

Productive Vocabulary Development in EFL Writing: A Corpus-based Study

Changzhong Shao

School of Foreign Languages, Linyi University, Linyi Shandong, China

shaochangzhong@lyu.edu.cn

Keywords: English writing; Productive vocabulary; Corpus; Longitudinal study

Abstract. Based on a self-built corpus, the present research carries out a longitudinal study method, which explores the developmental tendency of productive vocabulary in English writing from the aspect of lexical frequency profile and the lexical complexity of vocabulary breadth. The study examines the final examination compositions of 30 tertiary English majors in the first six semesters with the method of Lexical Frequency Profile. The study result indicates that the development of lexical complexity in productive vocabulary breath presents a straight-line progress, it does not exist the plateau phenomenon during the development of productive vocabulary.

Introduction

Vocabulary is the core problem in language acquisition and using. There is no doubt that vocabulary plays an important role in second language writing. Since 1980s, more and more scholars at home and abroad focus on the study about vocabulary in second language writing. What's more, with the applying of vocabulary measure instrument and analysis software, the corpus study in second language writing becomes more scientific and convenient. On the whole, the studies about the second language writing mainly focus on the breadth and depth of vocabulary. The breadth of vocabulary refers to the lexical complexity, and it is regarded as the important index (Engber, 1995: 139; Laufer & Nation, 1995: 307; Lemmouh, 2008: 163; Wan Lifang, 2010: 40-41). "The Lexical Complexity is the ratio of words beyond the most frequent 2000 words in the subjects' compositions" (Read, 2000: 203). Read (2000: 204) points out that using the out 2000 high-frequency words ratio can reflect the lexical complexity intuitively. The higher is the out 2000 high-frequency words ratio, the higher is the English level of the learners.

Literature Review

The second language vocabulary is usually divided into passive vocabulary and active vocabulary, or receptive vocabulary and productive vocabulary. The productive vocabulary involves in the breadth and depth of vocabulary.

Laufer (1998) investigated the development of learners' receptive vocabulary size, productive vocabulary size, and their relationship in Israel, the results of which indicate that during one year's study in class, receptive vocabulary size increases dramatically, while there is no significant improvement in productive vocabulary size. As a result, Laufer (1995) concluded that the development of productive vocabulary size lags far behind of its receptive counterpart, because for the respective use, the learners only need to know a few features of a word while for productive use, they need precise knowledge of the word. In some way, productive vocabulary is more difficult to learn than receptive vocabulary. The progress of productive vocabulary can reveal the nature of the second language learners' vocabulary development.

Some research was also made on the interrelation between the English level and the productive vocabulary development stages. Jullian (2000) found that second language learners' productive lexical development suffers a plateau after their language proficiency reaches the intermediate level.

This phenomenon is in conformity to what English teachers usually observe in classroom instruction settings.

However, the results of former studies differ greatly for most of them are based on a small amount of data. Nowadays with the development of advanced computer technology, corpus linguistic make it possible to do research on a large database and it has become a tendency that corpus is applied to second language learning and researching. It is hoped that the findings of the present study can explore the developmental tendency of the productive vocabulary based on a longitudinal corpus, and will shed some light on pedagogical approaches of productive vocabulary acquisition.

In previous study, various methods have been used to measure learners' productive vocabulary. Laufer and Nation (1995: 309-311) summarized the four most popular measures used in the description of productive lexicon: Lexical originality (LO), Lexical density (LD), Lexical sophistication (LS), and Lexical variation (LV). The Lexical Frequency Profile (LFP) was originally proposed Laufer and Nation (1995: 311). It is an instrument to analyze the subjects' productive vocabulary in compositions. LFP takes a text as raw input, and outputs a profile that describes the lexical content of the text in terms of frequency bands. The bands are described as follows (Meara, 2005:32):

The first band includes the most frequent 1,000 words of English.

The second band includes the second 1,000 most frequent words.

The third band includes words not in the first 2,000 words of English but which are frequent in upper secondary school and university texts from a wide range of subjects.

The breadth of productive vocabulary has two measure indexes, one is the lexical richness that is the distribution of the vocabulary or the vocabulary size, and it can be counted in the method of Lexical Frequency analysis. The other is the complexity of the vocabulary, it refers to the degree learners use the high-frequency vocabulary and low-frequency. Generally speaking, the high-frequency vocabulary emerging and using frequency is high, acquired firstly, the low-frequency vocabulary is on the contrary. Therefore, the quantity use the low-frequency can reflect the second language (especially the vocabulary) level of the learners in a certain degree. The complexity of the vocabulary can be measured by investigating the vocabulary rank that the low-frequency word exists in. The researchers at home and abroad have explored the development of the breadth of the productive vocabulary, but they are differs in conclusions.

Laufer (1995: 268) studied the effect of comprehension-based instruction on active (i.e. productive) vocabulary. The subjects were 37 adult EFL learners. They were asked to write a composition entitled 'How important is science to the modern world?' at the beginning of the experiment. Their compositions were analyzed in terms of the basic 2000 words and beyond 2000 words of LFP. The same texts were repeated at the end of the semester. The conclusion was that teaching vocabulary for comprehension would result in an increase in the passive (i.e. receptive) vocabulary, but this increase would not affect the growth of the active lexicon.

Laufer (1998) investigated the gains on the three types of EFL vocabulary knowledge, passive, controlled active and free active, in one year of school instruction. It also examined how these aspects of lexical knowledge were related to one another and what changes occurred in these relationships after one year. The results showed the passive vocabulary size progressed very well, controlled active vocabulary progressed too but less than the passive. Free active vocabulary (as measured by LFP) did not progress at all. Passive and controlled active size scores correlated with each other. Free active vocabulary, on the other hand, did not correlated with the other two types.

That is to say, the productive vocabulary of subjects does not increase with the rise of the receptive vocabulary. Therefore, we can infer that during the development of the productive vocabulary, the competence to produce the low-frequency words seems to fossilization.

Other research indicates that the development of the productive vocabulary goes through different stages. For example, Tan Xiaochen (2006) studies Chinese tertiary-level English majors' productive vocabulary development by analyzing 157 students' compositions in an academic term with the help of Range and Wordsmith. The results indicate a close relationship between English proficiency and

learners' productive vocabulary development, and learners' vocabulary development can be divided into three stages, the rapid period in low grade, the interim period in middle grade, and the maintenance period in high grade. What's more, the learners' vocabulary development in the breadth was generally concomitant with that in the depth.

Research Design

Based on the texted corpus, the present study aims to explore the productive vocabulary developmental characteristic of 30 English Educational majors from Grade One to Grade Three in the aspect of the breadth of vocabulary. Although the produce of language includes oral and written two presentations, specific on the produce of learners' language, the two methods have great difference in quantity and quality. The spoken communication is confined by the time limitation in the information extract and process, it cannot be modified before submit as the written language to guarantee the veracity. Therefore, compared with the spoken language, the composition method has the abundant vocabulary, and it has less language mistakes. The composition of students can present their productive vocabulary developmental character entirely.

Research Questions. The current study adopts a longitudinal research design that covers compositions finished by 30 English Education majors in six tests from Grade One to Grade Three in college. It aims to explore the productive vocabulary developmental characteristic of English learners' in their compositions finished under the testing conditions. The research addresses the following questions:

1) What is the variation characteristic of Lexical Frequency Profile in the writings of English learners from Grade One to Grade Three?

2) What is the Lexical Complexity developmental characteristic of the English learners from Grade One to Grade Three?

Samples. On the basis of self-built corpus, the subjects are 30 English Education majors, we chose their compositions in six final examinations from Grade One to Grade Three as our samples, 180 materials in all, among which all are effective. These compositions are collected in six final examinations of 30 English majors. In this way, it is guaranteed that there was a relatively large sample and stable development patterns can be found in the same learners.

Instrument. Range 32 (Nation & Coxhead, 2002) is a computer software that used to analyze the vocabulary breadth and vocabulary depth. It is designed by Paul Nation and A. Coxhead. It has acquired the approval of the researchers at home and abroad, it is the most influenced and used widely wordlist to measure the vocabulary breadth of the learners. Range can be used to compare a text against vocabulary lists to see what words in the text are and are not in the lists, and to see what percentage of the items in the text are covered by the lists. It can also be used to compare the vocabulary of two texts to see how much of the same vocabulary they use and where their vocabulary differs.

Range can also be used to compare the vocabulary of up to 32 different texts at the same time. For each word in the texts, it provides a range or distribution figure (how many texts the word occurs in), a headword frequency figure (the total number of times the actual headword type appears in all the texts), a family frequency figure (the total number of times the word and its family members occur in all the texts), and a frequency figure for each of the texts the word occurs in. It can be used to find the coverage of a text by a certain word lists, create word lists based on frequency and range, and to discover shared and unique vocabulary in several pieces of writing (Wang & Liang, 2007: 43-55).

Data Collection and Processing. The study aims at examining learns' productive vocabulary development in testing conditions. So compositions of 30 English majors are collected in the six final examinations. The compositions are then typed into the computer and checked to make sure there was no mistake in typing.

All the 180 compositions were typed into computer, and then each of them was analyzed by LFP. Laufer (1995) points out that the out 2000 high frequency words can be used to indicate the user's

LFP, this method is reliable and effective. The current study used the percentage of low frequency words to indicate subjects' productive vocabulary size, because in terms of LFP, "a higher percentage of low frequency words mean a larger productive vocabulary size" (Liu, 2004). In addition, the academic words and off-list words are regarded as together as the low frequency words rather than being distinguished when calculating the low frequency words.

Results and Discussions

Table 1 The Variation Characteristic of Lexical Frequency Profile

The Lexical Frequency Profile (Types)				
	Baseword1	Baseword2	Baseword3	Off the list
Grade One	64.75%	11.75%	7.35%	16.20%
Grade Two	61.50%	13%	8.40%	17.10%
Grade Three	58.05%	12.20%	10.65%	19%

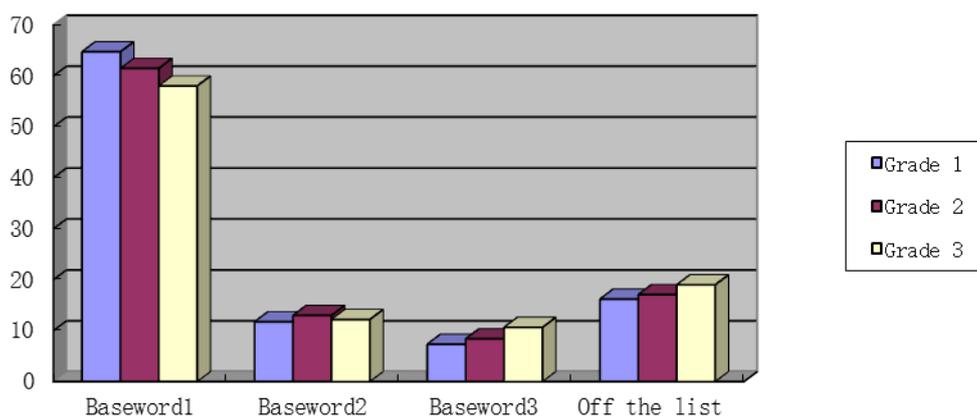


Figure 1. The LFP of the subjects' productive vocabulary

Fig. 1 demonstrates that the ratio of using high-frequency words decreases from Grade One to Grade Three. In general, the students use more on the first high-frequency words, and rely less on the words in other three wordlists. The mean rate of the subjects uses the first 1000 high-frequency words in grade one is 64.75%, 61.5% in grade two, and 58.05% in grade three. Compared with grade one, the rate decreases in grade two and grade three. The use of low-frequency words presents an increasing tendency from Grade One to Grade Three. The highest percentage that uses the high-frequency words is in grade one. And the highest ratio of academic words is in grade three. In the present study, the students in Grade Two and Grade Three use the highest percentage of wordlist2 and wordlist3 (academic word), it is consistent with the development of the second language acquisition. The result shows that the students' second language acquisition transits from the high frequency vocabulary to the low frequency vocabulary.

Table 2 The Developmental Characteristic of Lexical Complexity

	Types (Grade One)	Types (Grade Two)	Types (Grade Three)
Baseword3	7.35%	8.40%	10.65%
Baseword4	16.20%	17.10%	19%
Lexical Complexity	23.55%	25.50%	29.65%

From the lexical complexity result of Table 2, the types that out the wordlist1 and wordlist 2 are 23.55%, 25.50% and 29.65% from Grade One to Grade Three. The Table 3 further suggested that the lexical complexity in the English writings presents an increasing tendency from Grade One to Grade Three. The present research indicates that the English majors make a progress annually in the aspect of lexical complexity from Grade One to Grade Three. The result coincides with the related research results, such as Laufer (1994), Bao Gui (2008). However it is disaccord with the study result of Cui Yanyan and Wang Tongshun (2006). The different results of these researches may be influenced by the factors of the English level of the subjects, the requirement of collection corpus, the theme of the writings and the measured methods. According to the linguistic learning hypothesis theory (Wen, 2006, p.193-194), the linguistic skills learning experiences a period of obvious progress, it may be slow, stagnated and retrogress. Although the ratio of wordlist 3 and wordlist 4 is low during the period from Grade One to Grade Three, it does not exist the paused and declined plateau phenomenon. It predicts that the English majors have a great potential in the aspect of lexical complexity to improve the second language writings.

In current study, vocabulary ‘plateau’ does not appear at the period from Grade Two to Grade Three. This result does not confirm “Grade Two Phenomenon” proposed by Chen Yaping (1997). The development of lexical complexity presents linear developmental tendency, the students use the ratio of academic words increases obviously from Grade One to Grade Three.

Conclusion

By analyzing 30 English majors’ compositions in writing, the present study finds that the different developmental characteristic about the two aspects of productive vocabulary breadth. In current study, vocabulary “plateau” does not appear at the period from Grade Two to Grade Three. This result does not confirm “Grade Two Phenomenon” proposed by Chen Yaping (1997). The variation characteristic of lexical frequency profile indicates that with the rise of grade, the students gradually reduce to use the ratio of the high frequency words, and begin to use more academic words. The characteristic of the development of lexical complexity presents linear developmental tendency, the students use the ratio of academic words increases obviously from Grade One to Grade Three.

The present research also finds that with the improvement of English level, the ratio of using the first high-frequency words declines from Grade One to Grade Three and the use of low-frequency words increases year by year. The lexical complexity increases gradually with the rise of grade.

The important enlightenment of the data analysis result of the study to the English teaching is that productive vocabulary is an important character to reflect the learners’ English level. Therefore, to assist students to develop the productive vocabulary is an important task in English teaching. In general, to facilitate the development of the students’ productive vocabulary, the fundamental principle is to provide more chances of productive vocabulary, that is, the opportunity of output and using language. From the aspect of cognized theory, the significance of output to learning is that it can promote the learners to process language deeply. However, from the present English teaching situation in China, little attention is paid to the language output.

In the four years of academic learning in college, the first two years is the key period. Therefore, in grade two, it is necessary to keep cautious to the fossilization of second language and to utilize all possible resources to aid the learners to go beyond the “plateau” and to reach a new climax in the following college years.

Reference

- [1] Bao Gui, A Survey of the development of the lexical Richness in L2 compositions from a multidimensional perspective, *J. Computer-assisted Foreign Language Education*. 5 (2008) 38-44.
- [2] Chen Yaping, Psychological explanation and countermeasures to “fossilization”, *J. Foreign Language Teaching and Research*. 4 (1997) 56-60.
- [3] Cui Yanyan & Wang Tongshun, Development and relationships of receptive vocabulary size, productive vocabulary size and depth of vocabulary knowledge, *J. Modern Foreign Languages*. 4 (2006) 392-400.
- [4] Engber C A., The relationship of lexical proficiency to the quality of ESL compositions, *J. Journal of Second Language Writing*. 2 (1995) 139-155.
- [5] Jullian, P., Creating word-meaning awareness, *J. ELT Journal*. 1 (2000) 37-46.
- [6] Laufer B., The lexical profile of second language writing: Does it change over time, *J. RELC Journal*. 2 (1994) 21-33.
- [7] Laufer B. I. & S. P. Nation., Vocabulary Size and Use: Lexical Richness in L2 Written Production, *J. Applied Linguistic*. 3 (1995) 307- 322.
- [8] Laufer B. Beyond 2000: a measure of productive lexicon in a second language, in: L. Eubank, L. Selinker Benjamins & M. Sharwood-Smith (Eds), *The Current State of interlanguage*, John Benjamin, Amsterdam, 1995, pp. 265-272.
- [9] Laufer, B., The development of passive and active vocabulary in second language: Same or different? *J. Applied Linguistic*. 19 (1998) 255-271.
- [10] Lemmouh Z., The relationship between grades and the lexical richness of student essays, *J. Nordic Journal of English Studies*. 3 (2008) 163-180.
- [11] Liu Donghong, The influence of writing strategies and productive vocabulary size to writing quality, *J. Modern Foreign Languages*. 3 (2004) 302-310.
- [12] Lu Min, Development characteristics of productive vocabulary knowledge size: a research on English compositions, *J. Foreign Language Learning Theory and Practice*. 2 (2008) 10-13.
- [13] Meare, P., Lexical Frequency Profiles: a Monte Carlo analysis, *J. Applied Linguistic*. 26 (2005) 32-47.
- [14] Nation P. & Coxhead A., Retrieved Nov.10, 2015, from <http://www.vuw.ac.nz./lals/staff/Paul.Nation>.
- [15] Read J., *Assessing Vocabulary*, Cambridge University Press, Cambridge, 2000.
- [16] Read J., Research in teaching vocabulary, *J. Annual Review of Applied Linguistic*. 24 (2004) 146-161.
- [17] Tan Xiaochen, A research on the productive vocabulary development of English learners in China, *J. Foreign Languages Teaching and Research*. 3 (2006) 202-206.

- [18] Wan Lifang, A research on vocabulary richness in English writing of EFL learners in China, J. Foreign Language World. 1 (2010) 40-46.
- [19] Wang Lifei & Liang Maocheng, Computer-aided second language research methods and their applications, Foreign Language Teaching and Research Press, Beijing, 2007.
- [20] Wen Qiufang, The trend and characteristics of English majors' spoken words change, J. Foreign Language Teaching and Research. 3 (2006) 189-197.