

ANALYSIS ON THE DEVELOPMENT OF LEXICAL COMPLEXITY IN CHINESE SCIENCE STUDENTS' ENGLISH WRITING

Wanyi Du

Foreign Language Educational Centre, Dalian University of Technology, Liaoning, China

Received: 26-June-2019, **Revised:** 24-July-2019, **Accepted:** 29-July-2019, **Published:** 03-August-2019

ABSTRACT: This study analyzes the development characteristics of vocabulary complexity in written English of the students from key science and engineering universities from the aspects of the length of word, distribution of long words and parts of speech, type-and-token ratio (TTR) and Lexical Frequency Profile. The results show that, from the perspective of development, with the improve of English proficiency and age, the complexity of written vocabulary of science-and-engineering students will, in general, meet a gradually deepening and improving situation; the number of long words has increased significantly and the parts of speech tend to be more abundant; the rate and quantity of low-frequency words in writing are increasing.

Keywords: Science Students' English Writing, Lexical Complexity, Low-Frequency Words.

1. INTRODUCTION

Writing competence, as a key point of high-level foreign language learning, is an critical indicator for language competence assessment. It includes language competence and knowledge in several dimensions. However, the current study holds that vocabulary is still the central part of writing competence. The research shows that the lexical complexity and breadth of high-level second language learners (first language learners) are much higher than those of general-level learners. In the process of writing, second language learners also pay more attention to the use of vocabulary.

Typical indicators of lexical complexity are the length of word, number of long words and type-and-token ratio (TTR) (Xiaoling, 2007). In addition, Laufer and Nation's Lexical Frequency Profile is often used to analyze lexical complexity. The ratio of academic vocabulary to low-frequency vocabulary in a text is considered to be the main indicator of written vocabulary competence.

At present, most articles in China focusing on lexical complexity are conducted through cross-sectional empirical analysis. Cross-sectional study examines a learner's writing characteristics in the same period or stage, while longitudinal study of writing, also known as study of writing development, explores the a learner's writing characteristics at different stages of language development. At present, (Qin and Wen, 2007) have made a longitudinal empirical study on the features and characteristics of ability development of English majors' English writing, which is the first corpus-based study of the development system of English writing in China. However, the review found that the number of longitudinal study on vocabulary complexity is small and their objects are all English majors, and the development characteristics of vocabulary complexity of the students from science and engineering universities still remain a research gap. There are great differences in English learning between science-and-engineering students and English majors from the perspectives of teaching requirements, curriculum designs and teaching methods. Therefore, the written lexical complexity of science-and-engineering students still has research value.

2. RESEARCH DESIGN

From the perspective of lexical complexity, this study tends to explore the developmental characteristics of lexical complexity in writing. The questions to be answered in this study are as follows.

1. Specific manifestations of the development characteristics of lexical complexity in science-and-engineering students' written English, including the average length of word, number of long words, TTR, the distribution of high- and low-frequency words, nominalization and so on.
2. Does the lexical complexity in science-and-engineering students' writing improve with the increase of English proficiency and age?

In order to ensure the objectivity of the data, the author has built two written corpus in small scales. The corpus included the university entrance examination compositions of nearly 150 students majoring in Network Engineering in a key Science and Engineering University in 2008 (the beginning of grade one), and their compositions in CET4 (the beginning of grade two) and CET6 simulation test (the beginning of grade three). These students' average score of university entrance examination is 570, 510 in CET4 and 476 in CET6. Their English writing proficiency can generally reflect the overall English writing competence of the students from key science and engineering universities. In the process of corpus collection, the title, requirements and scoring criteria of the compositions have all been referred to the relevant instructions of CET4 and CET6. The compositions are completed within the prescribed time and retrieved on the spot without modification. All the compositions selected are more than 6 points (max.80 points). Detailed data of the corpus are as follows (Table 1):

Table 1. Introduction to the Corpus

| | Beginning of freshman year | Beginning of sophomore year | Beginning of junior year |
|------------------------|----------------------------|-----------------------------|--------------------------|
| Number of compositions | 518 | 502 | 463 |
| Number of types | 64436 | 67951 | 67296 |
| Number of tokens | 2286 | 4537 | 3587 |

Source: Results from self-constructed corpus

Firstly, the author uses two applications, text combining. exe and text sorting. exe, to process the text before the analysis, and uses the tagging software, TreeTagger, to tag the part of speech in the sorted text with codes. Then, the retrieval software AntConc written by Laurence Anthony and the hierarchical analysis software of vocabulary, AntWord Profiler, are used to complete the task of retrieval and hierarchical analysis.

3. DATA ANALYSIS WITH RESULTS

3.1. Analysis on Average Length of Words and Number of Long Words in Science-and-Engineering Students' Writings

Table 2. Comparisons between Average Length of Words and Number of Long Words

| | Beginning of freshman year | Beginning of sophomore year | Beginning of junior year |
|-------------------------|----------------------------|-----------------------------|--------------------------|
| Average Length of Words | 3.1 | 4.02 | 3.28 |
| 10-letter word | 745 | 1310 | 1309 |
| 11-letter word | 225 | 550 | 664 |
| 12-letter word | 136 | 332 | 153 |
| 13-letter word | 84 | 192 | 68 |
| 14-letter word | 26 | 43 | 22 |
| 15-letter word | 9 | 18 | 3 |
| 16-letter word | 4 | 15 | 4 |
| 17-letter word | 0 | 5 | 1 |
| 18-letter word | 0 | 1 | 2 |
| 19-letter word | 0 | 0 | 0 |
| 20-letter word | 0 | 1 | 0 |
| Total | 1229 | 2467 | 2226 |
| Standardization | 537 | 544 | 621 |

Source: Results from self-constructed corpus

The results of Table 2 show that, in the three small corpora, the average length of words gets the highest, 4.02, at the beginning of sophomore year. The author refers to the data of English majors' compositions of Written English Corpus of Chinese Learners (WECCCL), and the average length of words is 4.72 (Changsheng, 2007). Compared with the argumentative writing of English majors, for the key universities of science and Engineering in this study, the average length of words in the compositions of CET4 stimulation text is 4.02 which can be generally counted as ideal data. However, we find that the

average length of their written words at the beginning of junior year is 0.74, which is lower than that at the beginning of sophomore year. Viewing from the data of the average length of words, lexical competence does not seem to improve with the increase of English proficiency and age at the beginning of their junior year.

The number of long words is considered to be another feature of lexical complexity. We have counted the frequency of long words in the three corpuses and made them standardized (Table 2). It shows that the number of long words in science-and-engineering students' writings reaches 544 at the beginning of sophomore year, 621 at the beginning of junior year and 537 at the beginning of freshman year. From the total frequency of long words, the number of long words in sophomores' writing is not significantly higher than that in juniors'; in early junior year, their number of long words is 10 percent higher than that in sophomore year. Of course, the use of long words is directly related to the style, title and writing time of an article. In order to further understand the differences in the use of long words, the author further counts the specific part-of-speech distribution of long words (Table 3).

Table 3. Part-of-Speech Distribution of Long Words

| | Beginning of freshman year | Beginning of sophomore year | Beginning of junior year |
|----------------------|-----------------------------------|------------------------------------|---------------------------------|
| Noun | 81.5 % | 63.7 % | 61.2 % |
| Verb | 3.7% | 5.6% | 5.8% |
| Adjective/ Adverb | 12.6% | 28.4 % | 31.5% |

Source: Results from self-constructed corpus

In the compositions of juniors' entrance examination, the long words are mainly nouns, accounting for 81.5% (Table 3); other parts of speech are significantly fewer, while the lexical richness is also scarce. In the compositions of simulated CET4 in early sophomore year, the frequency of nouns decreases to 63.7%, and the frequency of adjectives and adverbs increases to 28.4%, such as especially, accordingly, increasing. In the compositions of simulated CET6, the frequency of long nouns decreases to 61.2%, while that of adjectives and adverbs increases to 31.5%, such as particular, consideration.

3.2. Type-and-Token Ratio (TTR) of Vocabularies in Science-and-Engineering Students' Writings

To a certain extent, the type-and-token ratio (TTR) can reflect the variability of the words used in a text. The higher the TTR of a text is, the more the changes in words are and the more the complexity is. Therefore, this study analyzed the lexical complexity characteristics of science-and-engineering students at different stages through the analysis of TTR.

Table 4. Comparisons of TTR

| | Beginning of freshman year | Beginning of sophomore year | Beginning of junior year |
|-------------------------|-----------------------------------|------------------------------------|---------------------------------|
| TTR | 3.55 | 6.68 | 5.33 |
| Standardized TTR | 5.51 | 9.83 | 7.92 |

Source: Results from self-constructed corpus

Data in table 4 shows that the standardized TTR of written vocabulary in sophomore year reaches the highest 9.83, while it is only 5.51 of university entrance examination, which is 4.32 higher. It shows that the students' written vocabulary has been greatly expanded within the first year of university, and TTR has nearly doubled. The standardized TTR of junior year is 7.92, slightly behind the data of sophomore year. It sounds a cautionary note again for the efficiency of the students' English learning after the CET4 test. According to the research of WECCL, the standardized TTR is 33.23 (Ping, 2009) and 12.28 (Changsheng, 2007). Viewing only from the figures, the TTR in the science-and-engineering students' writings in this study is greatly unsatisfactory. Because the accuracy of TTR data is closely related to the size of the corpus text, we can't arbitrarily judge the lexical complexity of science-and-

engineering students' second language writing; we can only infer that the lexical complexity of their written English is much lower than that of English majors.

3.3. Lexical Frequency Profile of Science-and-Engineering Students' Writings

The average length of word, the number of long words and TTR are all data of quantitative comparison, while the analysis on lexical frequency profile of written vocabulary at different stages is to accurately assess the vocabulary complexity of students from a qualitative way. The software AntWordProfiler written by Laurence Anthony, on the basis of Range 3.2 written by Nation et al, is a more convenient analysis software for lexical frequency profile. The results of data analysis are as follows.

Table 5. Hierarchical Comparison of Vocabularies (unit: type)

| | Beginning of freshman year | | | Beginning of sophomore year | | | Beginning of junior year | | |
|----------------------|----------------------------|-----------------|-------|-----------------------------|-----------------|-------|--------------------------|-----------------|-------|
| | Original | Standardization | | Original | Standardization | | Original | Standardization | |
| Basic vocabulary | 1000 | 437 | 80.7% | 1587 | 350 | 62 % | 1381 | 385 | 64.1% |
| Academic vocabulary | 228 | 99 | 18.2* | 552 | 121 | 21.5% | 417 | 116 | 19.3% |
| Technical Vocabulary | 118 | 5 | 9.1% | 420 | 93 | 16.5% | 360 | 100 | 16.6% |
| Total | 1346 | 541 | 100 % | 2559 | 564 | 100 % | 2158 | 601 | 100 % |

Source: Results from self-constructed corpus

Overall, the data show that, with the extension of English learning time and the improvement of English proficiency, the basic vocabulary (high-frequency vocabulary) in science-and-engineering students' writings has gradually decreased (Table 5), from 80.7% in freshman year to 62% in sophomore year and 64.1% in junior year. On the contrary, the low-frequency vocabulary (academic vocabulary and Technical Vocabulary) shows a gradually increasing trend. Especially for Technical Vocabulary, the vocabulary of junior year has reached 16.6%, which is 20 times higher than that of university entrance examination. The continuous increase of academic vocabulary and Technical Vocabulary has indicated that the students' lexical complexity keeps increasing gradually.

However, compared with the data of international second language learners, the percentage of low-frequency words in Chinese sophomores' writings is 38 percent, still being a lower data. In Laufer (1998) study, the percentage of low-frequency words (corresponding to more than 2000 high-frequency word families in his study) of free active vocabulary in Israeli senior three students' writings is nearly 70 percent. In this study, the actual number of low-frequency words (academic vocabulary and Technical Vocabulary) reaches 216 at the beginning of junior year, while the gap is still obvious when compared with the data of English majors (Changsheng, 2007). Therefore, the lexical complexity of Chinese science-and-engineering students' writing still has relatively great space for improvement.

4. CONCLUSION

This paper has conducted a contrastive analysis on the development characteristics of vocabulary complexity in written English of the students from key science and engineering universities at different stages with the hierarchical research method, in order to verify whether the vocabulary complexity in science-and-engineering students' writings increases with the improvement of their English proficiency, and the increase of the time of the learning English and their age. Through contrastive analysis on the indicators including the length of word, distribution of long words and parts of speech, type-and-token ratio (TTR) and Lexical Frequency Profile, we have found that:

1. Generally speaking, in the view of development, the written lexical complexity of science-and-engineering student has increased significantly from entering university to the beginning of junior year, and several indicators have shown a steady increase trend.
2. The improvement of written lexical complexity shows a straight upward trend especially from entering university to the beginning of sophomore year (before CET4 test); the average length of words and the number of long words increase, the part of speech of long words tends to be

richer, the ratio of low-frequency words continuously increases, and the ratio of high-frequency words suffers from a gradual decrease.

3. At the beginning of junior year (before CET6 test), the improvement of the written lexical complexity tends to be flat in quantity, but the further increase of the low-frequency words and the increasing richness of vocabularies of different parts of speech show that the improvement of written lexical complexity of science-and-engineering students tends to be more rational and balanced.

According to the results above, we can draw the following conclusions: with the improvement of second language proficiency and the increase of English learning time, the written lexical complexity of students in key science and engineering universities is also increasing. Because this study explores the written lexical complexity of science-and-engineering students from the perspective of longitudinal development, focusing on the lexical level, thus there are limitations in the research indicators and scopes. Follow-up research can consider to make the corpus more diversified and analyze the development characteristics of science-and-engineering-students' written language with the consideration of syntactic complexity. The data will be more convincing.

Through this study, the author believes that it is necessary to introduce the concept of lexical complexity into the teaching of College English Writing. Studies at home and abroad both show that, to some extent, there is a positive correlation between lexical complexity and writing grades (Xiaoling, 2007). In the long run, in the era of sharing academic achievements globally, English writing competence of the students from key science and engineering universities represents the international communication ability of Chinese scientific researchers. Glancing over the English abstracts of the science-and-engineering students' final dissertation in China, the common problem is that the language tends to be simple. In the English teaching for the science-and-engineering students, especially in the advanced stage in senior grade, we can try to practice teaching methods such as combining reading and writing and combining of writing and speaking, and get additional teaching measures such as English-English paraphrase and abstract writing, so as to train the students to gradually improve the written lexical complexity, reach better coherence of text and improve the formality of written style.

REFERENCE

- Changsheng, L. (2007). Do Chinese college students really need vocabulary complexity? A study based on SWECCCL- a learners' corpus. *Jiangsu Foreign Language Teaching and Research*, 1: 26-30.
- Laufer , B. (1998). The development of passive and active vocabulary in a second language : Same or different. *Applied Linguistics*, 19(2): 256-71.
- Ping, Z. (2009). Word Association Patterns in Chinese EFL; Learners' Mental Lexicon. Dongnan University Press.
- Qin, X. and Wen, Q. (2007). EFL Writing of College English Majors in China:A Developmental Perspective[M].Beijing:China Social Sciences Press. 56-95.
- Xiaoling, J. (2007). The development of language formality in compositions of english majors in China and its relationship with writing score. *Shandong Foreign Language Teaching Journal*, 6: 51-55.