

# Lexical Development of English Major Students

-- A Study Based on Written English Corpus

Liwei Zou

School of Languages and Literature, University of South China, Hengyang, Hunan, 421001, China

---

**Abstract:** Lexical evaluation has become undoubtedly one of the cornerstones in the study of second language learning. Not only can it cast some light on the process of second language learning, but also help us gain some insights into the shared characteristics of language learners. Based on a self-built corpus of 120 English timed compositions on the same topic written by 30 Chinese English majors throughout their four-year undergraduate study, this paper tends to explore the features of lexical development in English writing by English majors. The findings are as follows: a solid lexical development in general increases along with the grade, but the progress is not straight-line. Academic word lists present a linear increase; the lexical variation is also rising and learners have a wider range of words to choose from. However, compared with the lexical variation and lexical sophistication, lexical density shows a slower increase. In addition, some high-frequency words are overused by the students in their writings.

**Keywords:** Lexical development, Lexical richness, English majors.

---

## 1. Introduction

Since 1980s, lexical development in L2 writing has attracted the attention of domestic and foreign scholars and researchers. With the development and application of vocabulary measurement tools and software, the corpus study of L2 writing has become more scientific and feasible (Laufer 1994; Laufer & Nation 1995; Bachman 2000; Hu 2020). Lexical studies in L2 writing center on lexical breadth and lexical depth. Among them, the breadth of vocabulary of language learners mainly refers to the lexical richness, which is regarded as an important indicator of the overall level of second language writing (Laufer & Nation 1995; Lemmouh 2008). Taking English majors as the research subject, this paper examines the development and changes of lexical richness in the learners' English writing in four years. The purpose of this research is two-fold: firstly, using a self-built diachronic corpus to examine the learners' vocabulary from the perspective of output; and secondly, to explore the development and changes of learners' vocabulary output at various stages, and to reveal the dynamic characteristics of lexical development.

Lexical richness is regarded as an important index to predict the general language ability and writing level of learners on productive vocabulary (Laufer & Nation, 1995). Lexical richness is also called lexical diversity or lexical complexity. Read (2000) pointed out that lexical richness included four aspects: lexical variation, lexical complexity, lexical density and a small number of lexical errors. Laufer (2003) defined lexical richness as the lexical complexity and lexical variability. To gain insight into the lexical development, Cobb (2003: 401) argues that "the ideal approach would be to use a large number of corpora from the same set of learners over several years." In addition, the word frequency distribution which is based on the word frequency profile, can also reflect the characteristics of lexical richness. It is held that writing is the best way to evaluate the language level of learners as writing requires the support of grammar, vocabulary, and cultural knowledge of the target language,

especially lexical richness. Therefore, based on the above theoretical framework of lexical richness, this research measures the productive lexical richness from these three dimensions including lexical variation, lexical sophistication and lexical density to represent productive vocabulary richness of English major students.

## 2. Relevant Research

Studies on lexical richness in second language writing began in the 1980s, along three main lines and research findings vary with the different subjects selected in each category: (1) diachronic study of the same group. For example, Laufer (1994) conducted an in-depth diachronic study of 48 first-year English majors in Israel in one academic year to explore lexical richness in the English compositions. They collected the compositions at the end of the first and second semester and found that the subjects had made significant progress in terms of lexical sophistication and complexity during the follow-up survey, but the vocabulary variability did not significantly improve. Laufer (1998) further studied the characteristics of the productive vocabulary development of the tenth and eleventh grade English learners in Israel. The results showed that the eleventh grade students' productive vocabulary progress was not obvious, and there was a plateau phenomenon. In China, Wan Lifang (2010) studied 200 CET-4 (2003) and CET-8 (2005) writings of 100 English majors in four colleges in Shanghai and it was found that the complexity and diversity of vocabulary have improved as students developed to a higher level of English proficiency. (2) cross-sectional study of different groups. Leńko-Szymańska (2002) selected 100 freshmen and 67 senior students of college English majors in Poland as subjects to study lexical changes in second language writing. It was found that the writings of senior English majors are significantly better than that from freshman students. However, Wu Juan (2013) carried out a research based on SWECCCL (Chinese spoken and written language corpus) to study the compositions of four grades of students with different levels and revealed that the productive vocabulary might have fossilization or plateau phenomenon.

(3) the correlation study between lexical richness and writing quality. Engber (1995) studied 66 time-limited essays on the same topic written by international students at Indiana University, and found that vocabulary plays a very important role in the construction of meaningful texts. Liu Donghong (2003) stated that the output vocabulary had no direct effect on the writing quality after studying 57 college sophomores' compositions. Wen Qiufang and Wang Lifei (2007) found that lexical richness was positively correlated with the English level after investigating the compositions of 133 non-English majors in the second year in colleges and universities.

### 3. Research Design

Although the lexical richness in English writing has attracted the attention of many researchers at home and abroad, most of the relevant studies are cross-sectional studies or ones within a short period of follow-up research. This study adopts a longitudinal research method to examine the development and changes of vocabulary richness in English writings of 30 English majors in four years, and specifically explores the following questions: (1) What are the major features of lexical richness in English writings of English majors? (2) Is there any correlation between the development of lexical richness and the improvement of students' English proficiency?

This research data come from the writings of 30 English major students in four years in a Chinese domestic university. The students all have at least 6 years of experience in studying English in China before college admission. They were required to write an argumentative essay on a same topic every year during the four-year undergraduate study. The compositions should be completed within 45 minutes without consulting dictionary or any other reference materials. It took 4 years to collect altogether 120 valid compositions completed by students under the same writing requirement.

Research instruments of study on lexical richness were based on composition corpora and the vocabulary frequency software including Range 32 which can help calculate the number of words. First of all, the productive vocabulary of 120 writing texts was analyzed by the Range 32 text by text, and the productive lexical analysis report of each writing can be received. Secondly, the number of content words could be analyzed with the second time using Range 32 through the "stop list" function so as to measure the vocabulary density. Furthermore, variables of each composition were put into excel table to further analyze the lexical richness. Basic vocabulary list designed by Laufer and Nation(1995) was used in this study.

## 4. Results and Discussion

Data results and discussion about lexical richness in writings are presented in this part. The analysis of vocabulary development profile is unfolded from three dimensions: lexical variation, lexical sophistication and lexical density.

### 4.1. Features of Lexical Variation Development

Results show that the standardized TTR in the first- to fourth-grade students' compositions are 58.47%, 61.35%, 63.45% and 62.76%, respectively. As is shown in table 1, the students' lexical variation in writing is on the rise from the first year to the third year, but from the third year to the fourth year, there is a slight decline. An analysis of the writing texts has found that the length of the composition of the fourth

grade students increased significantly, and most compositions contain more than 400 words. This phenomenon shows that the growth rate of the fourth grade students' vocabulary has slowed down, or even has a downward trend. There are two main reasons accounting for the phenomenon of vocabulary plateau: insufficient language input restricts language output. Firstly, senior college students are facing greater employment pressure, so locating an ideal job takes up much more of their time in the fourth year at university. Secondly, they have fewer classes in the fourth year since the main academic tasks in the second semester were graduation practice and graduation thesis writing. Hence, the lack of classroom learning and teaching affects and even hinders the development of students' English vocabulary capacity to some extent.

**Table 1.** Statistical table of lexical variation (Type-Token Ratio)

	First grade	Second grade	Third grade	Fourth grade
Token	5639	7278	8754	10026
Type	867	1352	1378	1650
TTR	58.47%	61.35%	63.45%	61.76%

### 4.2. Features of Lexical Sophistication Development

Lexical sophistication, the distribution of low-frequency vocabularies in writing texts, is one of the indicators to evaluate the breadth of productive vocabulary. In this study, lexical sophistication refers to the basic vocabulary list made by Laufer and Nation (1995): the first 1000 vocabulary (hereafter referred to as V1) is the most frequently used words, and the second 1000 vocabulary (hereafter referred to as V2) is the second most common high-frequency word, and the third vocabulary list (hereinafter referred to as V3) is a list of academic vocabulary (Coxhead 2000), containing 3107 types and 570 word families, and the fourth vocabulary (hereinafter referred to as V4) ) are low-frequency words that are not in the first three word lists. As is shown in table 2, throughout the four years, lexical sophistication in students' compositions is on the rise. However, compared with the complex vocabulary of native speakers with the proportion of about 20% (Laufer and Nation, 1995), English majors used generally low level of academic vocabulary and not-in-the-list vocabulary.

**Table 2.** Statistical table of lexical sophistication (LS)

	Type-First grade	Type-Second grade	Type-Third grade	Type-Fourth grade
V3	3.23%	4.67%	5.82%	5.43%
V4	3.64%	4.83%	5.02%	6.03%
LS	6.87%	9.50%	10.84%	11.46%

We have also noticed that it is common for Chinese English majors to use high-frequency words or simply repeat the same word in the same composition, like, for example, "More and more students are reading on-line today.", "It's more likely for people to develop addiction.", "People should read more to expand their knowledge." These examples illustrate that students' writing is limited by vocabulary and they cannot use different synonyms to increase the lexical sophistication of the composition.

### 4.3. Features of Lexical Density Development

Table 3. Statistical table of lexical density

	Token	Number of Content Words	Lexical Density
First grade	5639	3127	57.95%
Second grade	7278	4275	58.87%
Third grade	8754	5267	59.26%
Fourth grade	10026	6032	59.78%

Lexical density refers to the percentage of content words in a composition. To a certain extent, lexical density reflects the information of students' written capacity. It can be seen from Table 3 that the development of vocabulary density in the compositions of freshman to senior year students basically shows a linear growth trend. The results show that it takes a relatively long process for students to effectively master content words. A close observation of the self-built corpus has found that a major factor contributing to the improvement of lexical variation but a lack of lexical density is the excessive use of function words by students, such as personal pronouns (we, I, you, my) and modal verbs (will, can, should, etc.), resulting in a reduced vocabulary density in the composition and a small amount of information. Overall, students have limited ability to learn and master content vocabulary. Therefore, when teaching vocabulary and guiding writings, teachers should raise the students' awareness of the importance of enriching the content vocabularies in their English compositions.

### 5. Conclusion

By tracking the English writings of 30 English majors in grades one to four, this study finds that the development trends of different dimensions of lexical richness in English writing are distinctive. Lexical variation in writing is closely related to students' English proficiency, but the growth trend is non-linear. Students experience a vocabulary plateau phenomenon in their senior year. Lexical sophistication increases as students progress through grades. The lexical density shows that students rely too much on high-frequency words in English writing. This trend gradually declines with the deepening of English learning, while the use of academic vocabulary is on the rise.

Based on the above findings, we believe that firstly, in the teaching of second language writing, teachers should actively encourage students not only to pay attention to the content, structure, grammar and semantics of the composition, but also attach importance to improving the lexical richness of the composition. Students should be motivated to use academic

vocabulary and formal language style correctly and try to avoid the interference of spoken style. Secondly, teachers should help students to convert receptive vocabulary into productive vocabulary to the maximum extent, and design writing exercises in a sequenced manner, so that students can use the newly-acquired vocabulary accurately.

### Acknowledgment

Project supported by 2017 University of South China Teaching Reform Research Project (Project Approval Number: 2017XJG-YB02, "Study on the English Productive Vocabulary of College Students Based on Self-built Corpus"); 2016 Foreign Languages Cooperation Project of Social Science Fund of Hunan Province. (Project Approval Number: 16WLH35).

### References

- [1] Bachman L F. Modern language testing at the turn of the century: Assuring that what we count counts [J]. *Language Testing*, 2000, (1):1-42.
- [2] Cobb T. Analyzing late interlanguage with learner corpora: Quebec replications of three European studies [J]. *The Canadian Modern Language Review*, 2003, (3): 393-423.
- [3] Engber C A. The relationship of lexical proficiency to the quality of ESL compositions [J]. *Journal of Second Language Writing*, 1995, (2): 139-155.
- [4] Laufer B. The lexical profile of second language writing: Does it change over time [J]. *RELC Journal*, 1994, (2): 21-33.
- [5] Laufer B. The development of passive and active vocabulary in a second language: Same or different? [J]. *Applied Linguistics*, 1998,(2):255-271.
- [6] Laufer B. Vocabulary acquisition in a second language [J]. *The Canadian Modern Language Review*, 2003, (4): 565-585.
- [7] Laufer B & Nation P. Vocabulary size: lexical richness in L2 written production [J]. *Applied Linguistics*, 1995, (3):306-322.
- [8] Lemmouth Z. The relationship between grades and the lexical richness of students essays [J]. *Nordic Journal of English Studies*, 2008, (3):163-180.
- [9] Leńko-Szymańska A. How to trace the growth in learners' active vocabulary? [A]. In Kettemann B & Morko G (eds.). *Teaching and Learning by Doing Corpus Analysis* [C]. Amsterdam: Rodopi,2002.
- [10] Liu Donghong. The role of vocabulary in English writing [J]. *Modern Foreign Languages*, 2003, (2): 181-186.
- [11] Read J. *Assessing Vocabulary* [M]. Cambridge: Cambridge University Press, 2000.
- [12] Wan Lifang. A study on lexical richness in second language writing of Chinese English majors [J]. *Foreign Language World*, 2010, (1):40-46.