

Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Corpus-based analysis of second language writing errors and their pedagogical implications

Yuehua Li¹, Lihao Han^{1,†}

1. Hebei Vocational University of Industry and Technology, Shijiazhuang, Hebei, 050091, China.

Submission Info

Communicated by Z. Sabir
Received September 29, 2024
Accepted January 2, 2025
Available online February 3, 2025

Abstract

Second language writing is not only a major problem for language learners but also a long-term focus of attention in the field of applied linguistics, and the error analysis of second language writing is of great significance for the development of language education at different stages. Multimodal discourse analysis is used to construct a corpus and propose a research methodology for second language teaching in this study. The errors in writing for second language teaching are investigated in detail based on the theory of error analysis in five steps: collecting samples, identifying errors, describing errors, interpreting errors, and evaluating errors. Through the investigation and analysis, it was found that among the four dimensions of functional adequacy, the task requirement dimension had the highest average score (5.71) and the most concentrated data, and the average scores of the comprehensibility (4.55) and coherence and articulation (4.35) dimensions decreased in order. Among the linguistic feature indicators, lexical complexity, lexical diversity, syntactic complexity and writing scores all have significant positive correlations, with correlation coefficients in the range of 0.26 to 0.35 and p-values less than 0.01. Among them, lexical diversity and writing scores have reached a moderate correlation coefficient of 0.35, which indicates that the more lexical varieties there are, the more low-frequency words are found, the more words are included in an average of each T unit, the more words, the higher the overall rating of the essay. The amount of writing errors in the student essay samples was all negatively correlated with the student's essay scores, and the p's were all less than 0.05, which is a very high correlation; the higher the number of writing errors, the lower the essay scores. This dissertation makes recommendations for teaching second-language writing and provides a reference for research on teaching writing in a second language based on the analysis results.

Keywords: Multimodal corpus; Discourse analysis; Error analysis theory; Second language writing; Pedagogical insights.

AMS 2010 codes: 00A35

[†]Corresponding author.

Email address: hanlh@hbcit.edu.cn

1 Introduction

Writing is a language output skill. For English majors, writing is often the most difficult of the four skills of listening, speaking, reading and writing, and writing is also the most demanding for learners [1-2]. In the process of foreign language writing, students not only have to master the basic knowledge of writing, such as spelling, punctuation and the ability to formulate sentences in foreign languages [3-4], but also need to express their thoughts creatively and logically in a foreign language thinking style, so students often make mistakes in spelling, vocabulary, grammar, sentence structure, discourse structure and so on, and it is important to analyze the errors in writing to improve the quality of writing [5-8].

Error analysis is a systematic study of errors made by learners in second and foreign language learning. It aims to understand how learners learn a language, the reasons why learners make mistakes in learning a language, and the common difficulties learners have in learning a language in order to arrange the teaching content and improve the teaching methods [9-12]. Error analysis is conducive to revealing the psychological process and law of students' foreign language learning, enriching and enriching the basic theory of foreign language teaching. The error refers to the phenomenon that language learners appear to deviate from the correct expression of the target language [13-15]. It reflects that the learner has not mastered or incompletely mastered a certain aspect of the target language, so the teacher should pay attention to correcting the students' errors in the teaching process. Errors exist for students because people's awareness of errors then has a gradual development process [16-19].

Writing errors are a common manifestation in the process of second language learning and an important stepping stone for learners to improve and enhance. This study is based on a multimodal corpus of second language writing errors, which can effectively identify problems in the process of students' second language learning. Then, an in-depth study of second-language writing errors is conducted based on the theory of error analysis. Then, an empirical study is carried out using M College as an example, and second language writing is analyzed in detail from the aspects of language characteristics and assignment scores. Finally, the analysis results provide reasonable suggestions from three aspects: enriching resources, strengthening strategy training, and cultivating habits.

2 Corpus-based Error Recognition in Second Language Writing

2.1 Analysis framework based on a multimodal corpus

As a new stage of corpus development, the multimodal corpus is increasingly widely used in the field of language research. Multimodal corpus refers to "a corpus that integrates textual corpus, audio corpus, and static and dynamic image corpus, and allows users to retrieve, statistically and other operations in a multimodal way." Some scholars believe that a multimodal corpus is a database that contains transcribed, processed and labeled linguistic texts and audio and video databases that are closely related to the texts, with the purpose of systematically studying the interaction between linguistic and non-linguistic symbols using empirical methods.

Nowadays, the common multimodal corpus annotation and retrieval software include Anvil, Transcriber, ELAN, Praat, and others. Among them, the ELAN corpus is the most popular one. Not only does the ELAN corpus process audio and video texts, but it also annotates paralinguistic and non-linguistic symbols and performs statistics and analysis on them. In ELAN software, researchers can also choose from five different modes according to their research needs. The annotation mode is used for detailed annotation of the corpus. Synchronization mode is used to synchronize audio, video,

and text. To convert audio and video content into text, the transcription mode is utilized. Segmentation mode divides the corpus into smaller units, which makes it easier for the researcher to analyze it locally. The linear interleaved mode can aid researchers in understanding the complex process of multimodal interaction by displaying information from multiple modalities on a single timeline. Therefore, the multimodal annotation software ELAN software was used to analyze the second language teaching, and a multimodal corpus containing multilevel annotation information was successfully created. ELAN software, as an important tool for multimodal corpus annotation and retrieval, has been widely recognized for its powerful editing and visualization processing functions. The application of multimodal corpora in reading and writing for second language teaching provides teachers with brand-new perspectives and tools.

The framework for multimodal discourse analysis is shown in Fig. 1. Multimodal discourse analysis has become an increasingly rich and diverse research field, with research methods ranging from a single perspective to the organic integration of multiple perspectives. The use of multimodal discourse analysis research in second language teaching quality classes based on multimodal corpora needs to be further explored. Early multimodal corpus research can be mainly divided into general-purpose corpora and specialized multimodal corpora, while multimodal special-purpose corpora in different domains represent specialized multimodal corpora, and the multimodal corpora are usually labeled with multimodal data such as syntactic features, turn-taking, gaze expression, and action behavior. In recent years, research based on multimodal corpora has been flourishing, and the application of multimodal discourse analysis methods has become more and more widespread in academia, which not only promotes the improvement and development of related theories but also provides strong support for the construction and research of multimodal corpora.

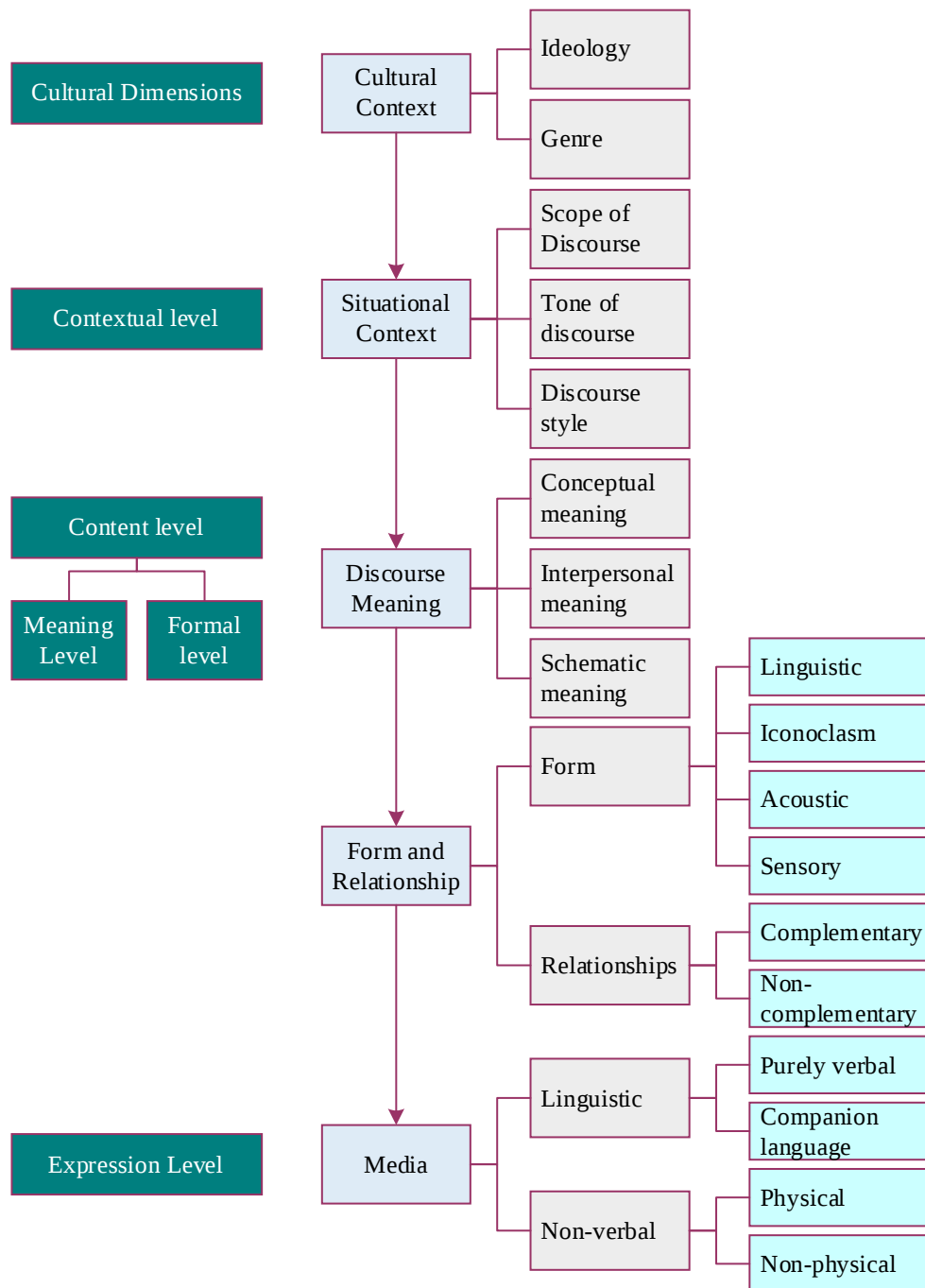


Figure 1. Multi-modal discourse analysis framework

2.2 A corpus-based theory for analyzing second language writing errors

In the process of learning a second language, mistakes are inevitable, which is a normal phenomenon. However, there has not been a unified standard for the concept of error nor a consistent definition to characterize it. Different scholars have different views on an error and give different definitions of “error”. This study focuses on the investigation of errors, which are defined as problems that occur in the use of a second language when learners do not have a good command of the language or when they misunderstand grammatical usage. The term “errors in writing” refers to errors in learners’ compositions that do not comply with the rules and norms of the second language.

Error analysis theory is shown in Figure 2, which is based on the theory of universal grammar and cognitive psychology, and divides error analysis into five steps: collecting samples, identifying errors, describing errors, explaining errors and evaluating errors.

Sample collection. This is started at the initial stage of the study, and the collection of samples is divided into natural sampling and induced sampling, which is conducive to the author's collection of the real language learning situation of the second language acquirers, including the collection of students' writing corpus.

Recognizing Errors. Identifying errors is to examine whether advanced speakers of the target language or native speakers of the target language conform to the grammatical norms of the target language when applying the target language, which is generally done from both grammatical and pragmatic aspects. It's important to check if the language is being used in accordance with the contextual rules.

Describing errors. To categorize the errors in written expressions collected from students. That is, the errors in the students' writing corpus are first identified, and then they are grouped according to different types of errors.

Explaining errors. Finding out the reasons for making errors, i.e., explaining why learners make these errors. Explaining the errors is an important step in error analysis because the ultimate goal of error analysis is to understand the outcomes of language learning for learners, so that teaching can be improved.

Evaluating the errors. The final step of error analysis involves using the appropriate methods to correct and improve the errors made by learners, which can help students improve their English writing.

In this study, the five steps of error analysis theory will be utilized to conduct a detailed study of writing errors in second language teaching.

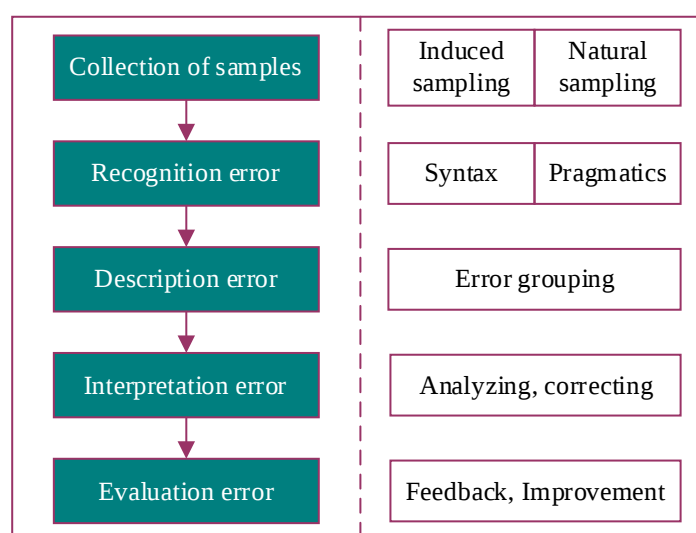


Figure 2. Error analysis theory

3 Second Language Writing Error Analysis Practice and Discussion

In order to fully understand the situation related to students' second language writing, this chapter designed the "Second Language Writing Scale" and measured the reliability of the questionnaire scale. The data collected from the students' writing samples and the questionnaire was processed and analyzed using the corpus. In this study, 276 questionnaires were randomly selected from two grades (freshman and sophomore) in College M. After removing the invalid questionnaires, 245 valid questionnaires were finally collected.

3.1 Corpus-based linguistic analysis of bilingualism

The analysis of the corpus was centered on three dimensions: overall rating of second language writing, linguistic features, and functional adequacy. All scores in this paper were manually scored by the raters, and the scoring criteria were determined by the Overall Writing Rating Scale and the Functional Adequacy Scale. The writing scores were categorized into five levels, with scores ranging from 1 to 9. Six levels make up the Functional Adequacy Rating Scale, with the lowest level being 1 and the highest being 6.

3.1.1 Corpus-based descriptive analysis of language

The results of the descriptive statistics of bilingual language are shown in Table 1. The average score of the test paper's writing performance is 6.25, with 5~6 categorized into three grades, indicating that the subjects' linguistic performance in writing basically reaches the intermediate level, but there is still a certain gap with the advanced level. Among the four dimensions of functional adequacy, the task requirement dimension had the highest mean score (5.71) and the most concentrated data, indicating that the subjects were generally able to fulfill the requirements of the writing task. The average scores of the comprehensibility (4.55) and coherence and articulation (4.35) dimensions decreased in order, and the average scores of the two dimensions were closer, which shows that the average level of language performance of the subjects in these two dimensions was closer to each other. Among the four dimensions, the average score of the content dimension is 0.88, which is the lowest. The value of the degree of dispersion is higher, and the average performance of the subjects is less satisfactory. The six dimensions of linguistic features were assessed in descending order as lexical diversity (18.59), syntactic complexity (13.52), lexical accuracy (0.98), grammatical accuracy (0.75), discourse coherence (0.62) and lexical complexity (0.23).

The mean value of the lexical accuracy measure for the bilingual corpus among the six assessment indicators of linguistic features is 0.98, with a standard deviation of 0.02. It indicates that the mastery and use of words in the writing process are relatively satisfactory, and most people are able to express themselves correctly using words. The mean value of vocabulary complexity is 0.25, and the degree of dispersion of vocabulary complexity dimension is close to that of vocabulary accuracy, which shows that the language performance in terms of word use is relatively similar.

Of the three measures, the lexical diversity dimension has the largest standard deviation and the highest degree of dispersion. The syntactic dimension was examined in this paper by examining syntactic complexity and grammatical accuracy. Syntactic complexity had a larger standard deviation and more discrete data than lexical complexity. Compared to lexical accuracy, the subjects performed poorly on the grammatical accuracy dimension. The standard deviation increased and the degree of dispersion increased. At the discourse level, the mean value of discourse coherence was 0.61, indicating that, on average, there were more than one explicit connecting means between every two T-units in the corpus, the standard deviation was smaller, and the data were more concentrated.

Table 1. Descriptive statistical results of the language of language

Table 1: Descriptive statistical results of the language of language						
Writing scores and functional scores						
	Writing score	Functional adequacy				
		Content	Task requirement	comprehensibility	Coherence and cohesion	
Mean value	6.25	3.83	5.71	4.55	4.35	
Standard deviation	1.06	0.88	0.72	0.94	0.85	
Six dimensions of language characteristics						
Vocabulary			Syntax		Discourse	
Vocabulary accuracy	Lexical complexity	Vocabulary diversity	Normal complexity	Grammar accuracy	Discourse coherence	
0.98	0.23	18.59	13.52	0.75	0.62	
Linguistic characteristics analyze descriptive data						
	Vocabulary accuracy	Lexical complexity	Vocabulary diversity	Normal complexity	Grammar accuracy	Discourse coherence
Mean value	0.98	0.25	19.52	13.32	0.75	0.61
Standard deviation	0.02	0.03	3.75	1.55	0.15	0.17

3.1.2 Correlation analysis of bilingual features and writing

The results of the correlation analysis between the variables of linguistic features and writing are shown in Table 2. Among the six linguistic feature indicators, lexical complexity, vocabulary diversity, syntactic complexity and writing scores have significant positive correlations, and the values of the correlation coefficients are in the range of 0.26 to 0.35, with p-values less than 0.01. Lexical diversity has a moderate correlation coefficient of 0.35 with the writing scores among them. According to the inference, the composition's overall score increases with the greater the vocabulary diversity, low-frequency words, and the number of words contained in each T-unit on average.

The correlations among the measures were further explored, and the results showed that there were 10 sets of variables with significant correlations among the six measures. For the vocabulary dimension, there were two-by-two correlations between the three measures of vocabulary accuracy, vocabulary complexity, and vocabulary diversity, with vocabulary accuracy and vocabulary complexity reaching a moderate level of correlation ($r = 0.34$, p-value less than 0.001). For the syntactic dimension, there was also a significant positive correlation between the variable's syntactic complexity and grammatical accuracy. For the accuracy dimension, there was a significant moderate negative correlation between lexical accuracy and grammatical accuracy ($r = -0.35$, $p < 0.001$), which means that a corresponding increase in lexical errors may accompany an increase in the grammatical accuracy of an essay. At the level of complexity, lexical complexity and syntactic complexity, on the other hand, showed a significant moderate positive correlation ($r = 0.35$, $p < 0.001$), i.e., compositions with high lexical complexity also have high syntactic complexity. The indicator of lexical diversity, on the other hand, showed a significant correlation with both grammatical dimension measures, and the correlation coefficient with grammatical accuracy, with an absolute value of 0.86, reached a large effect size and showed a negative correlation, i.e., compositions with a high variety of words are associated with an increased number of grammatical errors. A significant moderate positive correlation was found between it and syntactic complexity, which shows that the longer the average T-unit length of an essay is when more word variety is used. At the chapter level, discourse coherence and lexical variety showed a significant negative correlation ($r = -0.23$, $p < 0.05$), i.e., they showed a reciprocal relationship. In addition, discourse coherence and syntactic complexity showed a

significant positive correlation ($r=0.28$, $p<0.01$), i.e., compositions with a high level of syntactic complexity also had a high level of discourse coherence, and the two assessment metrics grew in conjunction with each other.

Table 2. The correlation analysis results of each variable and writing of language characteristics

Index	Writing score	Vocabulary accuracy	Lexical complexity	Vocabulary diversity	Syntax complexity	Grammar accuracy	Discourse coherence
Writing score	1						
Vocabulary accuracy	0.05	1					
Lexical complexity	0.26**	0.34***	1				
Vocabulary diversity	0.35***	0.26**	0.22*	1			
Syntax complexity	0.28**	-0.5	0.35***	0.33***	1		
Grammar accuracy	-0.18	-0.35***	-0.01	-0.86***	0.31**	1	
Discourse coherence	-0.12	-0.17	0.08	-0.23*	0.28**	0.38	1

3.2 Analysis of Second Language Writing Errors

In this study, second language writing was categorized into a total of 11 types of errors, including word form errors fm, lexical errors wd, syntactic errors in, verb phrase errors vp, noun phrase errors np, collocational errors cc, pronoun errors pr, prepositional errors pp, adjective errors aj, adverb errors ad, and conjunction errors cj. After the labeling of the students' composition samples, the amount of writing errors was determined through the two grades of the college students' composition samples, where writing errors were retrieved and counted separately, and the amount of writing errors in freshman student essays was labeled as M1, and the amount of writing errors in sophomore student essays was noted as M2.

3.2.1 Comparison of Differences in Writing Errors

The results of the comparison of the variability of writing errors between the two grades are shown in Figure 3, which shows visually and clearly that the two grades had the highest vocabulary error wd, 573 and 555, respectively. The error rate for sophomores was higher than for freshmen on seven dimensions: fm, wd, vp, np, cc, pp, and ad. However, on the remaining dimensions, sn, pr, and aj, freshmen had a greater amount of errors than sophomores the amount of errors is a bit more.

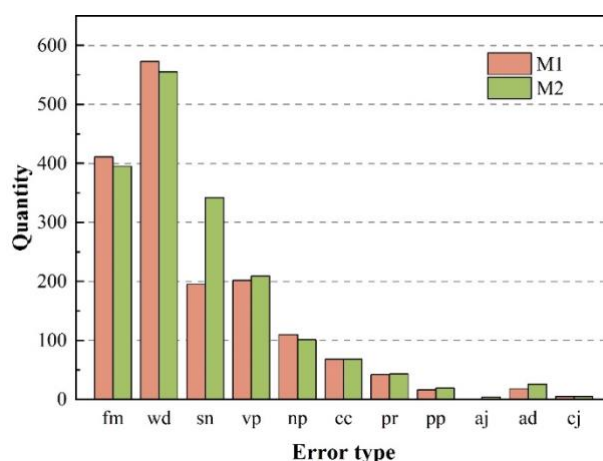


Figure 3. The comparison of the difference between the two grades was compared

In order to investigate the variability of writing errors at different levels among freshman and sophomore students, students who ranked in the top 50% of essay scores at each grade level were categorized as the high subgroup, and those who ranked in the bottom 50% were recorded as the low subgroup. The investigation focused on the differences in writing errors between high and low subgroups in the same grade level.

The analysis of the differences between the high and low subgroups in the freshman year is shown in Table 3. Freshman students made zero errors on conjunctions (aj) and errors on the other ten dimensions (fm, wd, sn, vp, np, cc, pr, pp, ad, cj). In addition, on all these dimensions, the means of the low subgroup of first-year students ranged from 0.081 to 6.951 were greater than the means of the high subgroup, which indicated that the low subgroup of first-year students had a greater amount of errors than the high subgroup on these ten dimensions.

From the results of the independent samples t-test, there is a significant difference between the freshman high and low subgroups in the seven dimensions of fm, wd, sn, vp, np, cc, and pr (Sig. (two-sided) < 0.05), which indicates that the amount of errors of the freshman low subgroups is much more than that of the freshman high subgroups in these seven dimensions. However, there is no significant difference between the freshman high and low subgroups in the three dimensions of pp, ad, and cj (sig. (two-sided) > 0.05), which means that the amount of errors in the freshman low subgroup is not much different from that of the freshman high subgroup in these three dimensions.

Table 3. Analysis of the difference between high scores and low group(M1)

Dimension	Scoring group		Low group		Independent sample t test	
	Mean	Standard deviation	Mean	Standard deviation	T	Sig.
fm	0.961	0.855	5.264	2.362	-10.445	0.000
wd	2.323	1.015	6.951	2.471	-12.431	0.000
sn	1.092	0.931	4.542	2801	-10.585	0.001
vp	0.401	0.286	2.981	0.969	-5.663	0.000
np	0.592	0.971	2.221	1.372	-6.821	0.000
cc	0.311	0.234	0.843	0.482	-2.301	0.001
pr	0.375	0.581	1.092	0.996	-3.721	0.000
pp	0.121	0.612	0.176	0.971	-1.603	0.101
aj	0.000	0.000	0.000	0.000	-	-
ad	0.155	2.889	0.182	0.368	-0.542	0.608
cj	0.002	0.001	0.081	0.293	-1.683	0.092

The analysis of variance between the high and low sophomore subgroups is shown in Table 4. Sophomore students had more or less errors on all nine dimensions of fm, wd, sn, vp, np, cc, pr, pp, and ad. In addition, the mean values of students in the lower subgroup of sophomores ranged from 0.411 to 7.714 on these nine dimensions, all of which were greater than the mean values of the upper subgroup of sophomores, which indicated that the amount of errors made by the lower subgroup of sophomores was greater than that made by the upper subgroup of sophomores on all of these nine dimensions. From the results of the independent samples t-test, the sophomore high and low subgroups of students differed significantly on the eight dimensions of fm, wd, sn, vp, np, pr, cc, and pp (sig. (two-sided) < 0.05), which means that the number of errors made by the sophomore low subgroup of students was much greater than that of the sophomore high subgroup of students on all of these eight dimensions. There was no significant difference between the sophomore high and low subgroups on aj, ad, and cj (sig. (two-sided) > 0.05). The amounts of errors made by the sophomore low subgroup of students were not significantly different from those made by the sophomore high subgroup of students on these dimensions.

Table 4. Analysis of the difference between high scores and low group(M2)

Dimension	Scoring group		Low group		Independent sample t test	
	Mean	Standard deviation	Mean	Standard deviation	T	Sig.
fm	1.102	0.87	6.571	2.702	-9.149	0.000
wd	3.092	1.602	7.714	2.388	-12.981	0.001
sn	0.832	1.339	2.771	2.496	-9.605	0.000
vp	0.932	1.145	2.951	2.1	-6.343	0.001
np	0.482	0.717	1.451	1.153	-7.578	0.000
cc	0.522	0.908	1.541	1.378	-7.654	0.000
pr	0.132	0.892	0.591	0.91	-6.528	0.000
pp	0.192	0.403	1.041	1.022	-2.478	0.000
aj	0.002	0.001	0.004	0.001	-0.871	0.095
ad	0.002	0.001	0.411	1.166	-0.924	0.087
cj	0.002	0.001	0.004	0.001	-0.574	0.095

3.2.2 Correlation analysis of bilingual writing errors

The results of the correlation analysis between writing errors and composition grades in the two grades are shown in Table 5. Through the correlation analysis, it can be seen that the amount of writing errors in the composition samples of the first-year students is negatively correlated with the students' composition grades, with a correlation coefficient of -0.964, which means that the correlation between them is extremely high. In addition, the amount of writing errors in the composition samples of sophomore students is also negatively correlated with their composition grades with a correlation coefficient of -0.941, which means that the correlation between them is extremely high. The higher the number of writing errors in the English composition samples of freshmen and sophomores, the lower their composition scores can be observed from the above.

Table 5. The analysis of the analysis of the error of writing and the achievement of the composition

Grade		Essay score	Error number
M1	Essay score	Correlation coefficient	1
		Sig	—
		N	115
	Writing error	Correlation coefficient	-0.964**
		Sig	0.000
		N	115
M2	Essay score	Correlation coefficient	1
		Sig	—
		N	103
	Writing error	Correlation coefficient	-0.941
		Sig	0.000
		N	103

3.3 Implications and Suggestions for Teaching Second Language Writing

By analyzing the students' second language corpus data and survey results, the current status of students' second language writing has become clearer, revealing the existence of common writing errors, and also discussing the relationship between writing errors and writing performance. In this part, this study will provide insights and suggestions on teaching second-language writing in terms of resource enrichment and strategy training.

1) Expanding Writing Resources

Teachers need to integrate learning materials creatively according to students' age and cognitive ability level, and utilize the background knowledge in textbooks to improve students' basic English cultural knowledge. In this case, students' basic knowledge of English should be broadened so that they can experience the cultural charm and atmosphere of the English language instead of mechanically learning to memorize vocabulary, collocations, grammatical rules, and so on. Secondly, teachers should provide a wide range of English materials. Learning a language can only be achieved with much input.

2) Strengthen writing strategy training

Enhance the training for leadership writing strategy and ensure that it is layered. Teachers should consciously strengthen the guidance and insight into the use of writing strategies, guide learners to develop the habit of independent learning and help learners form the awareness of active learning and use of writing strategies. Teachers should help students focus on using different writing strategies effectively in their daily learning at the same time. Ensure that you fully comprehend and pay attention to the significance of accumulating English materials, and emphasize the application of accumulated materials to writing practice. It is important for them to make full use of the vocabulary and sentence patterns they have learned, while avoiding errors that can disrupt coherence and repetitive use of simple vocabulary.

3) Cultivate the habit of correcting errors

After the writing is done, make sure to reflect and revise promptly and regularly ask teachers and classmates for comments and suggestions. In practice, teachers can plan the teaching of different factors of writing strategies in different learning stages, focusing on one writing strategy in each lesson, and after students have mastered all the strategies, gradually teaching them to use writing strategies comprehensively in the examination.

4 Conclusion

In this paper, we use multimodal corpus analysis to take samples from College M to study the performance of students' second language writing achievement on the four dimensions of functional adequacy and the task requirement dimension has the highest mean score (5.71) and the most concentrated data. The mean scores of the comprehensibility, coherence and articulation dimensions, on the other hand, decreased in order, with 4.55 and 4.35, respectively, and the mean scores were closer, indicating that the students were basically able to fulfill the requirements of the writing task and that the mean levels of linguistic performance on the dimensions of comprehensibility and coherence and articulation were closer to each other.

The mean values of vocabulary accuracy and complexity measures are 0.98 and 0.25, respectively, and the degree of dispersion of vocabulary complexity dimension is closer to that of vocabulary accuracy, which shows that most of the student's mastery and use of words in the process of writing is more satisfactory, but the language performance in terms of word use is more similar. Meanwhile, the lexical diversity dimension has the largest standard deviation and the highest degree of dispersion, while the syntactic complexity, lexical accuracy, and grammatical accuracy dimensions have relatively large discrete ranges. The mean value of discourse coherence is 0.61, and the standard deviation is smaller. The data is more concentrated.

In the correlation analysis, vocabulary complexity, lexical diversity, syntactic complexity and writing achievement all have significant positive correlations, with the values of the correlation coefficients in the interval of 0.26~0.35, and the p-values are all less than 0.01. Among them, vocabulary diversity has the highest correlation coefficient with writing achievement, which suggests that a rich variety of vocabulary reduces the frequency of low-frequency words and that the more words an average unit contains, the higher the second-language composition's overall level is. All writing errors were negatively correlated with students' composition scores, and the p's were all less than 0.05, indicating a very strong correlation.

Synthesizing the results of the above writing analysis, this paper puts forward suggestions from three aspects: expanding writing resources, strengthening writing strategy training and cultivating error correction habits, with a view to providing insights for second language writing teaching and improving students' writing level.

Funding:

- 1) The project organizer: China Metallurgical Education Association, the project title: The teaching reform of English corpus construction in metallurgical industry (project number: 2022YB33).
- 2) Science and technology research project of Hebei University: Mathematical and physical simulation of material movement and gas flow distribution of melting gasifier (project number: QN2023199).

- 3) Project supported by Natural Science Foundation of Hebei province: study on influence of Coke charging system on material movement and airflow distribution of melting gasifier (project number: E2017417008).

References

- [1] Reichelt, M., Lefkowitz, N., Rinnert, C., & Schultz, J. M. (2012). Key issues in foreign language writing. *Foreign Language Annals*, 45(1), 22-41.
- [2] Manchón, R. M., & Matsuda, P. K. (Eds.). (2016). *Handbook of second and foreign language writing* (Vol. 11). Walter de Gruyter GmbH & Co KG.
- [3] Liu, M., & Ni, H. (2015). Chinese University EFL Learners' Foreign Language Writing Anxiety: Pattern, Effect and Causes. *English Language Teaching*, 8(3), 46-58.
- [4] Bartan, Ö. Ş. (2017). The effects of reading short stories in improving foreign language writing skills. *The Reading Matrix: An International Online Journal*, 17(1), 59-74.
- [5] Sumarti, S., & Widodo, P. (2020). Writing Errors in Students' Foreign Language Acquisition. *Eralingua: Jurnal Pendidikan Bahasa Asing dan Sastra*, 4(2), 215-223.
- [6] Kazazoğlu, S. (2020). The impact of L1 interference on foreign language writing: A contrastive error analysis. *Journal of Language and Linguistic Studies*, 16(3), 1168-1188.
- [7] Amara, N. (2015). Errors correction in foreign language teaching. *The Online Journal of New Horizons in Education*, 5(3), 58-68.
- [8] Qadir, S. M., & Bostanci, H. B. (2023, December). An analysis of English as a foreign language undergraduate students' written errors. In *Forum for Linguistic Studies* (Vol. 5, No. 3, pp. 1881-1881).
- [9] Khansir, A. A. (2012). Error analysis and second language acquisition. *Theory and practice in language studies*, 2(5), 1027-1032.
- [10] Jobeen, A., Kazemian, B., & Shahbaz, M. (2015). The role of error analysis in teaching and learning of second and foreign language. *Education and Linguistics Research*, 1(2), 52-62.
- [11] Zafar, A. (2016). Error analysis: a tool to improve English skills of undergraduate students. *Procedia-Social and Behavioral Sciences*, 217, 697-705.
- [12] Mashoor, B. B. N., & Abdullah, A. B. (2020). Error analysis of spoken English language among Jordanian secondary school students. *International Journal of Education and Research*, 8(5), 75-82.
- [13] Zaki, M. (2015). Implications of Contrastive Analysis and Error Analysis in Second Language Acquisition. *Journal in English Language Teaching*, 2(1), 1-14.
- [14] Crosthwaite, P. R. (2013). An error analysis of L2 English discourse reference through learner corpora analysis. *Linguistic Research*.
- [15] Sompong, M. (2013). Error analysis. *Thammasat Review*, 16(2), 109-127.
- [16] Agustinasari, E., Simanjuntak, T., & Purwanto, M. B. (2022). A review on error analysis stages in teaching English structure. *Pioneer: Journal of Language and Literature*, 14(1), 253-268.
- [17] Hinkel, E. (2018). Error analysis. *The TESOL encyclopedia of English language teaching*, 1-5.
- [18] Sihombing, R., Nissa, A. K., & Estrelita, A. (2015). Students Written Production Error Analysis in the EFL Classroom Teaching: A Study of Adult English Learners Errors. *LLT Journal: A Journal on Language and Language Teaching*, 18(2), 125-132.
- [19] Karim, A., Mohamed, A. R., Ismail, S. A. M. M., Shahed, F. H., Rahman, M. M., & Haque, M. H. (2018). Error analysis in EFL writing classroom. *International Journal of English Linguistics*, 8(4), 122-138.