the theory of native language transfer to explore the root causes of these errors. Finally, based on the research results, this paper puts forward suggestions for improving English writing teaching, emphasizing that more attention should be paid to the contrastive learning of the mother tongue and the target language in teaching, clarifying the similarities and differences in the use of grammar in different languages, and suggesting that teachers strengthen targeted training and correction in corresponding aspects, so as to help learners improve their English writing ability.

## 4. Result and Discussion

### 4.1 Common Error Types

The results of the keyword search and quantity calculation for the marked errors are shown in Figure 1. In general, the most common error made by all Chinese English learners in this sample in writing is RP (punctuation needs Replacing), which appears 339 times in total. Other common errors include TV (wrong Tense of Verb), 279 times, S (Spelling error), 255 times, RV (Verb needs Replacing), 223 times, RT (Preposition needs Replacing), 214 times, and MD (Determiner Missing), 209 times. These findings are in line with those of Mohamed et al. (2004), who also found that verbs, prepositions, and spelling were the most problematic areas for Chinese learners of English.

Among all the errors related to word classes, errors related to verbs were the most numerous, with 844 errors, accounting for 44.4%. This type of error was mainly concentrated in tense and verb agreement. Tense errors occurred 278 times, and verb agreement errors occurred 91 times. Article-related errors (370 times) and preposition-related errors (356 times) followed closely, showing a high frequency of errors in these two grammatical categories. There were 330 errors related to nouns. This emphasis on verb-related errors corroborates the findings of Zhang and Park (2013), who highlighted that the incorrect use of verb forms and nouns was prevalent in Chinese learners' writing.

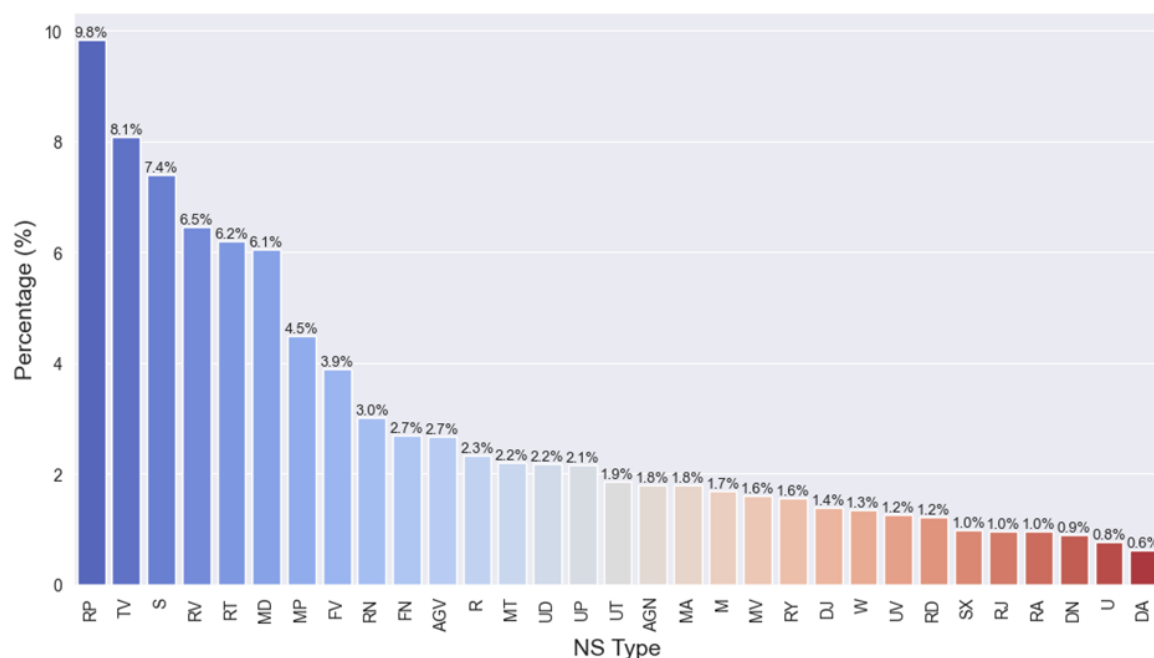


Figure 1. Proportional Breakdown of the Top 90% Error Types in Chinese English Learners' Writing

### 4.2 Specific Error Patterns

Learners' incorrect use of verbs is mainly reflected in tense and inconsistency, as shown in Table 1. Modal verbs are prone to tense errors, and learners often fail to correctly match modal verbs with the tense of the main clause. Learners also make mistakes in the tense of the be verb. Chinese verbs do not have complex tense changes like English. Verbs are usually not affected by tense, aspect, and voice. Therefore, learners may ignore the strict coordination rules between tense and modal verbs in English. In terms of verb inconsistency, the be verb is used most incorrectly. When the subject is plural, learners may use the singular be verb. This type of error is related to the fact that Chinese does not require subject-verb agreement. In Chinese, the form of the verb does not change whether the subject is singular or plural, like “我去” (I go), “我们去” (We go), “他去” (He goes), “她去” (She goes).

**Table 1.** Frequency of Verb Tense & Agreement Error Patterns

<b>Pattern</b>		<b>Count</b>	
Original	Corrected	Specific	Total
<b>Tense</b>			
is	was	12	
will	would	9	
could	can	9	52
have	had	8	
would	will	7	
are	were	7	
<b>Agreement</b>			
are	is	13	
is	are	12	41
was	were	10	
have	has	6	

Most of the errors in punctuation are shown in Table 2. Missing commas is the most common error, which occurred 97 times. The incorrect use of commas occurred 78 times, with commas being used where periods should be. It is worth noting that the incorrect use of commas is very common before pronouns such as *it, they, we, you, etc.* The redundant commas occurred 41 times. Some sentences did not end with periods, which occurred 30 times, and almost half of them occurred before *because*. In addition, some learners would omit the period. There were also cases where question marks were not used correctly, which occurred 9 times in this database. Chinese learners tend to ignore the use of commas in English writing because commas are used more frequently in Chinese and appear more in loose grammatical structures. In English, the use of commas is relatively more standardized and the scope of use is more limited than in Chinese.

**Table 2.** Frequency of Punctuation Error Patterns

<b>Pattern</b>	<b>Count</b>	
	Specific	Total
1. <b>adding ","</b>		97
	replacing	33
	. It	12
2. <b>"," to "."</b>	. They	9
	. We	9
	. You	8
	. The	7
3. <b>removing ","</b>		41
4. <b> "." to ","</b>	replacing	16
	, because	14
5. <b>adding "."</b>		15
6. <b> "." to "?"</b>		9
	<b>Sum</b>	<b>270</b>

Regarding spelling errors, "advertisement" was misspelled as "advertisment". This may be due to learners' weak grasp of spelling rules, resulting in the misspelling of words with similar syllables. When they wanted to write "then", they wrote "than". They wrote "though" instead of "thoug", and so on. These are homophone errors, and learners spell words based on pronunciation. More spelling errors can be seen in Figure 2. This is in line with De Wilde (2023), who found that young L2 English learners often rely on phonological representations when writing, leading to misspellings and vocabulary simplification in narrative tasks.

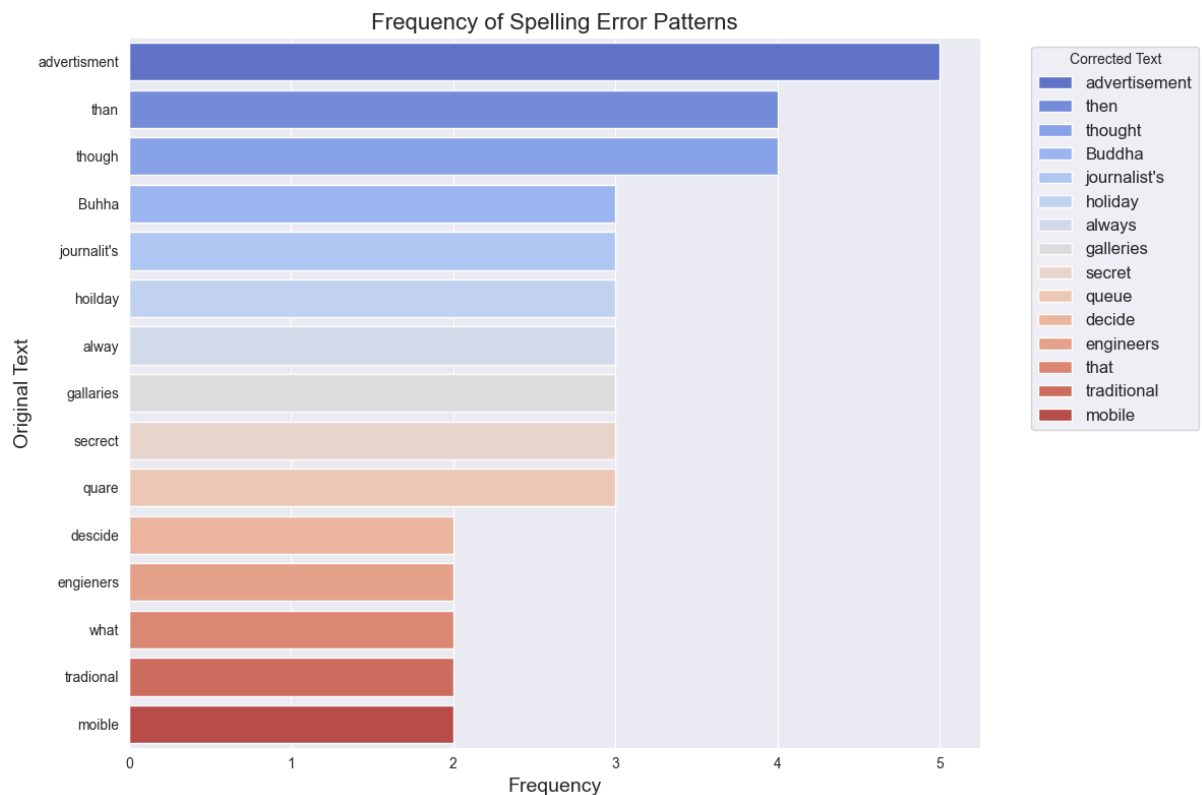
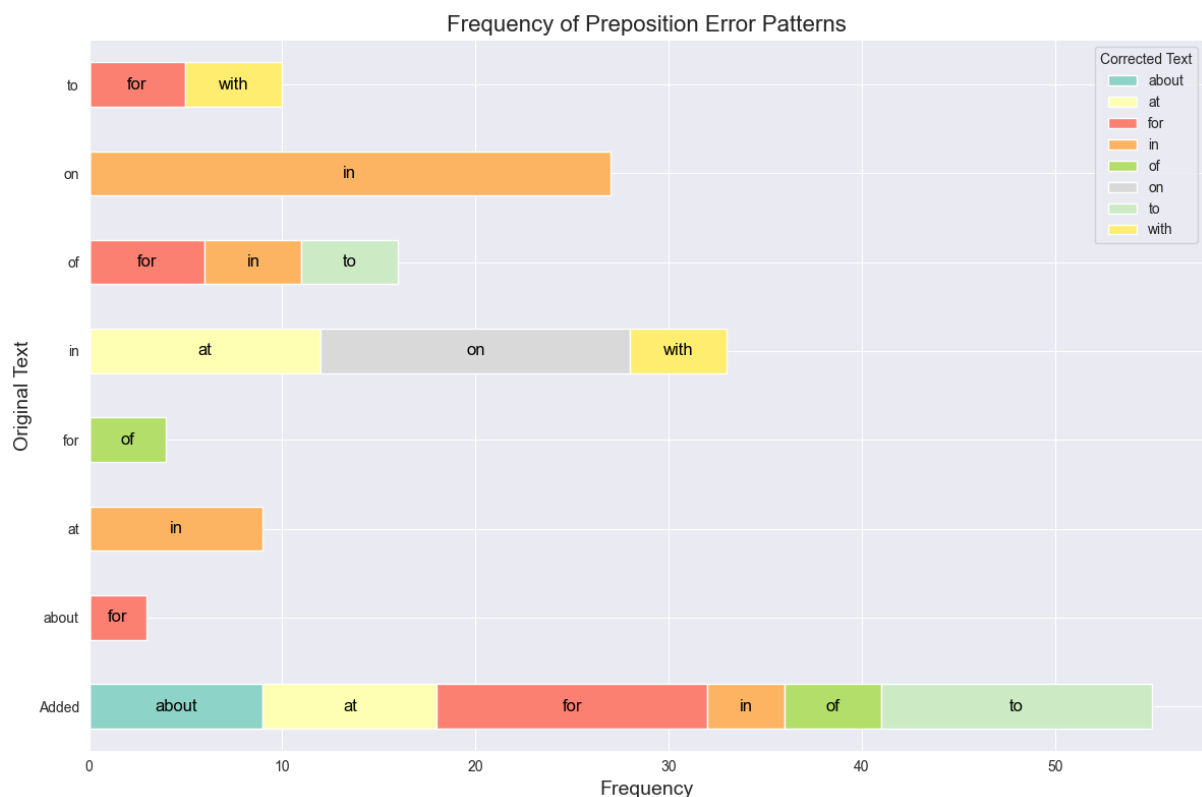


Figure 2. Frequency of Spelling Error Patterns

Determiners error patterns: missing “the” for 136 times and missing “a” for 53 times. Besides, “an”, “their”, “and my” are omitted by Chinese English learners. Some learners did not pay attention to the agreement of quantifiers; they used “this” and “that” when “these” and “those” were correct. This pattern echoes the findings of Ehsanzadeh and Dehnad (2024), who observed similar omission errors in determiner use among EFL medical students. Moreover, in English, definite articles, indefinite articles, and determiners have clear usage rules, while there is no similar article in Chinese, so learners ignored the articles. The inconsistency of quantifiers stems from the learners’ confusion about the usage of quantifiers and demonstratives.

In terms of the use of prepositions, the use of "in" and "on" is the most confusing. In all samples, "on" was used incorrectly as "in" 27 times, and "in" was used incorrectly as "on" 16 times. In addition, "for", "to", "at", and "about" are often forgotten. Figure 3 shows the specific errors and the corresponding correct forms.



**Figure 3.** Frequency of Preposition Error Patterns

Overall, the findings from this analysis correspond with the error taxonomies established in the literature. Mohamed et al. (2004), Zheng and Park (2013), Eng et al. (2020), Zhang et al. (2022), De Wilde (2023), and Ehsanzadeh and Dehnad (2024) all provide complementary evidence that Chinese learners of English frequently exhibit errors across morphosyntactic and lexical domains, rooted in cross-linguistic interference, incomplete acquisition, and rule overgeneralization. These findings collectively reinforce the necessity of targeted instruction in areas such as verb tense coordination, determiner usage, punctuation conventions, and prepositional collocations.

### 4.3 Pedagogical Implications

#### 4.3.1 Establishment of joint classroom

Chinese teachers and English teachers work together to systematically explain the differences in the use of punctuation marks between Chinese and English. There are some significant differences in punctuation rules between Chinese and English. The use of commas in Chinese is relatively flexible, and the use of commas and periods in Chinese is relatively subtle, such as "vividly." An idiom can also end with a period. However, in English writing, the use of punctuation marks is more strict, and the use of commas and periods has clearer specifications. Through communication and explanation between Chinese teachers and English teachers, teachers can help learners understand these differences so that they can avoid writing errors caused by negative transfer of their mother tongue.

#### 4.3.2 Focused training on common errors

The results of the corpus reveal that Chinese learners have some common errors in English writing, especially in verb tense, subject-verb agreement, spelling, punctuation, etc. Teachers can design targeted explanations and exercises based on common errors to help learners focus on solving these problems. For example, special tense comparison exercises can be designed based on tense errors to enable learners to master the correct use of tenses in different situations. For errors in subject-verb

agreement, teachers can design sentence conversion and error correction exercises to help students strengthen their understanding and application of subject-verb agreement rules. For spelling errors, teachers can explain spelling rules and combine them with vocabulary memory training to help students reduce common spelling errors. Through targeted and focused training, students can improve their accuracy in writing in a short period of time, reduce repetitive errors, and thus improve their overall language application ability.

#### *4.3.3 Cultivate awareness of mother tongue interference*

Teachers can guide students to realize that the second language will be interfered with by the mother tongue during teaching. In particular, the differences between Chinese and English in tense and verb agreement will cause Chinese learners to make relevant mistakes when doing English writing tasks. Teachers can emphasize that Chinese does not have clear tense changes, while English tenses need to express different dimensions of time through changes in verb forms. In addition, Chinese does not have definite and indefinite articles like English, and the use of prepositions is more flexible. Chinese learners use or omit these language elements incorrectly in English, which needs to be included in the teaching plan by teachers to cultivate learners' awareness of preventing negative transfer from their mother tongue.

#### *4.3.4 Provide personalized feedback*

Teachers can conduct in-depth analysis and correction of a certain type of errors in students' writing according to the actual situation of each student, so as to help students focus on solving specific problems and avoid general discussions. For example, teachers correct learners' tense usage errors in a writing task and let learners conduct relevant learning and corrections. Through this targeted feedback, students can not only be aware of their mistakes, but also get clear correction directions. Learners can consolidate the correct use of tenses through corresponding exercises, thereby improving the quality of their writing. This personalized feedback can not only save teachers' time in providing feedback, because teachers do not need to point out all the errors in students' writing one by one, but focus on a certain type of problem, thereby improving the efficiency of feedback, but also stimulate students' learning motivation, so that they can clearly understand their room for improvement in writing and focus more on self-improvement.

## **5. Conclusion**

This study explored the types and patterns of writing errors of Chinese English learners in the FCE corpus. The study found that learners have large loopholes in the use of verbs, especially in tense and subject-verb agreement. This is related to the fact that there are no strict requirements for tense and subject-verb agreement in Chinese. Learners are also not proficient in the use of punctuation marks, and there is a phenomenon of native language transfer. The misuse of punctuation marks is also the most frequent error among all error types. Similarly, since there is no clear restriction on the use of articles in Chinese, learners often miss articles in English. The incorrect use of prepositions is also relatively eye-catching, reflecting the characteristics of learners bringing some Chinese language habits into English writing. Based on this, this paper provides four suggestions for existing English teaching. First, establish a joint Chinese and English classroom. Second, conduct concentrated training on common errors. Third, cultivate awareness of native language interference. Fourth, require teachers to provide personalized feedback on students' writing errors.

There are some limitations to this study. First, compared with other corpus-based studies, the sample size is relatively small. Although the sample size is large enough to show statistical significance, due to the large number of error types, each error type may not be sufficient to reflect statistical significance. Secondly, although this study provides relevant suggestions for teaching, these suggestions are more based on the summary of analysis results and have not been verified in actual teaching practice. Therefore, the effects of the suggestions may vary in different teaching environments and

teaching implementations, and further experimental research is needed to verify their effectiveness and feasibility. Finally, due to time and resource constraints, this study did not take into account other factors affecting students' writing ability, such as the interference of psychological factors such as learning strategies and learning motivation. Future research can take a more comprehensive perspective and comprehensively consider these factors in order to better understand the multiple causes of writing errors and propose more precise and personalized teaching countermeasures.

Although this study has certain limitations, it can still provide important references and inspiration for future research. First, in terms of the analysis of common types and specific patterns, this study revealed that Chinese English learners are relatively weak in verb tense, subject-verb agreement, the use of punctuation marks, the omission of articles, and the misuse of prepositions. Second, this study provides specific suggestions for the teaching of English writing, and links the results found in the corpus to provide directions for adjusting teaching strategies from four aspects. Future research can also verify these teaching strategies and further optimize the improvement of Chinese English learners' writing skills. Therefore, this study still provides valuable real-world predictive data for understanding the common error patterns of Chinese learners in English writing, and also provides directions and inspiration for future research.

## References

- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied linguistics*, 15(2), 169-189.
- Chen, H. I. (2010). Contrastive learner corpus analysis of epistemic modality and interlanguage pragmatic competence in L2 writing. *Journal of Second Language Acquisition and Teaching*, 17, 27-51.
- Corder, S. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5(1-4), 161-170.
- Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System*, 26(2), 163-174.
- De Wilde, V. (2023). Lexical characteristics of young L2 English learners' narrative writing at the start of formal instruction. *Journal of Second Language Writing*, 59, 100960.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. Oxford University Press.
- Ehsanzadeh, S. J., & Dehnad, A. (2024). Analysis of high-frequency errors and linguistic patterns in FL medical students' English writing: Insights from a learner corpus. *BMC Medical Education*, 24(1), 1264.
- Eng, L. S., Luyue, C., & Lim, C. K. (2020). A comparison of the English grammatical errors of Chinese undergraduates from China and Malaysia. *International Journal of Instruction*, 13(1), 931-950.
- Gilquin, G. (2024). From second language acquisition research to foreign language teaching through the prism of corpora. *Ampersand*, 13, 100204.
- Granger, S. (2015). The contribution of learner corpora to reference and instructional materials design. In Granger, S., Gilquin, G. & Meunier, F. (eds.) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 486-510.
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO journal*, 20(3), 465-480.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. *Applied Corpus Linguistics: A Multidimensional Perspective/Rodopi*, 52, 123-145.
- Heydari, P., & Bagheri, M. S. (2012). Error analysis: Sources of L2 learners' errors. *Theory and practice in language studies*, 2(8), 1583-1589.
- Lan, G., Zhang, Q., Lucas, K., Sun, Y., & Gao, J. (2022). A corpus-based investigation on noun phrase complexity in L1 and L2 English writing. *English for Specific Purposes*, 67, 4-17.
- Mesham, S., Bryant, C., Rei, M., & Yuan, Z. (2023). An extended sequence tagging vocabulary for grammatical error correction. arXiv preprint arXiv:2302.05913.

- Mohamed, A. R., Goh, L. L., & Eliza, W. R. (2004). English errors and Chinese learners. *Sunway College Journal*, 1(1), 83-97.
- Rei, M. (2017). Semi-supervised multitask learning for sequence labeling. *arXiv preprint arXiv: 1704.07156*.
- Richards, J. C. (1970). A non-contrastive approach to error analysis. *TESOL Convention* (pp. 172-188). San Francisco.
- Satake, Y. (2020). How error types affect the accuracy of L2 error correction with corpus use. *Journal of second language writing*, 50, 100757.
- Stahlberg, F., & Kumar, S. (2020). Seq2Edits: Sequence transduction using span-level edit operations. *arXiv preprint arXiv: 2009.11136*.
- Tono, Y., Satake, Y., & Miura, A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL*, 26(2), 147-162.
- Ueno, S., & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics*, 3(3), 100076.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of second language writing*, 13(4), 257-283.
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2), 31-48.
- Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011, June). A new dataset and method for automatically grading ESOL texts. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies* (pp. 180-189).
- Zhao, W., Wang, L., Shen, K., Jia, R., & Liu, J. (2019). Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data. *arXiv preprint arXiv:1903.00138*.
- Zhang, L., Sun, S., & Yao, S. (2022). Error types of and strategies on learning Chinese connectives: a study on Chinese as a second language learners' writing. *Frontiers in Psychology*, 12, 790710.
- Zheng, C., & Park, T. J. (2013). An Analysis of Errors in English Writing Made by Chinese and Korean University Students. *Theory & Practice in Language Studies (TPLS)*, 3(8), 1342-1351.