2016 2nd International Conference on Social Science and Development (ICSSD 2016)
ISBN: 978-1-60595-356-4

A Research on Evaluation of Written Productive Vocabulary Based on Corpus

Yao-hua HUANG[1,a,*] and Lei-jiang SU[1,b]

[1]North China Electric Power University, Baoding 071003, Hebei, China

*[a]695062724@qq.com, [b]suleijiang@yahoo.com.cn

*Corresponding author

Keywords: Productive Vocabulary, Lexical Variation, Lexical Sophistication.

Abstract. Lexical Variation (LV) and Lexical Sophistication (LS) are selected in this corpus-based study as indexes to evaluate students' productive vocabulary in their writing. Data analysis has shown that writing quality is strongly influenced by students' written productive vocabulary, with LV and LS being major factors.

Introduction

Productive vocabulary has an essential influence on learners' writing ability[1]. Learners' ability to output vocabulary can be reflected by the evaluation of the productive vocabulary in the learners' written and spoken corpus[2]. In other studies of this kind, there are many indexes of evaluation on productive vocabulary, such as Lexical Originality (LO), Lexical Density (LD), Lexical Sophistication (LS), Lexical Variation (LV), and so on. This paper selects two, which are less controversial: Lexical Variation and Lexical Sophistication to study written productive vocabulary in a self-built corpus of students' compositions.

Basic Concepts

Productive Vocabulary

In vocabulary acquisition, it is generally believed that the acquisition of vocabulary is a progressive continuum, expanding from the acquisition of receptive vocabulary to the acquisition of productive vocabulary[8]. Receptive vocabulary is the word whose the most basic and core meaning can be understood by learners; whereas productive vocabulary is the vocabulary which can be used freely and spontaneously in learners' spoken and written expressions. Receptive vocabulary is often related with vocabulary size; while productive vocabulary is concerned with vocabulary's practical application, thus reflecting learners' vocabulary depth.

Lexical Variation

Lexical richness is generally calculated through type-token ratio (TTR). The formula is TTR=type/token. The TTR value is between 0 and 1. Greater absolute value means the lager vocabulary size. In other words, the richest information can be conveyed with the least words.

Lexical Sophistication

Laufer and Nation [9] pointed out that Lexical Frequency Profile (LFP) is the most reliable instrument to examine vocabulary sophistication. It focuses on the scale of frequency to determine the realm of each word. Laufer and Nation used readily proven four scales of frequency, that is, the most commonly used baseword 1 (approximately 1000 word family), commonly used baseword 2 (approximately 1000 word family), sub-commonly used baseword 3 and words off the list. In this paper, the evaluation of Lexical Frequency Profile is mainly obtained by referring to the frequency of productive vocabulary in learners' compositions in baseword 3, the sub-commonly used word list.

Experiment

Corpus and Instrument

Corpus samples in this study are 30 compositions written by sophomores of English majors under the same title within time limit. The measuring instruments used in this study are AWP (Ant Word Profiler) and RANGE 32.

Statistics

This paper uses AWP to calculate the type-token ratio, which reflects lexical variation in each composition. The results are shown in Table 1.

Table 1. Type-token ratio of each composition.

Essay	1	2	3	4	5	6	7	8	9	10
Type	76	83	86	87	105	65	94	85	96	106
Token	168	143	158	158	173	128	158	144	157	182
TTR	0.45	0.58	0.54	0.55	0.61	0.51	0.59	0.59	0.61	0.58
Essay	11	12	13	14	15	16	17	18	19	20
Type	70	63	85	95	94	78	76	108	97	91
Token	116	122	142	153	149	120	117	165	156	142
TTR	0.60	0.52	0.59	0.62	0.63	0.65	0.65	0.65	0.62	0.64
Essay	21	22	23	24	25	26	27	28	29	30
Type	66	59	68	93	88	62	82	79	89	96
Token	109	112	128	182	132	134	192	160	156	180
TTR	0.61	0.53	0.53	0.51	0.67	0.46	0.43	0.49	0.57	0.53

This paper adopts Range software to analyze lexical frequency profile, and then to calculate the ratio of low-frequency words in each composition. In order to avoid misspelled words in compositions are mistakenly regarded as the low-frequent vocabularies, the author makes manual correction of spelling errors in each composition before the software calculation. The results are shown in Table 2:

Table 2. Low-frequency word ratio in each composition.

Essay	1	2	3	4	5	6	7	8	9	10
type/ type%	5/ 5.7	11/ 10.5	12/ 12.8	14/ 16.3	9/ 13.6	11/ 12.9	9/ 11.5	12/ 12.8	17/ 16.7	5/ 7.3
Essay	11	12	13	14	15	16	17	18	19	20
type/ type%	2/ 2.4	5/ 6.4	11/ 11.3	6/ 7.4	7/ 7.4	15/ 15.0	9/ 9.6	8/ 8.8	4/ 5.7	8/ 9.1
Essay	21	22	23	24	25	26	27	28	29	30
type/ type%	5/ 8.4	4/ 4.7	1/ 1.1	5/ 5.2	3/ 4.8	5/ 6.9	3/ 4.0	7/ 9.0	5/ 5.7	13/ 12.8

Then, make a correlation analysis between type-token ratio and score of each sample composition with Person correlation analysis in SPSSV13.0, and the results are shown in Table 3:

Table 3. Correlations between Sugeno result and score.

		Score	Integrated result
Score	Pearson Correlation	1	.615(**)
	Sig. (2-tailed)		.000
	N	30	30
Sugeno result	Pearson Correlation	.615(**)	1
	Sig. (2-tailed)	.000	
	N	30	30

** Correlation is significant at the 0.01 level (2-tailed)

Sig. (2-tailed)= .000 < 0.05, indicates that type-token ratio is significantly related to scores of sample compositions.

In addition, make a correlation analysis between the ratio of low-frequency words and score of each sample composition with Person correlation analysis, and the results are shown in Table 4:

Table 4. Correlations between Sugeno result and score.

		Score	Integrated result
Score	Pearson Correlation	1	.717(**)
	Sig. (2-tailed)		.002
	N	30	30
Sugeno result	Pearson Correlation	.717(**)	1
	Sig. (2-tailed)	.002	
	N	30	30

** Correlation is significant at the 0.01 level (2-tailed)

Sig. (2-tailed)= .002 < 0.05, indicates that the ratio of low-frequency words is significantly related to scores of sample compositions.

Summary

Productive vocabulary is an essential factor affecting learners' communicative competence. This paper analyzes students' written productive vocabulary in terms of lexical variation and lexical sophistication. The significant correlation between lexical variation, lexical sophistication and scores of sample compositions indicates that vocabulary is an essential factor affecting learners' writing competence.

Due to limited conditions, this paper still has some limitations. Corpus of larger scale should be used to make more profound research on writing test discrimination by examining it from more angles (like vocabulary, syntax, discourse, and so on) in the future.

Acknowledgement

This research was financially supported by "The Fundamental Research Funds for the Central Universities".

This research was also financially supported by "Education and Teaching Reform Research Project in North China Electric Power University Science and Technology College".

References

[1] G.H. Ma and Q.F. Wen, The relationship of L2 learners' linguistic variables to L2 writing ability, Foreign Language Teaching and Research. 31 (1999) 34-39.

[2] S.P. Duan, L2 lexical competence and its assessment, Journal of PLA University of Foreign Languages. 32 (2009) 51-54.

[3] B. Laufer and Z. Goldstein, Testing vocabulary knowledge: size, strength and computer adaptiveness, Language Learning. 54 (2004) 399-436.

[4] B. Laufer and P. Nation, Vocabulary size and use: lexical richness in L2 written production, Applied Linguistics. 16 (1995) 307-322.