International Journal of Social Science and Education Research Studies

ISSN(print): 2770-2782, ISSN(online): 2770-2790

Volume 03 Issue 05 May 2023

DOI: https://doi.org/10.55677/ijssers/V03I5Y2023-27, Impact Factor: 5.574

Page No : 949-953

A Study on the Regional Differences of Lexical Richness in High School English Writing

Wen Jinxin
School of Foreign Languages, CWNU, Nanchong, China

ABSTRACT

The cornerstone of studying and writing English is lexicon. For lexical evaluation and writing evaluation, lexical richness is a crucial factor. This paper aim to examine the differences in lexical richness among 284 similar writing essays from Jiangsu, Sichuan, and Yunnan using Treetagger, PowerConc, and Range32 from four dimensions: lexical density, lexical diversity, lexical complexity, and lexical error. According to the research, Jiangsu province has a somewhat higher level of lexical diversity and complexity than Sichuan and Yunnan regions, but both have slightly lower lexical density and error rates. This study makes recommendations for vocabulary teaching based on this.

Published Online: May 30, 2023

Keywords:
Lexical richness, senior high school students, region, English writing

1. INTRODUCTION

Lexicon is the cornerstone and primary component of learning English, so it is natural that lexical assessment plays a significant role in the study of second language acquisition. Corpus has become an innovative tool for English learning and teaching, and the development of lexical measurement and analysis software has provided feasibility for lexical assessment. A crucial component of lexical evaluation is lexical richness. Lexical density, lexical diversity, lexical complexity, and lexical errors are the four variables that Read (2000) sees as directly reflecting lexical richness. Numerous scholars have studied lexical richness in recent years. However, these studies pay attention to the variations in lexical richness with an emphasis on the university level. Few studies have examined the variations in lexical richness between regions and high school stages in writing. In addition, there has long been a disparity in educational standards between the southwest region and the eastern coastal areas, and it is indisputable that English teaching in the southwest region is of lesser quality and effectiveness than that in the coastal districts. As a result, this study chooses 284 essays from Pigaiwang (142 from each of Jiangsu, Sichuan, and Yunnan regions), uses the corpus tools Treetagger, PowerConc, and Range32 to explore the differences in the richness of English writing lexicon between regions, and hopes to offer new insights and ideas for teaching English writing vocabulary in the Sichuan and Yunnan regions.

II. LITERATURE REVIEW

Definition of Lexical Richness

The controversy over the definition of lexical richness mainly lies in the dimensions it encompasses. Firstly, lexical sophistication, lexical individuality, lexical density, and lexical variation were the four subcategories that Linnarud (1986) divided lexical richness into. Secondly, according to Laufer (1991) and Nation (1995) , there are four dimensions for assessing lexical richness: lexical variability, lexical density, lexical lexical sophistication, and lexical originality. Thirdly, the variability of the lexicon with errors, the variability of the lexicon without errors, the proportion of lexicon errors, and the density of the lexicon are all factors that Ngber (1995) put in the measurement range of lexical richness. Fourthly, Wolfe-Quintero, Inagaki, and Kim (1998) contended that lexical richness describes the sophistication and variety of an L2 learner's productive vocabulary. It has come to be recognized as a key construct in L2 teaching and research since it directly links to the learner's ability for effective oral and written communication (Lu, 2012). Fifthly, for lexical richness, sometimes referred to as lexical diversity or lexical complexity (Read, 2000; Daller et al., 2003). However, according to Read , lexical originality is not a good way to assess how well learners are learning new vocabulary. Instead, lexical richness should be evaluated in terms of four different factors: lexical diversity (type/token ratio), lexical complexity (percentage exceeding the top 2000 vocabulary),

Corresponding Author: Wen Jinxin

Cite this Article: Wen Jinxin (2023). A Study on the Regional Differences of Lexical Richness in High School English Writing. International Journal of Social Science and Education Research Studies, 3(5), 949-953

lexical density (percentage of meaningful words in the total vocabulary), and a small number of lexical errors (Read, 2000).

In order to assess the differences of lexical richness in Jiangsu as well as Sichuan and Yunnan, this paper will use Read's definition of lexical richness, which includes lexical density, lexical diversity, lexical complexity, and lexical errors.

Previous Research

Research on the richness of lexicon in English writing can be mainly divided into two categories: The first one is research on measurement tools; the second is the study of lexical richness of different participants.

From the perspective of measurement tools, there is relatively little research on this topic in China, mainly including Bao Gui and Wang Xia (2005) researched on the use of Range. Via Range, they compared the variations in the amount of the productive vocabulary and the language of the learners, and discovered a propensity for vocabulary overuse. Range has excellent potential, operability, and openness characteristics in the perspective of Bao Gui and Wang Xia. Also, it is acceptable for teachers to assess students' productive vocabulary. The research on measurement tools abroad mainly includes Richard and Malvern (1997), Vermeer (2000), and Jarvis (2002).

About participants of lexical richness, college students with and without an English major as well as those learning a second language and those learning their native tongue make up the majority Meanwhile, the research mainly involves a diachronic study of the development of lexical richness and the relationship between lexical richness and writing quality. For example, Zhu Huimin and Wang Junju (2013) studied 120 argumentative papers written by 30 English majors in the four years from their freshman year to their senior year to explore the development characteristics of students' changes in lexical diversity, lexical density, lexical complexity, word length and word frequency distribution. Zhu Huimin and Liu Yanmei (2021) used visual data processing technology and interviews to conduct a case study on the dynamic development characteristics and influencing factors of lexical complexity of two non English majors. What’s more, Wan Lifeng (2010) used 200 CET-4 and CET-8 essays from 100 English major students in Shanghai as research materials to explore the development trend of lexical diversity, complexity, and errors. The research results showed that lexical richness affects writing quality. In addition, Wang Haihua and Zhou Xiang (2012) employed software such as Range, AntConc, Gotagger, and SPSS to study the developmental characteristics of 30 non English major students in four dimensions: lexical complexity, lexical diversity, lexical density, and lexical errors, as well as their relationship with writing quality.

III. RESEARCH DESIGN

Research Questions

This paper is based on the four dimensions of Read and aims to identify the differences in lexical richness between high school students in Jiangsu as well as Sichuan and Yunnan. Research questions are as follows:

1.What are the differences in lexical density between high school students in Jiangsu, Sichuan, and Yunnan regions?

2.What are the differences in lexical diversity between high school students in Jiangsu, Sichuan, and Yunnan regions?

3.What are the differences in lexical complexity between high school students in Jiangsu, Sichuan, and Yunnan regions?

4.What are the differences in lexical errors among high school students in Jiangsu, Sichuan, and Yunnan regions?

Research Materials

The research corpus of this study are 284 essays written by high school students on the same topic from Pigaiwang . Among them, there are 60 compositions each in the first year, 42 compositions each in the second year, and 40 compositions each in the third year in Jiangsu as well as Sichuan and Yunnan regions. The writing theme for the same topic is "Work Hard To Make Our Dreams Come True", with a requirement of around 300 words and appropriate content and language. In view of the fact that this paper studies the richness of lexicon, the author omits the essays that do not meet the requirements of word number, language or content fragmentation, punctuation confusion and other problems, and finally selects 284 articles randomly from the articles that meet the requirements as corpus.

Research Instruments and Measures

This paper uses three tools, namely Treetagger, PowerConc, and Range32. Firstly, TreeTagger is an automatic part of speech coder that supports part of speech tagging in four languages, including English, German, French, and Italian. It also has the function of word form restoration. Secondly, PowerConc, designed by professors Xu Jiajin, Liang Maocheng and Jia Yunlong (2012), is an integration of a series of software developed by the former Beijing Foreign Language Database Linguistics Team, which expands and optimizes the traditional functions of vocabulary indexing, vocabulary generation, subject word calculation, etc.. Thirdly, Range32 was developed by Professor Paul Nation and is designed based on word frequency analysis. It comes with several basic word lists, with a base table based on word families. The inflection and derivative forms of lexicon are grouped into the same word family. The concept is that once learners master words such as accept, they will naturally use accepted and accepted (Tang & Liang, 2021). The lexicon used in Range32 in this study is the basic lexicon of Laufer and Nation (1995), including the most commonly used high-frequency words, the second most commonly used high-frequency words, the academic vocabulary, and the first three off list words.

Firstly, this paper uses Linnarud's (1986) measurement method in the lexical density dimension to calculate the percentage of content words in the total number of words in the composition. Secondly, lexical diversity refers to the use of various different words such as synonyms, superlatives, and other related words in writing, while avoiding the repeated use of certain words (Read, 2000). Therefore, the measurement indicators for the lexical diversity in this study are set as class symbols and the ratio of class symbols to form symbols. Thirdly, the lexical complexity dimension refers to the ability to appropriately use low-frequency words related to the theme and style in the text, rather than just using commonly used high-frequency words (Wang & Zhou, 2012). Therefore, its measurement indicators consider Liu Donghong's (2003) method of calculating the ratio of "academic words" and "off table words" beyond the 2000 high-frequency words (the most frequently used first 1000 words and the second most frequently used 1000 words) in the Range to the total form symbol. Lastly, lexical errors are an important dimension in lexical richness and an important source of information for English learning and teaching; This study uses Zhang Huiping's (2020) method of measuring lexicon bias rate to screen for misspelled words in the "off table words" of Range, and counted the bias rate of each grade and region.

Research Procedures

The research steps of this paper mainly include: firstly, collecting high school English writing essays with the same topic from Pigaiwang, with the title requirement of "Work Hard To Make Our Dreams Come True". The specific writing topic is not limited, and the genre is not limited; The required word count is 300-500 words; Remove articles that do not meet the writing requirements and randomly select 284 articles by grade and region from those that meet the requirements. Secondly, 284 essays will be entered in the form of txt. and each corpus document will be encoded to create a corpus. Then, use TreeTagger software to encode each corpus, use PowerConc to search for nouns, verbs, adjectives, adverbs, pronouns, and numerals, and use Range32 to obtain relevant information such as lexical complexity, and lexical errors. Finally, discuss the results and draw conclusions.

IV. RESULTS AND DISCUSSION

Differences of Lexical Density

The measurement results of lexical density in this study are obtained by the ratio of the number of content words to the number of tokens. From Table 1, it can be seen that the lexical density of Jiangsu, Sichuan, and Yunnan is 62.60% and 62.98%, respectively. The difference in lexical density between the two regions is not significant, with Jiangsu slightly lower than Sichuan and Yunnan. The reason for this result may be that Jiangsu Province students with higher English proficiency have a richer vocabulary and pay more attention to simple expression of meaning when writing, while Sichuan and Yunnan regions have to use longer sentences and more words to express the same meaning.

Table 1. Differences in lexical density

Measures	Jiangsu			Sichuan and Yunnan		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Token	21251	13546	14358	21410	14997	14305
Content word	13316	8475	8984	13737	9486	8800
Lexical density	62.60%			62.98%		

Differences of Lexical Diversity

The diversity of lexicon is represented by the ratio of standard types to tokens. Table 2 shows that the lexical diversity in Jiangsu is 44.58%, while in Sichuan and Yunnan is 42.63%. There is a difference between the two, with the former being higher than the latter. The reason may be that the areas where students in Sichuan and Yunnan are located are relatively underdeveloped, with relatively poor educational resources. English education is mostly for exam taking, and the learning content only revolves around textbooks. Students' vocabulary expansion ability is low, and the possibility of using repetitive words is high. However, Jiangsu students have stronger abilities and have the ability to pay attention to the diverse changes in lexicon, avoid repeating the same words, and have a higher ability to produce vocabulary. Also they have more opportunities to interact with the English language environment and can continuously accumulate more authentic vocabulary.

Table 2 Differences in lexical diversity

Measures	Jiangsu			Sichuan and Yunnan		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Token	21251	13546	14358	21410	14997	14305
Type	9522	6516	6244	9128	6401	6089
Lexical diversity	44.58%			42.63%		

Differences of Lexical Complexity

This research measures lexical complexity through the ratio of academic words and off table words to tokens. From Table 3, it can be seen that the lexical richness of Jiangsu students and Sichuan and Yunnan students is 4.56% and 3.90%, respectively. The latter focuses on the most commonly used 1000 words in writing, while the former not only uses these 1000 words, but also extensively uses secondary high-frequency words and academic vocabulary. The reason for this result may be that English education in Sichuan and

Yunnan regions is at a relatively low level, which makes students seek stability in writing and try to avoid using unfamiliar words, instead turning to familiar high-frequency words; On the contrary, students in Jiangsu have a wider range of knowledge, diverse sources of information, and more diverse learning objectives for English, providing more opportunities to understand and master academic and low-frequency words.

Table 3 Differences in lexical complexity

Measures	Jiangsu			Sichuan and Yunnan		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Token	21251	13546	14358	21410	14997	14305
Academic lexicon	1.20%	1.48%	1.38%	0.97%	1.15%	1.34%
Off table lexicon	2.94%	3.70%	2.99%	2.38%	2.67%	3.19%
Lexical complexity	4.56%			3.90%		

Differences of Lexical Errors

The measurement of lexical errors in this study is obtained by comparing the frequency of errors in off table words to the ratio of tokens. From Table 4, it can be seen that the lexical error rate of Jiangsu students is 0.92%, while that of Sichuan and Yunnan students is 0.66%. The lexical error rate of Jiangsu high school students is higher than that of Sichuan and Yunnan high school students. The reason may be that Jiangsu students pay more attention to higher dimensions such as sentence diversity, discourse logic, and language fluency when writing, and neglect the accuracy of vocabulary spelling, resulting in slightly more lexical errors; However, students from Sichuan and Yunnan have relatively lower writing abilities than the former, with a focus on accuracy and a tendency not to use low-frequency words, reducing lexicon error rates.

Table 4. Differences in lexical errors

Measures	Jiangsu			Sichuan and Yunnan		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Token	21251	13546	14358	21410	14997	14305
Error frequency	181	142	122	147	75	113
Error rate	0.92%			0.66%		

V. CONCLUSION

This study uses 284 essays on the same topic writing of high schools in Jiangsu, Sichuan, and Yunnan regions as research corpus, and employs corpus tools Treetagger, PowerConc, and Range32 to explore the differences in lexical richness between the two regions from four dimensions: lexical density, lexical diversity, lexical complexity, and lexical errors. This research has found that there are differences in all four dimensions between the two regions. In terms of lexical density and lexical error rate, Sichuan and Yunnan regions are slightly lower than Jiangsu regions; In terms of lexical diversity and complexity, Jiangsu region is slightly higher than Sichuan and Yunnan regions.

Based on the above results, the author believes that there is still room for improvement and development in vocabulary teaching in both regions. For the eastern coastal areas represented by Jiangsu, English education is relatively developed. However, when paying attention to the more diverse and complex use of vocabulary, teachers should not ignore lexical accuracy and should prepare drills for students to practice. Accuracy is the foundation of vocabulary learning and development, as well as the foundation of English writing. For the southwestern region represented by Sichuan and Yunnan, teachers should not only focus on the quantity and accuracy of vocabulary output, but also strive to help students expand their knowledge and expose them to more low-frequency words to a certain extent; At the same time, teachers can improve the quality of their vocabulary production by providing students with more opportunities to access native language materials.

Due to the limited proficiency of the author, this paper has some shortcomings: firstly, when measuring lexical complexity, this study did not consider incorrect and correct off table words separately; Secondly, this study directly used the vocabulary provided by Range32, which is not entirely suitable for studying Chinese students' writing; Finally, the sample size involved in this study is still small. Future researchers can create their own lexicon lists according to the specific research situation, process lexicon in more detail, and expand the sample size.

REFERENCES

1. BAO Gui & WANG Xia. 2005. Use of RANGE in Assessing L2 Productive Vocabulary. 《Technology Enhanced Foreign Language, (04): 54-58.
2. Daller, H, Van Hout R. & Treffers-Daller J. 2003. Lexical richness in the spontaneous speech of bilinguals. Applied Linguistics, (2), 197-222.
3. Engber, C A. 1995. The relationship of lexical proficiency to the quality of ESL compositions. Journal of Second Language Writing, (2), 139-155.
4. Jarvis, Scott. 2002. Short texts, best-fitting curves and new measures of lexical diversity. Language Testing, (19), 57-84.
5. Laufer, B. 1991. The development of L2 lexis in the expression of the advanced learner; The Modern Language Journal, (4), 440-448.

6. Laufer, B., & Nation, P. 1995. Vocabulary size: Lexical richness in L2 written production. Applied Linguistics, (3), 307-322.
7. Linnarud, M. 1986. Lexis in Composition: A Performance Analysis of Swedish Learners&apos; Written English. Lund, Sweden: Gleerup.
8. LIU Donghong. 2003. The Influence of Vocabulary Size on EFL Writing. Modern Foreign Languages, (02): 180-187.
9. Read, J. 2000. Assessing Vocabulary. Cambridge University Press.
10. Richards, Brian., & David. Malvern. 1997. Quantifying lexical diversity in the study of language development. Reading:The University of Reading New Bulmershe Papers.
11. TANG Meihua & LIANG Maocheng. 2021. A Study of Level Difference in College English Textbooks' Lexical Complexity. Foreign Language Education in China, 4(01): 61-68.
12. Vermeer, Anne. 2000. Coming to grips with lexical richness in spontaneous speech data. , (17), 65-83.
13. WAN Lifang. 2010. An Empirical Investigation into Lexical Diversity of Chinese English Majors' TEM Writings. Foreign Language World, (01): 40-46.
14. WANG Haihua & ZHOU Xiang. 2012. A Longitudinal Study on the Features of Lexical Richness in Writing by University Non-English Majors. Foreign Languages and Their Teaching, (02): 40-44.
15. Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. 1998. Second language development in writing: Measures of fluency, accuracy and complexity. Honolulu:University of Hawai'i, Second Language Teaching and Curriculum Center.
16. ZHU Huimin & LIU Yanmei. 2021. A Case Study on the Dynamic Development of Lexical Complexity in Second Language Writing. Shandong Foreign Language Teaching, 42(05): 54-64.
17. ZHU Huimin & WANG Junju. 2013. Developmental features of lexical richness in English writing: A self-built corpus-based longitudinal study. Foreign Language World, (06): 77-86.
18. Xiaofei, Lu. 2012. The relationship of lexical richness to the quality of esl learners' oral narratives. Modern Language Journal, 96(2), 190-208.
19. XU Jiajin & JIA Yunlong. 2012. The Design and Development of R-gram Based Corpus Analysis Tool 'Power Conc'. Technology Enhanced Foreign Language, (149): 58.
20. ZHANG Huiping. 2020. On Developmental Features of Lexical Richness in EFL Writing by Chinese Beginner Learners of English. Modern Foreign Languages, (04): 529-540.